

List of Science Papers - #1



UEDIN for the Astrodeep project



ASTRODEEP

"Unveiling the power of the deepest images of the Universe"

THEME [SPA.2012.2.1-01]

[Exploitation of space science and exploration data]

Grant agreement for: Collaborative project

Grant agreement no: 312725

Deliverable 5.1

ABSTRACT

In this document we present the scientific publications obtained within ASTRODEEP during the period 01/01/2013 – 31/12/2013.

Prepared by: J. Dunlop.

Approved by: AEC

Date: 30 December 2013

Summary of Science Papers

Due to the inevitable gestation period, science papers that have genuinely benefitted from ASTRODEEP funding have only been completed and submitted towards the end of 2013. Nevertheless, we can already report on 6 refereed journal papers that have now been submitted for publication. These papers all present significant scientific advances based on early versions of catalogues within the key survey fields that are the focus of the ASTRODEEP project. In several cases the analysis undertaken en route to science results has revealed both the strengths and deficiencies of existing catalogues, reaffirming the central premise of ASTRODEEP that high-quality data products can only be produced and validated in conjunction with cutting-edge research. Below we provide a brief summary of the key results from each of these papers, describing in each case the status of the new datasets from which they were derived.

Bowler, Dunlop et al., 2013, MNRAS, submitted (arXiv:1312.5643)

“The bright end of the galaxy luminosity function at $z \sim 7$: before the onset of mass quenching?”

This paper presents the first science results based on the second data release (DR2) of the UltraVISTA survey, and the latest (DR10) release from the UKIDSS UDS survey. These two surveys are providing the key near-infrared imaging across the two largest (degree-scale) fields to be analysed within the ASTRODEEP project. However, the proper exploitation of these data in the study of high-redshift galaxies requires that they be combined in a resolution-matched way with optical (Subaru and CFHT) and mid-infrared (Spitzer) data. In this study a first version of such a combined multi-wavelength catalogue has been used to produce a new sample of ~ 35 luminous galaxies at redshift $z \sim 7$. The most important result is that such galaxies appear to be more numerous than expected based on the standard (Schechter function) extrapolation from the number density of fainter $z \sim 7$ galaxies (as revealed by smaller-area deeper surveys with HST). Interestingly, this suggests the form of the bright-end of the galaxy luminosity function at $z \sim 7$ simply mirrors the high-mass end of the dark-matter halo mass function, and hence that the physical mechanism which produces the very steep decline in the number of luminous galaxies seen in the low-redshift Universe has yet to impact on the galaxy population at $z \sim 7$ (within the first billion years of cosmic time). While undertaking the analysis reported in this paper, a problem was uncovered with the calibration of the UltraVISTA DR1 Y-band data, and this has now been corrected in DR2 (once again emphasizing the importance of catalogue production proceeding in parallel with scientific validation).

We attach here the submitted paper.

Rogers, McLure, Dunlop et al., 2013, MNRAS, submitted (arXiv:1312.4975)

“The colour distribution of galaxies at redshift five”

This paper presents new results on the colours of high-redshift galaxies, this time at $z \sim 5$, where the higher signal-noise and additional rest-frame coverage of the galaxy spectral energy distributions enables more detailed study of the controversial issue of the galaxy colour-magnitude relation (CMR) than at higher redshifts. Crucially, in order to achieve maximum dynamic range in luminosity, this work relies on combining ground-based and space-based (HST) imaging, a key focus of AstroDEEP. The galaxy samples and the associated galaxy photometry was derived from the HUDF12 imaging in the Hubble Ultra Deep Field, the HST CANDELS imaging in the GOODS North and South fields, and the ground-based imaging in

the UKIDSS UDS field. The paper presents two key results. First, it reaffirms and clarifies the form of the colour-magnitude relation at $z \sim 5$. Second, it presents the first statistically-significant evidence for a scatter in intrinsic colour at the redshifts, a variation which appears to increase with increasing UV luminosity. The conclusion is that the increasing width of the intrinsic galaxy colour distribution and the CMR itself are both plausibly explained by a luminosity independent lower limit on UV slope, combined with an increase in the fraction of red galaxies with increasing UV luminosity.

We attach here the submitted paper.

Castellano, Sommariva, Fontana et al. 2013, A&A, submitted

“Constraints on the star-formation rate of $z \sim 3$ LBGs with measured metallicity in the CANDELS GOODS-South field”

This paper is focused on the properties of galaxies at somewhat lower redshift, at $z \sim 3$ which corresponds to the peak epoch of star-formation activity in the Universe. The paper exploits the latest photometry in the GOODS South field to enable detailed fitting of the spectral energy distributions (SEDs) of a subset of $z \sim 3$ galaxies which possess spectroscopic redshifts. The power of this approach is that the spectroscopy not only provides the reassurance of accurate redshifts, but also helps to break some the degeneracies which often plague the extraction of robust physical parameters from SED fitting based purely on photometry. One key conclusion is confirmation that the appropriate metallicity to use in the fitting is ~ 0.1 times the solar value. Adopting this metallicity, and then fitting a range of star-formation histories to the combined optical-infrared photometry, yields star-formation rates for these objects which suggest that the star-formation density at $z \sim 3$ is even higher than previously thought. We note that this paper also makes use of the new near-infrared data in the GOODS-South field provided by members of our team through the VLT Hawk-I HUGS program, and thus provides another example of the importance of coherently combining ground-based and space-based data.

We attach here the submitted paper. A revised version addressing referee’s comments will be submitted soon.

Sommariva, Fontana et al., 2013, A&A, submitted

“A mass threshold in the number density of passive galaxies at $z \sim 2$ ”

This paper again utilizes the new HUGS K-band data in tandem with existing HST and ground-based data, but this time in both the GOODS-South and UDS fields, and in the study of passive galaxies at redshifts $z \sim 2$. Once again the importance of properly matching the space-based and ground-based photometry is emphasized, as here use is made of well-established BzK colour-criteria to select galaxies at $z \sim 2$ in which star-formation activity has been quenched. The key breakthrough enabled here is that, by extending the study of BzK selected passive galaxies a magnitude fainter than was previously possible, this work has revealed a clear turnover in the number density of quiescent galaxies below a stellar mass threshold of $M^* \sim 10^{10.8}$ solar masses. This result confirms (and better quantifies) previous suggestions that passive galaxies are confined to the high-mass regime at these redshifts, and comparison with the predictions of various current models of galaxy formation shows that none of them can properly reproduce this key feature of the evolving galaxy population. This work thus demonstrates that the evolving mass function of passive galaxies continues to present one of the key challenges for proposed models of galaxy formation and evolution.

We attach here the submitted paper, including revision after referee's report.

Fontana, Dunlop et al., 2013, A&A, submitted

"The Hawk-I UDS and GOODS Survey (HUGS): Survey Design and Deep K-band Number Counts"

This paper presents the survey design of the near-infrared Hawk-I UDS and GOODS Survey (HUGS) that was utilized in the afore-mentioned two studies. It also exploits these new, deepest-ever K-band data to explore the K-band galaxy number counts to previously unexplored depths. This is a crucial dataset for many studies of distant galaxies because HST is not capable of imaging at wavelengths longer than 1.6 microns. By providing deep, high-quality (0.4 arcsec seeing) imaging at 2.2 microns, HUGS fills a crucial gap in spectral coverage between the HST and Spitzer data in two of the CANDELS survey fields. This paper demonstrates that the HUGS K-band data are well-matched in depth to the CANDELS HST H-band imaging, validating the approach taken in designing the survey. It also shows that the slope of the deep K-band galaxy number counts depends sensitively on the assumed distribution of galaxy sizes, with potential impact on the estimated extra-galactic background light

We attach here the submitted paper.

Vanzella, Fontana et al., 2013, ApJ, 2013, submitted

"LBT/MODS1 spectroscopic confirmation of two faint sources at $z=6.4$ magnified by the CLASH / Frontier Fields cluster MACSJ0717.5+3745: towards the characterization of star-forming galaxies at the epoch of reionization"

This paper presents some of the very first results from analysis of the first HST Frontier Fields dataset. Specifically, the paper presents spectroscopic results from the Large Binocular Telescope which show that two of the faint galaxies gravitationally lensed by the galaxy cluster MACS0717.5+3745 lie at redshift $z = 6.4$. This redshift information, in tandem with the deep multi-wavelength HST imaging enables study of the SEDs of two faint $z \sim 6.4$ galaxies, galaxies which have lower luminosities than those which can be studied without the aid of gravitational lensing. These galaxies are, as expected, compact and of low stellar mass. However, perhaps the most interesting finding is that they seem to be extremely blue, displaying a UV continuum shape that appears to reflect the presence of extremely young and extremely low metallicity stellar populations.

We attach here the submitted paper.

The bright end of the galaxy luminosity function at $z \simeq 7$: before the onset of mass quenching?

R. A. A. Bowler^{1*}, J. S. Dunlop¹, R. J. McLure¹, A. B. Rogers¹, H. J. McCracken²,
 B. Milvang-Jensen³, H. Furusawa⁴, J. P. U. Fynbo³, Y. Taniguchi⁵, J. Afonso^{6,7},
 M. N. Bremer⁸, O. Le Fèvre⁹

¹*SUPA†, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ*

²*Institut d’Astrophysique de Paris, UMR 7095 CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, 75014 Paris, France*

³*Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark*

⁴*Astronomical Data Center, National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan*

⁵*Research Institute for Space and Cosmic Evolution, Ehime University, 2-5 Bunkyo-cho, Matsuyama 790-8577, Japan*

⁶*Centro de Astronomia e Astrofísica da Universidade de Lisboa, Observatório Astronómico de Lisboa, Tapada da Ajuda, 1349-018, Portugal*

⁷*Faculdade de Ciências da Universidade de Lisboa, Ed. C8. Campo Grande, 1749-016, Lisbon, Portugal*

⁸*H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL*

⁹*Laboratoire d’Astrophysique de Marseille, CNRS and Aix-Marseille Université, 38 rue Frédéric Joliot-Curie, 13388 Marseille Cedex 13, France*

ABSTRACT

We present the results of a new search for bright star-forming galaxies at $z \simeq 7$ within the UltraVISTA DR2 and UKIDSS UDS DR10 data, which together provide 1.65 deg^2 of near-infrared imaging with overlapping optical and *Spitzer* data. Using a full photometric redshift fitting analysis to identify high-redshift galaxies and reject contaminants, we have selected a sample of 34 luminous ($-22.7 < M_{UV} < -21.2$) galaxies in the redshift range $6.5 < z < 7.5$. Crucially, the deeper imaging provided by UltraVISTA DR2 confirms all of the robust objects previously uncovered from the DR1 data by Bowler et al. (2012), validating our selection technique. Our new expanded galaxy sample includes the most massive galaxies known at $z \simeq 7$ with $M_* \sim 10^{10} M_\odot$, and the majority are resolved in the ground-based imaging, consistent with a larger intrinsic size ($r_{1/2} \simeq 1 - 1.5 \text{ kpc}$) than that found for less massive galaxies at comparable redshifts (a result supported by the *HST* WFC3/IR imaging of four of our galaxies). From our final robust sample, we determine the form of the bright end of the rest-frame UV galaxy luminosity function (LF) at $z \simeq 7$ using a $1/V_{\text{max}}$ estimator. Our results now provide strong evidence that the bright end of the $z = 7$ LF does not decline as steeply as predicted by the Schechter function fitted to fainter data. We consider carefully, and exclude the possibility that this result could be a consequence of either gravitational lensing, or significant contamination of our high-redshift galaxy sample by active galactic nuclei (AGN). Rather our results favour a double power-law form for the galaxy LF at high redshift or, perhaps more interestingly, a LF which simply follows the form of the dark-matter halo mass function at bright magnitudes. This suggests that the physical mechanism which inhibits star-formation activity in massive galaxies (i.e. AGN feedback or some other form of “mass quenching”) has yet to impact on the observable galaxy LF at $z \simeq 7$, a conclusion supported by the estimated masses of our brightest galaxies which have only just reached a mass comparable to the critical “quenching mass” of $M_* = 10^{10.2} M_\odot$ derived from studies of the mass function of star-forming galaxies at lower redshifts.

Key words: galaxies: evolution - galaxies: formation - galaxies: high-redshift.

1 INTRODUCTION

The study of galaxies at high redshift is crucial for understanding the early and subsequent stages of galaxy evolu-

* E-mail: raab@roe.ac.uk

† Scottish Universities Physics Alliance

tion in the Universe. Within the last decade, the number of galaxies known at $z > 6$ has increased to samples of hundreds, led by observations taken with the Wide Field Camera 3 on the *Hubble Space Telescope* (WFC3/*HST*). The key feature of WFC3 that makes it so successful at detecting high-redshift galaxies is the unrivalled sensitivity in the near-infrared, which allows the detection of Lyman-break galaxies (LBGs) at $z > 6.5$ by the redshifted rest-frame UV light from these star-forming galaxies. The accurate selection of LBGs at high redshifts relies on a measurement of the strong spectral break at the wavelength of the Lyman- α line (1216Å), produced by absorption from the integrated neutral Hydrogen along the line of sight. The resulting spectral energy distribution (SED) can then be identified in multiwavelength imaging as an “optical-dropout” galaxy with either colour-colour selection or a SED fitting analysis.

The deepest near-infrared image ever taken in the *Hubble* Ultra Deep Field (HUDF), now reaches depths of $m_{AB} \simeq 30$ over 4.5 arcmin² (UDF12 observing programme; Koekoemoer et al. 2013). The imaging confirms that the low-luminosity, early galaxies uncovered in the HUDF are compact (half-light radius, $r_{1/2} < 0.5$ kpc; Ono et al. 2013) and have similar colours to local star-forming galaxies (rest-frame UV slope $\beta \simeq -2$, where $F_\lambda \propto \lambda^\beta$; Dunlop et al. 2013). The detection of an increasing number of galaxies at $z = 6-8$, including the first sample of galaxies at $z > 9$ (Ellis et al. 2013; Oesch et al. 2013), within the UDF12 and other surveys (Zheng et al. 2012; Coe et al. 2013) has allowed the determination of the rest-frame UV luminosity function at high redshift. The LF characterises the number density of galaxies per comoving volume element as a function of luminosity, and hence is an important measurement for charting the evolution of galaxies (e.g. Bouwens et al. 2011). A commonly used parameterisation of the LF, which well describes the number densities of galaxies at low redshift, is the Schechter function, where $\phi(L) = \phi^*(L/L^*)^\alpha e^{-L/L^*}$. The Schechter function form exhibits a power-law slope to faint luminosities described by the index α , and an exponential cut-off at luminosities exceeding the characteristic luminosity L^* . The extremely faint galaxies ($M_{UV} > -17$) detected within the ultra-deep imaging of the UDF12 programme have constrained the faint-end slope of the $z \simeq 7$ LF to be very steep with $\alpha = -1.9$ at $z = 7$ (McLure et al. 2013; Schenker et al. 2013). The slope of the faint-end of the LF is key for ascertaining the role LBGs play in reionizing the Universe, as only by extrapolating the number densities of galaxies beyond the faintest galaxy detected even in the HUDF, can the ionising photon budget be met by early galaxies (Robertson et al. 2013).

However, the form of the LF, and the characteristic break luminosity, cannot be accurately constrained using galaxy samples from the HUDF alone; samples of intrinsically-rarer bright galaxies are required. There are several key programmes from *HST* that have detected significantly brighter galaxies at $z \simeq 7$ around the apparent break luminosity; the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS, Grogin et al. 2011; Koekoemoer et al. 2011) is extremely powerful at detecting galaxies analogously to those in the HUDF, the Cluster Lensing and Supernova Survey with Hubble (CLASH, Postman et al. 2012; Zheng et al. 2012; Coe et al. 2013) uses foreground clusters to detect gravitationally lensed galaxies, and the

Brightest of the Reionizing Galaxies (BoRG, Trenti et al. 2011; Bradley et al. 2012) programme has allowed the detection of bright galaxies specifically at $z = 8$. Despite the clear success of these *HST* programmes at selecting galaxies at $z > 6.5$, the brightest $z \simeq 7$ galaxies detected in the CANDELS imaging to date are only slightly brighter than the characteristic luminosity ($L \sim 3L^*$, McLure et al. 2013). To detect the very brightest galaxies, degree-scale surveys are needed and these can only be completed efficiently from the ground with wide field-of-view near-infrared cameras.

Theoretically, the form and redshift dependence of the LF contains important information on the key physical processes that govern early galaxy formation and subsequent evolution. To match the observed luminosity- and mass-functions at $z = 0$, simulations of the build-up of galaxies require that, at some stage, the growth of galaxies in the most massive dark-matter haloes be suppressed by some mechanism, such as feedback from a central AGN. In effect, the challenge is to match the steep exponential decline of the galaxy stellar mass (and luminosity) functions at high mass/luminosity, as parameterised through the Schechter function described above. Interestingly, recent studies of the stellar mass function of *star-forming galaxies* have now shown that the characteristic mass above which this steep decline sets in appears to be essentially independent of redshift out to $z \simeq 3$ (e.g. Ilbert et al. 2013). This has led some authors to infer the presence of a characteristic “quenching mass”, i.e. a stellar mass above which a galaxy is likely to have its star-formation activity strongly suppressed by some physical mechanism. As discussed in Peng et al. (2010), current data indicate that the quenching mass above which galaxies rapidly cease forming stars and leave the “main-sequence” of star-forming galaxies is $M_* \sim 10^{10.2} M_\odot$. If this mass quenching, whatever its physical origin, really does set in at a physical threshold which is independent of redshift, then we might reasonably expect the form of the galaxy UV LF to start to diverge from a simple Schechter function at very early epochs (when very few galaxies will have grown to the relevant stellar mass). The ability to test such key ideas, and potentially better constrain the (still unclear) physical origin of “mass quenching”, provides additional strong motivation for determining the form of the bright end of the galaxy UV LF at the highest redshifts.

Further interest in the detailed properties of $z \simeq 7$ galaxies has been generated by the follow-up *HST* and Atacama Large Millimeter/Submillimeter Array imaging of the spectroscopically confirmed Lyman- α emitter (LAE) at $z = 6.595$ (Ouchi et al. 2013), nicknamed ‘Himiko’, that was first discovered by Ouchi et al. (2009b). When observed at ground-based resolution this galaxy appears as a single bright extended source with $m_{AB} \simeq 25$ (at $\lambda \simeq 1\mu\text{m}$), however *HST*/WFC3 imaging reveals the galaxy to be an apparent triple merger system, where each component has roughly the characteristic luminosity L^* . Although an extreme galaxy, Himiko illustrates the potential for detailed study of rare and bright $z \simeq 7$ galaxies that can only be efficiently detected in ground-based narrow- or broad-band survey fields.

Here we extend the work presented in Bowler et al. (2012) using the second data release (DR2) of UltraVISTA, which provides deeper imaging in the Y , J , H and K_s -filters, in strips covering approximately half of the full 1.5 deg² of

the UltraVISTA DR1 data (McCracken et al. 2012). The improved photometry over 70% of the field searched in Bowler et al. (2012) allows us to check the previous candidates and hence validate our selection methodology. We also incorporate the UKIDSS Ultra Deep Survey (UDS), which has a comparable depth and area of J , H and K -band imaging to the UltraVISTA DR2. By combining the deep near-infrared survey data with the optical and mid-infrared data currently available in the UltraVISTA and UDS fields, we have assembled the widest area (1.65 deg^2) of imaging available for the robust selection of $z \simeq 7$ Lyman-break galaxies.

We begin with a summary of the multiwavelength data utilised here from the UltraVISTA and UDS fields in Section 2, followed by the details of our candidate selection in Section 3. The resulting sample of galaxies is presented in Section 4, with physical properties derived from our SED fitting analysis in Section 5. In Section 6 we investigate the sizes of the galaxies in our sample, including an analysis of *HST* imaging of four galaxies in our sample that lie within the region of the COSMOS field imaged as part of the CANDELS programme. We calculate the binned luminosity function for our sample in Section 7, where we also carefully consider the potential effect of gravitational lensing. We discuss our results and compare them to previous work at $z = 5$ and $z = 6$ in Section 8, which also includes a prediction of the level of contamination of our sample by high-redshift quasars. The astrophysical implications of our results are considered further in Section 9, and our conclusions are summarised in Section 10. All magnitudes quoted are in the AB system (Oke 1974; Oke & Gunn 1983) and we assume a cosmology with $H_0 = 70 \text{ kms}^{-1}\text{Mpc}^{-1}$, $\Omega_m = 0.30$ and $\Omega_\Lambda = 0.70$ throughout.

2 DATA

The two multiwavelength survey fields analysed in this paper contain a wealth of observations from X-ray to radio wavelengths. In the following section we describe the specific datasets utilised here for the selection of $z > 6.5$ galaxies, most importantly the near-infrared data from the UltraVISTA DR2 and UDS DR10. The coverage maps of the different wavelength data are shown in Figs. 1 and 2, and a summary of the available broad- and narrow-band filters utilised to image each field can be found in Table 1. The final area of overlapping multiwavelength imaging over the two fields, excluding large stellar diffraction haloes that were masked, comprised 1.65 deg^2 , with 0.62 deg^2 from the UltraVISTA ‘ultra-deep’ survey, 0.29 deg^2 from the UltraVISTA ‘deep’ component and 0.74 deg^2 in the UDS field.

2.1 The COSMOS/UltraVISTA field

2.1.1 UltraVISTA near-infrared imaging

The analysis presented in this paper relies on the first and second data releases of the ongoing UltraVISTA survey¹, which consists of Y , J , H and K_s imaging with the Visible and Infrared Camera (VIRCAM) on the VISTA telescope within the Cosmological Evolution Survey (COSMOS) field.

¹ http://www.eso.org/sci/observing/phase3/data_releases/

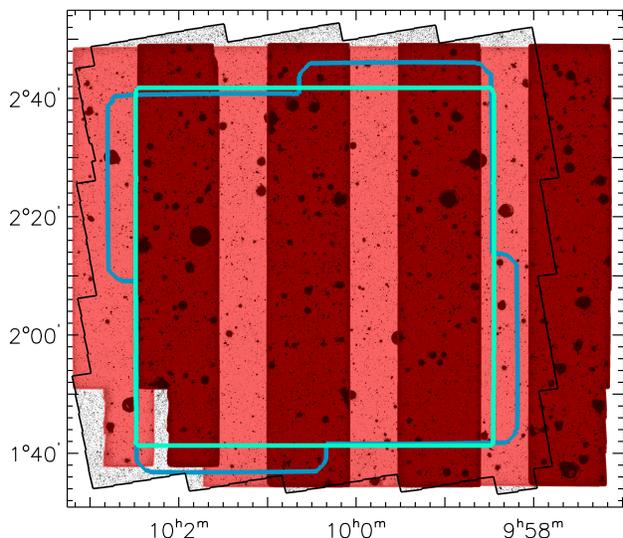


Figure 1. The footprint of the UltraVISTA/COSMOS field showing the key multiwavelength data used here. The large red rectangle shows the year-one 1.5 deg^2 of near-infrared imaging from UltraVISTA DR1, with the deeper strips comprising the second data release shown in dark red. The Subaru z' -band mosaic, formed from four individual Suprime-Cam pointings, is indicated by the blue outline, and the 2 deg^2 of *HST*/ACS I_{814} -band imaging from the COSMOS survey is shown as the large jagged outline. The overlap with the DR2 strips in dark red and the central green square, which is the 1 deg^2 area of a single pointing of CFHT/MegaCam, defines the $\simeq 0.7 \text{ deg}^2$ area utilised in this study.

The first data release, described in detail by McCracken et al. (2012), provided near-infrared imaging over the maximum area of the programme covering 1.5 deg^2 . DR2 provides deeper data in strips that cover $\sim 70\%$ of the the full field as shown in Fig. 1. Integration times for the DR2 Y , J , H and K_s bands range from 29-82 hours per pixel, compared with 11-14 hours per pixel from DR1. Throughout this paper we refer to the DR2 imaging within the strips as the ‘ultra-deep’ part of the survey, and the DR1 imaging over the full field as the ‘deep’ part.

2.1.2 Auxiliary optical and mid-infrared imaging

The auxiliary data used in this paper is described in full by Bowler et al. (2012), but here we briefly describe the key datasets that are shown in Fig. 1. The UltraVISTA survey lies within the multiwavelength imaging taken as part of the COSMOS survey (Scoville et al. 2007b), which covers a total of 2 deg^2 on the sky. Specifically we use optical imaging from the CFHTLS T0006 data release, which defines the maximal area of our search centred on RA $10^{\text{h}}00^{\text{m}}28^{\text{s}}.00$, Dec. $+2^{\circ}12'30''$, with additional deep z' -band data from Subaru Suprime-Cam. Mid-infrared imaging over the COSMOS field by *Spitzer*/IRAC exists from two programmes; the *Spitzer* Extended Deep Survey (SEDS; Ashby et al. 2013) and the *Spitzer* Large Area Survey with Hyper-Suprime-Cam (SPLASH, PI: Capak). The SPLASH data consist of 438 individual exposures in the $3.6\mu\text{m}$ and $4.5\mu\text{m}$ bands, available as calibrated Level-2 files on the *Spitzer* Legacy Archive. We created a mosaic of the SPLASH images by

first background subtracting the frames using SEXTRACTOR with a large background mesh size, before combining the frames using the software package SWARP. The SEDS data was also background subtracted and incorporated into the SPLASH mosaic using SWARP. The photometric and astrometric accuracy was confirmed by comparing to the shallower *Spitzer*/IRAC imaging across the field from the S-COSMOS survey (Sanders et al. 2007). Finally the field is also covered to single-orbit depth in the I_{814} -band by the *HST* Advanced Camera for Surveys (ACS; Koekemoer et al. 2007; Scoville et al. 2007a; Massey et al. 2010).

2.1.3 Data processing and consistency

All images were resampled to the pixel grid of the CFHTLS data (0.186-arcsec/pixel) using the IRAF package SREGISTER, once the astrometric solution had been matched to that of the UltraVISTA Y -band image using the IRAF package CCMAP.

Zeropoints of the full set of multiwavelength imaging were checked by inspecting the colours of flat-spectrum objects defined by the colour bridging the central band. For example, a sample of objects was extracted with aperture corrected flat $z' - J$ colours (e.g. $|z' - J| < 0.05$), and the $z' - Y$ colours of these objects were examined with the expectation that they should also be close to zero. The optical bands were found to have zeropoint offsets of < 0.05 mag, however when comparing the DR1 UltraVISTA data utilised in Bowler et al. (2012) to the optical imaging we found the Y -band magnitudes to be brighter than expected by 0.06 mag. The re-reduction of the full 1.5 deg^2 field encompassing the strips for the second data release, which included improved Y -band calibrations, entirely compensates for this observed colour difference in the DR1. Hence when comparing the depths of the full-field of imaging in Table 1 and the magnitudes of the 10 objects selected in Bowler et al. (2012) in Section 4.3, note that there are small changes in the Y -band photometry as a result of this zeropoint change. The zeropoints of the individual Subaru z' -band tiles were also adjusted slightly ($\Delta m < 0.1$) to be consistent with the single pointing of z -band imaging from CFHT/MegaCam.

2.2 The UKIDSS Ultra Deep Survey field

2.2.1 UKIRT and VISTA near-infrared imaging

In this study, we use the 10th data release (DR10) of near-infrared imaging in the UDS field, from the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007). The UKIDSS UDS consists of deep imaging ($m_{AB} \sim 25, 5\sigma$, see Table 1) in the J, H and K -bands over an area of 0.8 deg^2 centred on RA $02^{\text{h}}17^{\text{m}}48^{\text{s}}$, Dec. $-05^{\circ}05'57''$ (J2000). Data release 10 was made public in January 2013 and is available on the WFCAM science archive².

The UDS field lies within the *XMM-Newton* large-scale structure field, where there exists Y -band imaging from the VISTA Deep Extragalactic Observations survey (VIDEO; Jarvis et al. 2013) as can be seen in Fig. 2. The VIDEO survey is a public survey with VISTA that aims

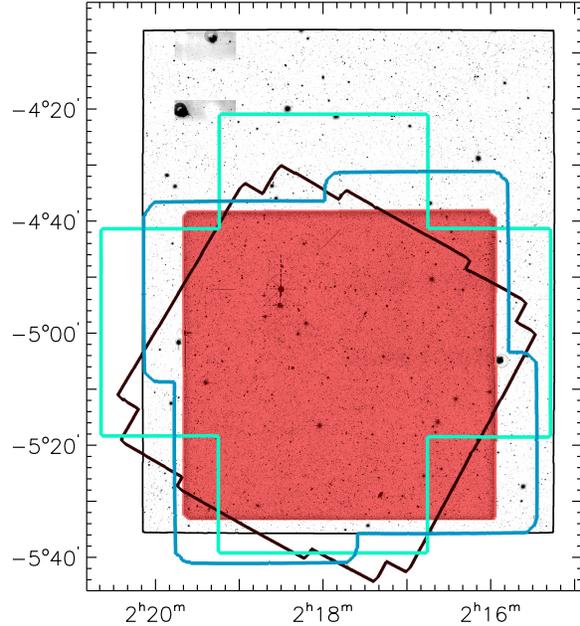


Figure 2. The footprint of the UDS field, showing the UKIRT near-infrared data as the red square sitting within the large rectangle of Y -band imaging from the VISTA VIDEO survey. Data from Subaru Suprime-Cam are shown as the green and blue outlines, where the blue outline defines a z' -band mosaic of four pointings as in Fig. 1. The green cross-shaped outline shows the B, V, R, i and z' -imaging from the original Subaru programme (Furusawa et al. 2008), where the mosaic is formed from five separate pointings. Finally, the footprint of *Spitzer* data from SpUDS is shown as the black diamond. The total area of overlapping Subaru optical and UKIRT near-infrared imaging is $\simeq 0.8 \text{ deg}^2$.

to cover 12 deg^2 in the Z, Y, J, H and K_s -bands over three separate fields. Imaging in the Y -band is key for the robust selection of $z \simeq 7$ galaxies in the UltraVISTA and UDS datasets, as the filter straddles the position of the spectral break and hence can separate genuine high-redshift galaxies from dwarf stars and low-redshift galaxy contaminants that can have identical, red, $z' - J$ colours. Furthermore, the presence of cross-talk artefacts (described further in Section 3.3) in the UKIRT J, H and K imaging makes the presence of a detection in a bluer band independent from the UKIRT data an essential condition for confirming the reality of the high-redshift candidates. When compared to the UltraVISTA Y -band imaging however, the current release of the VISTA VIDEO data is substantially shallower by 1 mag (see Table 1), which reduces the capabilities of the data for selecting $z \sim 7$ sources and makes the selection function for the UDS field different to the UltraVISTA field.

2.2.2 Auxiliary optical and mid-infrared imaging

Optical imaging over the field was provided by the Subaru Suprime-Cam as part of the Subaru/*XMM-Newton* Deep Survey (SXDS; Furusawa et al. 2008). We also obtained ad-

² <http://surveys.roe.ac.uk/wsa/>

ditional z' -band data in four Suprime-Cam pointings, each with 8–15 hrs of integration time (Furusawa et al. in preparation). The astrometry of the individual tiles was matched to that of the UKIRT J -band image using the IRAF package CCMAP. The tiles were background subtracted using SEXTRACTOR (Bertin & Arnouts 1996) and combined into a mosaic with the software SWARP (Bertin et al. 2002), where overlapping regions were combined with the WEIGHTED keyword using weight maps produced by SEXTRACTOR. Finally, a science image on the pixel scale of the binned UKIRT imaging (0.2684-arcsec/pixel) was created from the mosaic (on the native Subaru pixel size of 0.202-arcsec/pixel) using the IRAF package SREGISTER.

In addition to the broad-band filters presented above, we also included data taken with the NB921 filter on the Subaru Suprime-Cam in our analysis (Sobral et al. 2012). The NB921 filter is positioned to the red side of the Subaru Suprime-Cam z' -band filter and hence allows better constraints on the photometric redshift for objects at $6.5 < z < 7.0$. The individual Suprime-Cam tiles were combined into a mosaic covering the full UDS field using the same method as for the z' -band mosaic described above.

The UDS field is covered by *Spitzer*/IRAC data from the *Spitzer* UKIDSS Ultra Deep Survey (SpUDS) programme (PI Dunlop), which has comparable depth to S-COSMOS (see Table 1) with an integration time per pointing of 1200 seconds. The central ~ 600 arcmin² of the UDS field is covered by deeper *Spitzer* imaging from SEDS, and we used a mosaic that included both the SpUDS and SEDS data for our analysis³, taking into account the varying depths across the field using local depths.

2.2.3 Data processing and consistency

All images were mapped onto the astrometric solution and pixel grid of the binned UKIRT J -band imaging, with a pixel scale of 0.2684 arcsec/pixel. Consistency within < 0.05 mag was found between the zeropoints, with the exception being the VISTA VIDEO Y -band image which we found to be ~ 0.1 mag brighter than predicted from the sample of objects with flat $z' - J$ colours. We also observed an identical offset when comparing magnitudes with the Very Large Telescope/HAWK-I Y -band imaging in the field (Fontana et al. in preparation). Hence we shifted the zeropoint of the VIDEO imaging to produce fainter magnitudes by 0.1 mag, with the expectation that future reduction of the data will largely correct this offset with improved calibration, as was the case with the UltraVISTA Y -band data.

2.3 Image depths

The careful determination of imaging depths across all bands is crucial to obtain accurate errors for use in the SED fitting analysis, and also in making appropriate preliminary magnitude cuts in the selection process. However, defining global limiting depths for data over degree-scale fields in the optical and near-infrared becomes problematic, as the combined pointings may have different integration times and seeing. Therefore we obtained local depths over each image

from the clipped median absolute deviation of the 200 closest apertures to each point. Apertures were placed randomly on the blank regions of the images that had been background subtracted using SEXTRACTOR, where blank regions were defined using the segmentation map of each image. The medians of the local depths for the imaging utilised here are presented in Table 1, within the 1.8-arcsec diameter circular aperture used for the photometry in this paper.

2.4 Determination of the enclosed flux

We expect the high-redshift galaxies detected here to be close to unresolved in the available ground-based imaging, which typically has seeing of 0.8 arcsec (see Section 6). However, the variations in seeing throughout the multiwavelength data, along with potential variations across each individual mosaic, result in a different fraction of the enclosed flux in a given fixed circular aperture, which must be corrected for to ensure accurate colours and magnitudes. To extract a point spread function (PSF) for each image, we selected stars using the BzK-diagram (uzK-diagram for the UltraVISTA dataset) as defined in Daddi et al. (2004) and extracted a postage-stamp of each star from a background-subtracted image before centring using the centroid coordinates from SEXTRACTOR. In a range of magnitude bins, these stars were then further background-subtracted and normalised, and a median was taken to form a high signal-to-noise PSF. The curve of growth was then measured on the stack for each magnitude bin (typically from $m_{AB} = 18 - 25$, $\Delta m = 1.0$), and the curves were visually inspected to exclude bins where the stars were saturated and to identify any possible trend with magnitude. Enclosed flux values in a 1.8-arcsec diameter aperture were typically 80% across the optical to near-infrared data, with the exception of the Y and J -band imaging from VISTA VIRCAM which has extended wings (as noted in McCracken et al. 2012) and hence encloses only $\sim 70\%$. For the IRAC imaging, the enclosed flux values quoted in the *Spitzer*/IRAC handbook were used.

3 CANDIDATE SELECTION

3.1 Initial detection and photometry

The primary catalogues for the UltraVISTA and UDS fields were created using SExtractor v2.8.6 (Bertin & Arnouts 1996), run in the dual-image mode to create multiwavelength catalogues. The UltraVISTA catalogue was selected as in Bowler et al. (2012) in a $Y + J$ inverse-variance weighted stack, with additional objects included from Y and J -selected catalogues to ensure we are sensitive to very blue and red spectra up to $z = 7.5$. In the UDS field, where the J -band imaging is substantially deeper than the Y -band, we combined Y and J -band selected catalogues to form a master catalogue rather than creating a $Y + J$ stacked image. Magnitudes were measured in circular apertures of diameter 1.8-arcsec to deliver high signal-to-noise whilst ensuring that the measurements are robust against any astrometric differences between bands (typically $< 0.1''$). We also simultaneously created catalogues with photometry measured in 1.2-arcsec diameter circular apertures, which were used for

³ <http://www.cfa.harvard.edu/SEDS/data.html>

Table 1. The median 5σ limiting magnitudes for the relevant optical and near-infrared data used in this study, obtained from the median of local depths calculated from apertures placed in blank regions of each image (see Section 2.3 for more details). All ground-based magnitudes were calculated within the 1.8-arcsec diameter circular aperture used for photometry here. For the *HST*/ACS I_{814} depth, we used a 0.6-arcsec diameter aperture and the IRAC 3.6 μm and 4.5 μm values were calculated in a 2.8-arcsec diameter aperture. We present the depths of the near-infrared data from UltraVISTA separately for the ‘ultra-deep’ and ‘deep’ parts of the survey. Note that the depths for the ‘deep’ part shown here are ~ 0.4 mags deeper than the results from the original data release presented in Bowler et al. (2012) and McCracken et al. (2012). This is a result of improvements in the photometric calibration of the VISTA data, the smaller apertures used (1.8-arcsec diameter as opposed to 2-arcsec results presented previously) and an improved global depth derived from the local depths.

COSMOS/UltraVISTA			UKIDSS UDS			
Filter	$m_{5\sigma,AB}$ deep	$m_{5\sigma,AB}$ ultra-deep	Source	Filter	$m_{5\sigma,AB}$	Source
u^*	27.2		CFHT/MegaCam	B	27.9	Subaru/Suprime-Cam
g	27.3		CFHT/MegaCam	V	27.6	Subaru/Suprime-Cam
r	27.0		CFHT/MegaCam	R	27.3	Subaru/Suprime-Cam
i	26.7		CFHT/MegaCam	i	27.2	Subaru/Suprime-Cam
I_{814}	26.7		<i>HST</i> /ACS			
z	25.5		CFHT/MegaCam			
z'	26.7		Subaru/Suprime-Cam	z'	26.5	Subaru/Suprime-Cam
Y	25.1	25.8	UltraVISTA	$NB921$	26.1	Subaru/Suprime-Cam
J	24.9	25.3	UltraVISTA	Y	24.8	VISTA VIDEO
$Y + J$	25.3	25.9	UltraVISTA	J	25.7	UKIRT/WFCAM
H	24.6	24.9	UltraVISTA	H	25.1	UKIRT/WFCAM
K_s	24.1	25.0	UltraVISTA	K	25.3	UKIRT/WFCAM
J_{125}		26.8	<i>HST</i> /WFC3			
H_{160}		27.0	<i>HST</i> /WFC3			
3.6 μm	25.3		<i>Spitzer</i> /SPLASH	3.6 μm	24.4, 25.2	<i>Spitzer</i> /SpUDS, SEDS
4.5 μm	25.1		<i>Spitzer</i> /SPLASH	4.5 μm	24.2, 25.0	<i>Spitzer</i> /SpUDS, SEDS

SED fitting of stellar templates with the aim of increasing the signal-to-noise for true point sources.

The *Spitzer*/IRAC measurements were made in 2.8-arcsec diameter circular apertures, using images that had been background subtracted using a large filter size by SEXTRACTOR. The majority of our high-redshift candidates presented in the next section are isolated and hence the large aperture photometry is sufficiently accurate. However, when the photometry is confused, we have excluded the 3.6 μm and 4.5 μm bands from the SED fitting process. We have flagged confused and hence unreliable IRAC magnitudes (which translate into unreliable stellar mass estimates) in Tables 2, 3 and 4.

3.2 UltraVISTA DR2 selection

The initial catalogue for the UltraVISTA dataset consisted of 278916 objects within the $\simeq 0.7$ deg² ‘ultra-deep’ part of UltraVISTA imaging that overlaps with the multiwavelength auxiliary data shown in Fig. 1. We then required that an object be detected at greater than 5σ significance in either the Y or J -bands, where the 5σ limit here was taken as the median local depth from the deepest of the three strips ($Y = 25.8$, $J = 25.4$ mag, 5σ , 1.8-arcsec diameter circular aperture). Using our local depth estimates to compensate for the large diffraction halos around stars in the CFHT/Mega-Cam imaging, we then required the candidate to be undetected in the u^*gr_i -bands at the 2σ -level. The result of the described cuts was a sample of 1188 galaxy candidates, which was further reduced to 589 candidates with the removal of artefacts in the UltraVISTA imaging around

the region of missing data (seen in Fig. 1) and in the haloes of bright stars.

3.3 UKIDSS/UDS selection

The raw catalogue from the UDS dataset consisted of 248191 objects over the full area of J , H and K -band imaging shown in Fig. 2. We then applied the initial criteria that the candidates must be brighter than the 5σ -level in the J -band and undetected at 2σ -significance in the i -band leaving 24797 sources. The available Y -band imaging in the UDS field is of insufficient depth to be competitive with the UltraVISTA Y -band imaging for the selection of galaxies, however it is essential to remove potential ‘cross-talk’ artefacts that occur only within the UKIRT imaging and hence can closely mimic a z -dropout LBG at $z \geq 7$. Cross-talk appears at constant pixel separation from all objects in the UKIRT/WFCAM imaging, which is 128 pixels on the native 0.4-arcsec/pixel scale (51.2-arcsec), and can appear many multiples of this distance away from the source object. Although the brightest occurrences are easily distinguishable from real objects by their ‘doughnut’ appearance, for fainter objects or artefacts a large distance away from the source, it can be very difficult if not impossible to distinguish cross-talk from a high-redshift galaxy. Hence, we apply the condition that candidates must be brighter than the 2σ -level in the VISTA VIDEO Y -band and clearly visible in either the Y -image or the z' -band imaging to ensure a detection in data obtained from independent telescopes. By further insisting that the objects must lie within the region of overlapping Subaru optical and UKIRT near-infrared data (see Fig. 2), and are

further than 100 pixels from the edge of the UKIRT imaging where the signal-to-noise drops significantly, we remove the majority of objects leaving only 202.

3.4 Visual inspection

In the final step before photometric redshift fitting, the candidates were visually inspected in the z' , Y and J -images to remove obvious artefacts such as diffraction spikes, remaining cross-talk in the case of the UDS, and sources within the bright haloes around stars in the VISTA imaging. The i -band images were also inspected, and objects with any detection here (that may have escaped the 2σ -level cut applied above) were removed as indicative of galaxies at $z < 6.5$. The final samples containing the remaining high-redshift candidates for SED fitting consisted of 400 galaxies in the UltraVISTA data and 36 in the UDS field. All near-infrared photometry was corrected to a constant enclosed flux level of 84% (the enclosed flux in a 1.8-arcsec diameter aperture for the UltraVISTA CFTHLS z -band imaging), using the enclosed flux values derived for each image as described in Section 2.4. The 1σ -errors on the photometry were calculated from the nearest 200 blank apertures to each object, using the method described in Section 2.3.

3.5 Photometric redshift analysis

The final step in selecting our sample of $z \sim 7$ galaxies involves fitting the available multiwavelength data points using a photometric redshift fitting routine. By incorporating optical, near- and mid-infrared photometry, we can select good high-redshift galaxy candidates and identify possible low-redshift galaxy interlopers (where the Balmer or 4000Å break is confused with the Lyman-break) or galactic dwarf stars whose spectrum peaks in the near-infrared.

We fitted Bruzual & Charlot (2003) models assuming a Chabrier (2003) initial-mass function, using the Le Phare photometric redshift code (Arnouts et al. 1999; Ilbert et al. 2006)⁴. At each redshift, stellar populations were constrained to be older than 10 Myr and younger than the age of the Universe. We fitted models with an exponentially-decreasing star-formation history with a characteristic timescale, $50 \text{ Myr} \leq \tau \leq 10 \text{ Gyr}$, for two metallicities ($Z = 1/5 Z_{\odot}$ and Z_{\odot}). Note that a constant star-formation and burst model can be closely reproduced by the longest and shortest age τ models respectively. Internal dust reddening was calculated using the Calzetti et al. (2000) attenuation law, and parameterised by values of the rest-frame V -band attenuation in the range $0.0 \leq A_V \leq 4.0$. Absorption by the intergalactic medium was applied using the prescription of Madau (1995).

The presence of a Lyman- α emission line within the spectrum can significantly alter the photometric redshift derived (up to $\Delta z \sim 0.5$) when fitting to broad-band photometry, and hence can cause genuine galaxies at $z \geq 6.5$ to be excluded from our sample. In addition to the models described above, we also separately fitted templates where Lyman- α emission was added to the full template set, with rest-frame equivalent width in the range $0.0 \leq EW_0 \leq 240 \text{ \AA}$. The

continuum level was estimated from the mean value of the continuum in the wavelength range $\lambda = 1250\text{--}1300 \text{ \AA}$ before the reddening was applied.

Contamination by cool galactic stars can be a significant problem when using ground-based data for high-redshift studies (Dunlop 2013), particularly when there is insufficient wavelength sampling of the SED around the predicted Lyman-break. To ascertain how well our galaxy candidates could be described by stellar templates, we fitted the reference stellar spectra from the SpeX library⁵ with spectral types from M4 to T8. The dwarf spectra do not extend to the wavelengths of the *Spitzer*/IRAC bands and so these were excluded during the fitting processes, although they can be taken into account in the selection via the mid-IR colours (see Bowler et al. 2012).

Using the redshift- χ^2 distributions from our photometric redshift fitting procedure, we required an acceptable fit above $z = 6$ (determined by $\chi^2 < 10$ and 11.3 for the UDS and UltraVISTA respectively, which corresponds to 2σ significance given the degrees of freedom in the fitting). In an effort to remove low-redshift contaminants from the sample we also excluded all objects that had a low-redshift solution ($z < 4.5$) within $\Delta\chi^2 = 4$ of the high-redshift ($z > 6$) solution. In the final stage of selection we performed careful visual checks of a stack of the optical bands blueward of the z' -band to ensure there was no residual optical flux that would imply a lower redshift solution. These further steps resulted in samples of 53 and 18 remaining candidates in the UltraVISTA and UDS datasets respectively.

Armed with a reduced sample of candidates that were acceptable as $z > 6$ objects, we performed further SED fitting including Ly α -emission and stellar templates. We also included photometry measured in the NB921 filter for candidates within the UDS field, as the narrow-band sits half-way through the z' -band filter and, with or without allowing for Ly α emission in the SED, allows a more precise determination of the photometric redshift. At this stage we also fitted to the photometry including the IRAC bands at 3.6 μm and 4.5 μm , to exclude dusty low-redshift solutions which have SEDs that rise rapidly to longer wavelengths and also to obtain a more accurate estimate of the galaxy masses. The results of careful consideration of the SED fits, along with final visual identification and removal of subtle near-infrared artefacts, resulted in the sample of 34 galaxies presented in the next section.

4 CANDIDATE GALAXIES

The observed photometry for the final sample of 34 galaxies in the UltraVISTA and UDS fields is presented in Table 2 and the photometric redshifts and best-fitting parameters, such as the rest-frame equivalent width for Ly α and the dust attenuation, are presented in Table 3. In Figure A1 we present postage-stamp images of each candidate and the best-fitting galaxy and star SEDs. All the following tables, SED fits and postage-stamps show the candidates ordered by best-fitting photometric redshift without Ly α emission included in the fitting, where candidates have been split by

⁴ <http://www.cfht.hawaii.edu/~arnouts/lephare.html>

⁵ <http://pono.ucsd.edu/~adam/browndwarfs/spexprism/>

Table 2. The photometry for the sources in our final UltraVISTA and UDS samples is presented in the upper and lower sections of the table respectively. The magnitudes presented here were based on 1.8-arcsec diameter circular apertures except for the *Spitzer*/IRAC channels where 2.8-arcsec diameter circular apertures were used. The photometry has been corrected to the 84% enclosed flux level using the appropriate point-source correction. Where the flux is below 2σ significance, as defined by the local depth derived for each object, we replaced the magnitude with the limiting 2σ magnitude. The errors displayed were derived from the median absolute deviation of the fluxes from the closest 200 blank sky apertures. The presence of a dagger symbol in the right-hand column indicates that the [3.6 μ m] and [4.5 μ m] photometry is confused, and the number corresponds to the Bowler et al. (2012) galaxies with the order consistent with table 2 in Bowler et al. (2012). The ‘Himiko’ galaxy from Ouchi et al. (2013) referenced in the text is marked with an ‘H’.

ID	RA	DEC	z'	NB921	Y	J	H	K	3.6 μ m	4.5 μ m	B12
136380	09:59:15.89	+02:07:32.0	> 27.9	-	25.5 $^{+0.2}_{-0.1}$	25.3 $^{+0.3}_{-0.2}$	25.1 $^{+0.3}_{-0.2}$	25.6 $^{+0.4}_{-0.3}$	> 25.5	> 25.3	
28495	10:00:28.13	+01:47:54.4	> 28.1	-	25.4 $^{+0.2}_{-0.2}$	25.1 $^{+0.2}_{-0.2}$	25.0 $^{+0.3}_{-0.2}$	25.2 $^{+0.3}_{-0.2}$	24.8 $^{+0.3}_{-0.3}$	24.2 $^{+0.2}_{-0.2}$	
268511	10:00:02.35	+02:35:52.4	> 27.6	-	25.0 $^{+0.3}_{-0.2}$	25.0 $^{+0.5}_{-0.3}$	25.4 $^{+0.6}_{-0.4}$	> 25.0	> 25.6	> 25.2	6
268037	09:59:20.69	+02:31:12.4	> 27.6	-	25.1 $^{+0.2}_{-0.2}$	25.5 $^{+0.4}_{-0.3}$	> 25.6	24.9 $^{+0.2}_{-0.2}$	> 25.5	> 25.2	
65666	10:01:40.69	+01:54:52.5	> 27.6	-	24.9 $^{+0.2}_{-0.2}$	24.7 $^{+0.2}_{-0.2}$	24.6 $^{+0.3}_{-0.2}$	24.8 $^{+0.3}_{-0.3}$	> 25.1	24.6 $^{+0.3}_{-0.2}$	4
211127	10:00:23.77	+02:20:37.0	> 27.6	-	25.1 $^{+0.2}_{-0.2}$	25.3 $^{+0.3}_{-0.2}$	> 25.8	25.3 $^{+0.4}_{-0.3}$	24.6 $^{+0.2}_{-0.2}$	23.9 $^{+0.2}_{-0.2}$	†
137559	10:02:02.55	+02:07:42.0	> 27.5	-	25.4 $^{+0.2}_{-0.1}$	25.6 $^{+0.4}_{-0.3}$	25.5 $^{+0.6}_{-0.4}$	26.1 $^{+0.7}_{-0.4}$	> 25.3	> 25.1	
282894	10:00:30.49	+02:33:46.3	> 27.7	-	25.5 $^{+0.3}_{-0.2}$	25.8 $^{+0.6}_{-0.4}$	> 25.9	25.5 $^{+0.6}_{-0.4}$	25.0 $^{+0.4}_{-0.3}$	24.7 $^{+0.3}_{-0.2}$	
238225	10:01:52.31	+02:25:42.3	> 27.6	-	25.0 $^{+0.2}_{-0.2}$	25.0 $^{+0.2}_{-0.2}$	25.0 $^{+0.3}_{-0.2}$	24.9 $^{+0.3}_{-0.2}$	24.7 $^{+0.3}_{-0.2}$	> 25.2	3
305036	10:00:46.79	+02:35:52.9	> 27.7	-	25.3 $^{+0.1}_{-0.1}$	25.2 $^{+0.2}_{-0.2}$	24.9 $^{+0.3}_{-0.3}$	25.1 $^{+0.4}_{-0.3}$	23.9 $^{+0.2}_{-0.2}$	23.9 $^{+0.2}_{-0.2}$	†
35327	10:01:46.18	+01:49:07.7	27.5 $^{+0.5}_{-0.4}$	-	25.3 $^{+0.2}_{-0.2}$	25.8 $^{+0.5}_{-0.3}$	25.7 $^{+0.7}_{-0.4}$	> 25.5	> 25.3	> 25.3	
304416	10:00:43.37	+02:37:51.6	26.6 $^{+0.3}_{-0.2}$	-	24.3 $^{+0.1}_{-0.1}$	24.2 $^{+0.1}_{-0.1}$	24.1 $^{+0.1}_{-0.1}$	24.2 $^{+0.1}_{-0.1}$	23.4 $^{+0.2}_{-0.2}$	23.4 $^{+0.2}_{-0.2}$	1†
185070	10:00:30.19	+02:15:59.8	27.4 $^{+0.4}_{-0.3}$	-	25.4 $^{+0.2}_{-0.1}$	25.3 $^{+0.2}_{-0.2}$	25.5 $^{+0.6}_{-0.4}$	25.1 $^{+0.2}_{-0.2}$	23.7 $^{+0.2}_{-0.2}$	24.7 $^{+0.3}_{-0.2}$	†
169850	10:02:06.48	+02:13:24.2	26.1 $^{+0.1}_{-0.1}$	-	24.5 $^{+0.1}_{-0.1}$	24.5 $^{+0.1}_{-0.1}$	24.6 $^{+0.2}_{-0.2}$	24.6 $^{+0.2}_{-0.2}$	> 25.1	> 25.1	2
304384	10:01:36.86	+02:37:49.2	26.4 $^{+0.3}_{-0.2}$	-	25.0 $^{+0.4}_{-0.1}$	24.6 $^{+0.2}_{-0.2}$	24.9 $^{+0.4}_{-0.3}$	24.8 $^{+0.4}_{-0.3}$	24.6 $^{+0.3}_{-0.2}$	24.7 $^{+0.3}_{-0.2}$	5†
279127	10:01:58.50	+02:33:08.5	26.1 $^{+0.1}_{-0.1}$	-	24.8 $^{+0.1}_{-0.1}$	24.6 $^{+0.2}_{-0.1}$	25.4 $^{+0.3}_{-0.3}$	25.5 $^{+0.5}_{-0.3}$	25.3 $^{+0.6}_{-0.4}$	25.0 $^{+0.4}_{-0.3}$	
170216	10:02:03.82	+02:13:25.1	26.8 $^{+0.3}_{-0.3}$	-	25.5 $^{+0.3}_{-0.2}$	26.0 $^{+0.6}_{-0.4}$	> 25.6	25.4 $^{+0.4}_{-0.3}$	> 25.1	> 25.0	
104600	10:00:42.13	+02:01:57.1	26.3 $^{+0.2}_{-0.1}$	-	25.0 $^{+0.2}_{-0.1}$	25.0 $^{+0.2}_{-0.2}$	24.7 $^{+0.2}_{-0.2}$	25.4 $^{+0.4}_{-0.3}$	23.5 $^{+0.2}_{-0.2}$	23.6 $^{+0.2}_{-0.2}$	†
268576	10:00:23.39	+02:31:14.8	26.8 $^{+0.3}_{-0.2}$	-	25.5 $^{+0.3}_{-0.2}$	25.6 $^{+0.3}_{-0.3}$	> 25.8	25.9 $^{+0.6}_{-0.4}$	24.4 $^{+0.2}_{-0.2}$	> 25.5	
2103	10:01:43.13	+01:42:55.0	26.1 $^{+0.2}_{-0.2}$	-	25.1 $^{+0.2}_{-0.2}$	25.3 $^{+0.4}_{-0.3}$	> 25.3	24.9 $^{+0.4}_{-0.3}$	24.8 $^{+0.4}_{-0.3}$	> 25.4	
179680	09:58:39.76	+02:15:03.3	26.0 $^{+0.2}_{-0.2}$	-	25.0 $^{+0.2}_{-0.2}$	24.9 $^{+0.2}_{-0.2}$	24.8 $^{+0.3}_{-0.2}$	24.6 $^{+0.3}_{-0.2}$	24.0 $^{+0.2}_{-0.2}$	24.5 $^{+0.2}_{-0.2}$	†
18463	09:58:49.36	+01:46:02.1	26.3 $^{+0.2}_{-0.1}$	-	25.4 $^{+0.1}_{-0.1}$	> 26.0	> 25.8	25.9 $^{+0.7}_{-0.4}$	24.8 $^{+0.3}_{-0.2}$	24.7 $^{+0.3}_{-0.2}$	†
122368	10:01:53.46	+02:04:59.9	26.5 $^{+0.3}_{-0.2}$	-	25.5 $^{+0.2}_{-0.2}$	> 26.2	> 25.9	> 25.8	24.9 $^{+0.5}_{-0.3}$	> 25.0	
583226	10:00:46.89	+01:58:46.9	26.4 $^{+0.2}_{-0.2}$	-	25.5 $^{+0.2}_{-0.2}$	25.7 $^{+0.6}_{-0.4}$	25.2 $^{+0.5}_{-0.3}$	25.6 $^{+0.7}_{-0.4}$	24.5 $^{+0.2}_{-0.2}$	> 25.4	
82871	10:01:43.04	+01:58:01.1	25.9 $^{+0.1}_{-0.1}$	-	25.1 $^{+0.1}_{-0.1}$	25.6 $^{+0.3}_{-0.2}$	25.4 $^{+0.3}_{-0.3}$	25.2 $^{+0.4}_{-0.3}$	25.0 $^{+0.4}_{-0.3}$	> 25.2	
68240	09:59:16.85	+01:55:22.1	25.4 $^{+0.1}_{-0.1}$	-	24.7 $^{+0.1}_{-0.1}$	24.6 $^{+0.2}_{-0.1}$	24.9 $^{+0.3}_{-0.2}$	24.7 $^{+0.2}_{-0.2}$	23.7 $^{+0.2}_{-0.2}$	24.0 $^{+0.2}_{-0.2}$	
271028	10:00:45.17	+02:31:40.2	25.9 $^{+0.1}_{-0.1}$	-	25.5 $^{+0.2}_{-0.1}$	25.2 $^{+0.3}_{-0.3}$	25.3 $^{+0.6}_{-0.4}$	25.4 $^{+0.3}_{-0.2}$	24.3 $^{+0.2}_{-0.2}$	24.7 $^{+0.3}_{-0.2}$	†
30425	10:00:58.01	+01:48:15.3	25.2 $^{+0.1}_{-0.1}$	-	24.8 $^{+0.2}_{-0.1}$	24.6 $^{+0.1}_{-0.1}$	25.0 $^{+0.4}_{-0.3}$	25.1 $^{+0.4}_{-0.3}$	23.5 $^{+0.2}_{-0.2}$	24.9 $^{+0.6}_{-0.4}$	9
234429	09:58:36.65	+02:24:56.4	25.7 $^{+0.3}_{-0.2}$	-	25.3 $^{+0.3}_{-0.3}$	> 25.6	> 25.4	> 25.2	> 25.2	> 25.3	
328993	10:01:35.33	+02:38:46.3	25.6 $^{+0.1}_{-0.1}$	-	25.4 $^{+0.5}_{-0.3}$	25.5 $^{+0.5}_{-0.3}$	> 25.4	> 25.4	24.2 $^{+0.2}_{-0.2}$	24.1 $^{+0.2}_{-0.2}$	
35314	02:19:09.49	-05:23:20.6	26.7 $^{+0.2}_{-0.2}$	> 26.8	25.2 $^{+0.5}_{-0.3}$	25.1 $^{+0.2}_{-0.1}$	25.3 $^{+0.3}_{-0.3}$	25.5 $^{+0.2}_{-0.2}$	25.3 $^{+0.5}_{-0.4}$	> 24.7	
118717	02:18:11.50	-05:00:59.4	26.8 $^{+0.6}_{-0.4}$	26.5 $^{+0.4}_{-0.3}$	25.0 $^{+0.5}_{-0.3}$	25.3 $^{+0.2}_{-0.2}$	25.3 $^{+0.3}_{-0.3}$	25.0 $^{+0.2}_{-0.2}$	23.8 $^{+0.2}_{-0.1}$	23.6 $^{+0.2}_{-0.2}$	†
88759	02:17:57.58	-05:08:44.8	25.8 $^{+0.1}_{-0.1}$	24.0 $^{+0.1}_{-0.1}$	25.5 $^{+0.7}_{-0.4}$	25.1 $^{+0.1}_{-0.1}$	25.5 $^{+0.4}_{-0.3}$	24.9 $^{+0.2}_{-0.1}$	23.9 $^{+0.2}_{-0.2}$	24.7 $^{+0.7}_{-0.4}$	H
87995	02:18:50.86	-05:08:57.8	26.3 $^{+0.2}_{-0.1}$	25.6 $^{+0.1}_{-0.1}$	25.0 $^{+0.5}_{-0.3}$	25.3 $^{+0.2}_{-0.1}$	25.3 $^{+0.3}_{-0.2}$	25.3 $^{+0.3}_{-0.2}$	24.9 $^{+0.5}_{-0.3}$	24.8 $^{+0.6}_{-0.4}$	†

field. With the goal of including all potential $z > 6.5$ galaxies, in the following tables we have included galaxy candidates that have best fitting photometric redshifts in the range $6.5 < z < 7.5$ only with the inclusion of Ly α emission in the SED, in addition to the galaxies that do not require Ly α to be robustly at $z > 6.5$. The presence of the spectroscopically confirmed galaxy ‘Himiko’ at $z = 6.595$ in the sample motivates this inclusive approach, as without the inclusion of Ly α in the SED, the galaxy would have been excluded from the sample on the basis of a best fitting photometric redshift of $z = 6.38^{+0.03}_{-0.05}$.

4.1 UltraVISTA DR2

From the 0.62 deg² of the ‘ultra-deep’ UltraVISTA DR2 data, we found 29 candidate $6.5 < z < 7.5$ galaxies. These candidates are listed in the upper part of Tables 2 and 3 and include seven of the eight candidates presented in Bowler et al. (2012) that lie within the regions of the image covered by the ultra-deep data. The table also includes a single object from Bowler et al. (2012) that lies outside the new ultra-deep UltraVISTA imaging we analyse here, but is confirmed as a high-redshift galaxy with the improved photometry now available (see further discussion in Section 4.3). Of the 29 candidates within the ultra-deep strips, 11 have best-fitting photometric redshifts at $z > 6.5$ only when Ly α -emission is included in the templates, although in some cases the candidate still has a reasonable probability of being at

Table 3. The best-fitting photometric redshift and model parameters for the DR2 UltraVISTA and UDS samples are presented in the upper and lower sections of the table respectively. Where the IRAC photometry is considered confused with a nearby object we exclude the [3.6 μ m] and [4.5 μ m] bands from the fitting procedure, and present the resulting χ^2 value labelled with a dagger to illustrate that there are different degrees of freedom, and hence acceptable χ^2 values, for these objects. We order the galaxies by the best-fitting photometric redshift in column 2. Columns 7-10 of the table show the redshift, χ^2 -value, rest-frame equivalent width and A_V value when we introduce the possibility of Ly α -emission in the fits. The candidate labelled ‘H’ is Himiko, which has an extremely bright NB921 flux that can only be well-fitted with models including Ly α emission, hence the unacceptable χ^2 value for continuum-only fitting seen in column three. The best-fitting stellar template, where we fit spectral types M4–T8, is given in column 11 with the χ^2 in column 12. Where a FWHM value has been measured, corresponding to an object selected by SExtractor in that band, it is displayed in the three columns on the right-hand side. In the final column, we flag the candidates from Bowler et al. (2012) that have been re-selected here, with a number corresponding to the order of the candidates in Bowler et al. (2012).

No Ly α			With Ly α							Star		FWHM			
ID	z	χ^2	A_V /mag	Z / Z_\odot	z	χ^2	EW $_0$ /Å	A_V /mag	Z / Z_\odot	Stellar Type	χ^2	z' /arcsec	Y	J	B12
136380	7.21 ^{+0.10} _{-0.21}	1.7	0.0	1.0	7.24	1.6	10	0.0	1.0	T3	17.2	-	1.0	1.3	
28495	7.19 ^{+0.10} _{-0.14}	3.5	0.1	1.0	7.69	1.7	180	0.0	1.0	T1	23.2	-	2.1	2.2	
268511	7.12 ^{+0.14} _{-0.11}	2.6	0.0	0.2	7.25	2.2	80	0.0	0.2	T8	9.2	1.7	1.4	0.5	6
268037	7.07 ^{+0.14} _{-0.12}	9.5	0.2	1.0	7.55	8.6	200	0.3	1.0	T8	22.5	-	2.0	0.7	
65666	7.04 ^{+0.16} _{-0.11}	5.5	0.4	1.0	7.04	5.5	0	0.4	1.0	T4	25.0	1.7	1.1	1.6	4
211127	7.03 ^{+0.12} _{-0.11}	3.2 [†]	0.0	0.2	7.20	2.4	160	0.0	0.2	T8	29.0	-	1.5	1.1	
137559	7.03 ^{+0.14} _{-0.16}	1.9	0.0	0.2	7.15	1.8	60	0.0	0.2	T8	9.8	-	1.7	-	
282894	7.01 ^{+0.14} _{-0.15}	8.5	0.0	0.2	7.31	6.6	240	0.4	0.2	T8	12.5	-	1.6	0.7	
238225	6.98 ^{+0.12} _{-0.12}	3.1	0.5	1.0	7.01	3.0	10	0.5	1.0	T3	23.2	1.0	1.0	1.6	3
305036	6.95 ^{+0.23} _{-0.22}	0.7 [†]	0.0	1.0	7.04	0.7	40	0.0	1.0	T3	11.6	-	-	-	
35327	6.88 ^{+0.10} _{-0.13}	1.5	0.0	0.2	7.05	0.8	100	0.0	0.2	T8	19.0	-	1.5	-	
304416	6.85 ^{+0.09} _{-0.17}	2.4 [†]	0.0	1.0	6.85	2.4	0	0.0	1.0	T3	27.2	-	1.7	1.9	1
185070	6.77 ^{+0.14} _{-0.19}	1.1 [†]	0.1	0.2	7.01	1.1	90	0.2	0.2	T2	32.0	-	1.3	1.3	
169850	6.70 ^{+0.05} _{-0.06}	4.6	0.2	1.0	6.86	4.4	50	0.2	1.0	M6	28.0	1.4	1.7	2.2	2
304384	6.64 ^{+0.12} _{-0.22}	1.9 [†]	0.8	0.2	6.65	1.9	0	0.8	0.2	T3	6.3	1.3	1.1	1.7	5
279127	6.59 ^{+0.05} _{-0.06}	8.8	0.0	0.2	6.59	8.8	0	0.0	0.2	M6	25.9	2.6	2.8	1.9	
170216	6.55 ^{+0.14} _{-0.17}	2.5	0.2	1.0	6.95	1.6	200	0.5	1.0	M6	10.6	1.5	1.2	-	
104600	6.54 ^{+0.07} _{-0.08}	10.1 [†]	0.4	1.0	6.54	10.1	0	0.4	1.0	M6	16.7	1.9	1.5	1.7	
268576	6.51 ^{+0.14} _{-0.12}	9.2	0.0	0.2	6.93	8.6	180	0.0	0.2	M6	13.4	1.4	1.4	1.2	
2103	6.41 ^{+0.11} _{-0.12}	7.0	0.0	1.0	6.90	6.3	240	0.6	1.0	M7	17.3	1.1	2.8	1.3	
179680	6.40 ^{+0.16} _{-0.26}	3.7 [†]	0.4	0.2	6.93	3.0	240	1.0	0.2	M6	20.3	1.6	1.0	0.9	
18463	6.38 ^{+0.07} _{-0.10}	7.3 [†]	0.0	0.2	6.71	6.1	170	0.0	0.2	M7	37.4	1.3	1.6	-	
122368	6.36 ^{+0.14} _{-0.16}	6.4	0.0	0.2	6.94	5.4	240	0.0	0.2	M7	19.5	1.8	1.7	-	
583226	6.33 ^{+0.15} _{-0.16}	5.0	0.0	1.0	6.66	4.7	140	0.0	1.0	M7	10.7	1.5	1.7	1.9	
82871	6.31 ^{+0.09} _{-0.08}	6.1	0.0	1.0	6.81	4.0	240	0.1	1.0	M7	29.3	1.0	1.1	1.0	
68240	6.29 ^{+0.08} _{-0.10}	3.7	0.0	0.2	6.71	3.1	200	0.0	0.2	M7	20.6	1.8	2.0	2.3	
271028	6.21 ^{+0.10} _{-0.19}	0.5 [†]	0.7	1.0	6.50	0.3	110	0.8	0.2	M7	8.2	1.3	1.8	1.6	
30425	6.20 ^{+0.10} _{-0.08}	15.3	0.0	1.0	6.59	13.6	150	0.0	0.2	M7	31.9	1.3	2.1	1.3	9
234429	6.11 ^{+0.22} _{-0.21}	4.3	0.1	0.2	6.64	3.1	240	0.0	1.0	M7	16.1	1.1	0.0	-	
328993	6.01 ^{+0.19} _{-0.21}	0.9	0.5	0.2	6.58	0.1	240	0.1	1.0	M7	21.7	0.9	-	0.7	
35314	6.69 ^{+0.10} _{-0.07}	6.9	0.2	0.2	6.90	6.6	80	0.3	0.2	M6	18.7	2.0	-	1.1	
118717	6.51 ^{+0.05} _{-0.08}	1.0 [†]	0.0	0.2	6.61	1.0	30	0.0	0.2	M9	6.8	0.9	-	1.4	
88759	6.38 ^{+0.03} _{-0.05}	55.8	0.0	1.0	6.52	12.7	150	0.1	1.0	M8	109.8	2.2	-	2.1	H
87995	6.48 ^{+0.03} _{-0.23}	1.4 [†]	0.0	1.0	6.59	1.0	60	0.0	0.2	M8	29.5	0.8	-	1.3	

$z > 6.5$ with continuum-only fitting (as illustrated by the error bars presented and the χ^2 vs. redshift insets in the SED figures). The majority of the candidates are detected in the *Spitzer*/IRAC bands; in the cases where the photometry was contaminated by nearby lower-redshift objects we have excluded these bands from the fitting procedure and flagged the object in Tables 2, 3 and 4. When excluding probable galactic stars from the sample based on the stellar template fitting, the size information obtained from the full-width at half-maximum (FWHM) measurement described

in Section 6 was taken into account. Candidates excluded as stars were retained for size comparisons with the final sample as shown in Fig. 4.

4.2 UDS

Within the 0.74 deg² of overlapping optical and near-infrared imaging in the UDS field, we found four candidates for galaxies at $z > 6.5$ as listed in the lower section of Tables 2 and 3. Of the four candidate $z > 6.5$ galaxies, two

have best-fitting models at $z > 6.5$ only when the fitting allows Ly α emission, one of which is the previously identified and spectroscopically confirmed galaxy ‘Himiko’ (Ouchi et al. 2013) at $z = 6.595$. For Himiko, we find a best-fitting photometric redshift of $z = 6.52$ when including the narrow-band NB921 photometry and allowing for Ly α emission in the fits, and a value of $z = 6.38^{+0.03}_{-0.05}$ with continuum-only fits. The slightly lower redshift compared to that found spectroscopically for Himiko ($z = 6.52$ vs. $z = 6.595$), is a result of the exact redshift being very sensitive to the calibration between the z' and the NB921, as the strength of any line emission must be able to reproduce the $z' - \text{NB921}$ colour observed. The effect is compounded by the slightly longer-wavelength peak of the Ly α line compared with the central wavelength of the narrow-band filter, which is optimised to select LAEs with $z = 6.56 \pm 0.05$. We find a best-fitting $EW_0 = 160 \pm 60 \text{ \AA}$ for Himiko, which although large compared to that found spectroscopically ($\sim 80 \text{ \AA}$), is consistent with the original estimate from the broad- and narrow-band photometry made by Ouchi et al. (2009a) of $EW_0 = 100^{+302}_{-43} \text{ \AA}$.

4.3 Bowler et al. (2012) candidates revisited

In Bowler et al. (2012) we presented ten candidate $6.5 < z < 7.5$ galaxies in the first year ‘deep’ component of the UltraVISTA survey, within the full 1 deg^2 of overlapping data shown in Fig. 1. Using the new reductions of the full field that contain the DR2 imaging, we obtained improved photometry and photometric redshifts for all ten galaxies, which we include in the Appendix. Although only eight of the original ten candidates lie within the ultra-deep strips, the imaging available for the remaining two objects includes more exposures in the DR2 release than those utilised in Bowler et al. (2012), and improvements in the calibration and pipeline has resulted in deeper data by up to 0.35 mag in the case of the H -band. The depth improvements in the Y - and J -bands over the full area are moderate however ($< 0.1 \text{ mag}$) and hence we do not repeat the selection process over the full 1 deg^2 here. The improved photometry shown in Table B1 was taken from the $Y + J$ selected catalogue created as part of the selection process described in Section 3.1. When comparing the photometry between DR1 and DR2, the Y -band magnitudes are fainter by $\gtrsim 0.1 \text{ mag}$, a consequence of the zeropoint shifts in the data reduction process in part, with the remaining magnitude offset a likely result of the use of smaller apertures here (1.8-arcsec diameter vs. 2.0-arcsec) on resolved galaxies (see Section 6).

The four candidates in the top ‘robust’ category are all confirmed as high-redshift galaxies by the new imaging, which forms effectively an independent check of the photometry given the difference in integration times between DR1 and DR2. Comparing the best-fitting parameters, all four have lower best-fitting redshifts by $\Delta z \leq 0.1$. Inspection of the magnitudes and SED plots between the DR1 and DR2 data indicates that the shift in best-fitting redshift is a consequence of the drop in Y and J -band magnitudes closer in-line with the z' -band photometry, requiring the Lyman-break to additionally fill the z' -band to reproduce the now bluer $z' - Y$ colour. The stellar fits for these candidates get worse for three of the four ‘robust’ candidates, with the χ^2

value for object ID 218462 only marginally decreasing from $\chi^2 = 23.9$ to $\chi^2 = 23.2$.

Of the four out of six further candidates from Bowler et al. (2012) that lie in the deep strips, two are present in our final sample (277880 and 28400) with consistent photometry and resulting photometric redshift results as presented. The weakest candidate from Bowler et al. (2012), with ID 2233, was not present in the new $Y + J$ selected catalogue. Visual inspection shows 2233 to be faint in the near-infrared imaging, but bright and elongated in the z' -band. There is a tentative detection in the smoothed i -band image suggesting, when coupled with the photometric redshift analysis, that this object is likely at a slightly lower redshift of $6.0 < z < 6.5$. The remaining candidate (271105), which had the lowest χ^2 value for the stellar fits in Bowler et al. (2012), is best fitted as a star, with the characteristic ‘hook’ in the SED clearly seen in Fig. B1.

Only two of the original ten candidates (268511 and 95661) do not lie within the deep strips, however we remeasured the photometry using the improved data reductions of the full field. Candidate 95661 was included in the Bowler et al. (2012) sample as a result of the large errors on the photometric redshift of $z = 6.13^{+0.38}_{-0.27}$, but is now excluded from our new sample as a $z < 6.5$ galaxy with best-fitting photometric redshift of $z = 6.25^{+0.10}_{-0.13}$. Finally, object 268511 remains a good $z > 6.5$ LBG candidate, although as it is effectively only detected in the Y -band there is still the possibility of it being a transient. Hence, we include candidate 268511 in the final sample as our only candidate from the region of the full 1 deg^2 of overlapping multiwavelength imaging not covered by the ultra-deep DR2 imaging.

In conclusion, of the ten candidate $z > 6.5$ galaxies presented previously in Bowler et al. (2012), all four of the ‘robust’ candidates are reconfirmed here. Two of the three ‘insecure/contaminant’ galaxies are still present in the sample, with the third candidate being excluded as a star based on the improved photometry. Only one of the three ‘insecure’ candidates is retained in our final $z \simeq 7$ sample, with the two excluded galaxies now confirmed to lie at a slightly lower redshift in the range $6.0 < z < 6.5$. We present the improved photometry and derived SED properties for all ten candidates in the Appendix, along with the SED fits and postage-stamp images of the three candidates from Bowler et al. (2012) that are not included in the final sample in this paper.

5 GALAXY PROPERTIES

The extensive multiwavelength photometry available within the UltraVISTA and UDS fields allows an investigation of the physical properties of the galaxies, which further exploits the SED fitting analysis utilised in the selection of our sample. Here we present the derived physical properties including stellar masses, star-formation rates (SFRs) and specific star-formation rates (sSFRs). The section ends with an investigation into the rest-frame optical colours of the galaxies including the potential effects of nebular line emission.

Table 4. The physical properties of the galaxies in our $z \simeq 7$ sample derived from the best-fitting SED models. The stellar mass, SFR and sSFR were determined from the best-fitting τ -model with the errors derived from the points with $\Delta\chi^2 = 1$ from the minimum χ^2 value, after minimising over all other variables. A sSFR of 100 Gyr^{-1} corresponds to the maximum attainable in the set of models we use, which always corresponds to an uncertain mass estimate. We also present the SFR_{UV} derived using the Madau, Pozzetti, & Dickinson (1998) prescription, which converts the observed rest-frame UV flux into a SFR directly. The absolute rest-frame UV magnitudes of our galaxies were measured on the best-fitting model (corrected to total magnitudes using a point-source correction) in the rest-frame, when observed through a top-hat filter of width $\Delta\lambda = 100\text{\AA}$ centred on 1500\AA . Note that the absolute magnitudes presented here have not been corrected for gravitational lensing magnification, as discussed in Section 7.3. In the final two columns we present the rest-frame UV slope index β , calculated by fitting a power law to the measured $YJHK$ or JHK photometry. A missing value for β indicates that the object has large uncertainties in one or more bands and hence a secure value could not be obtained.

ID	$\log(M_*/M_\odot)$	SFR $/M_\odot\text{yr}^{-1}$	sSFR $/\text{Gyr}^{-1}$	SFR_{UV} $/M_\odot\text{yr}^{-1}$	M_{1500} $/\text{mag}$	β_{YJHK}	β_{JHK}	B12
136380	$8.4^{+0.6}_{-0.1}$	27^{+22}_{-16}	100^{+0}_{-88}	19	-22.0	-2.4 ± 0.4	-2.4 ± 0.6	
28495	$10.1^{+0.1}_{-0.7}$	23^{+190}_{-15}	2^{+68}_{-1}	21	-22.1	-2.0 ± 0.4	-2.0 ± 0.5	
268511	$8.5^{+0.5}_{-0.1}$	31^{+9}_{-19}	100^{+0}_{-87}	20	-22.1	-3.5 ± 0.9	-3.9 ± 1.2	6
268037	$8.6^{+0.5}_{-0.2}$	43^{+29}_{-31}	100^{+0}_{-92}	18	-21.9	-1.8 ± 0.4	-0.8 ± 0.6	
65666	$9.2^{+0.4}_{-0.3}$	89^{+87}_{-72}	54^{+46}_{-49}	27	-22.4	-2.0 ± 0.3	-2.1 ± 0.5	4
211127	$8.8^{+1.3}_{-0.4}$	17^{+62}_{-11}	28^{+72}_{-27}	18	-21.9	-2.7 ± 0.5	-2.0 ± 0.8	†
137559	$8.4^{+0.9}_{-0.1}$	20^{+28}_{-14}	91^{+9}_{-86}	14	-21.7	-2.8 ± 0.6	-2.8 ± 0.4	
282894	$9.9^{+0.2}_{-0.3}$	8^{+25}_{-5}	$1^{+5}_{-0.5}$	12	-21.5	-2.8 ± 1.0	-	
238225	$9.0^{+0.6}_{-0.2}$	110^{+48}_{-98}	100^{+0}_{-98}	25	-22.3	-1.9 ± 0.4	-1.8 ± 0.5	3
305036	$10.4^{+0.6}_{-1.6}$	10^{+459}_{-8}	$0.4^{+100}_{-0.3}$	20	-22.1	-1.6 ± 0.3	-1.8 ± 0.6	†
35327	$8.3^{+0.8}_{-0.1}$	21^{+21}_{-14}	100^{+0}_{-95}	15	-21.7	-3.5 ± 1.1	-2.9 ± 1.1	
304416	$10.5^{+0.5}_{-1.0}$	26^{+560}_{-16}	$0.9^{+99}_{-0.7}$	47	-23.0	-1.9 ± 0.2	-1.9 ± 0.3	1†
185070	$10.5^{+0.2}_{-1.6}$	10^{+270}_{-9}	$0.3^{+100}_{-0.2}$	17	-21.9	-1.8 ± 0.3	-1.7 ± 0.5	†
169850	$8.9^{+0.2}_{-0.1}$	78^{+16}_{-38}	100^{+0}_{-69}	35	-22.7	-2.2 ± 0.2	-2.2 ± 0.3	2
304384	$9.4^{+1.4}_{-0.4}$	250^{+120}_{-250}	100^{+0}_{-100}	26	-22.3	-1.9 ± 0.4	-2.3 ± 0.6	5†
279127	$9.2^{+0.2}_{-0.6}$	21^{+54}_{-9}	12^{+88}_{-8}	25	-22.3	-2.8 ± 0.4	-3.8 ± 0.8	
170216	$9.0^{+0.7}_{-0.8}$	15^{+86}_{-12}	17^{+83}_{-15}	11	-21.4	-2.0 ± 0.6	-0.8 ± 1.2	
104600	$8.9^{+1.3}_{-0.2}$	74^{+84}_{-67}	100^{+0}_{-100}	20	-22.1	-2.2 ± 0.3	-2.4 ± 0.5	†
268576	$9.6^{+0.3}_{-0.3}$	4^{+15}_{-1}	$1^{+6}_{-0.6}$	12	-21.5	-2.5 ± 0.6	-2.7 ± 1.0	
2103	$9.4^{+0.3}_{-0.7}$	10^{+130}_{-2}	4^{+96}_{-2}	16	-21.8	-2.0 ± 0.6	-1.2 ± 0.9	
179680	$10.5^{+0.6}_{-1.5}$	22^{+610}_{-21}	$0.6^{+99}_{-0.5}$	20	-22.1	-1.5 ± 0.4	-1.5 ± 0.6	†
18463	$8.2^{+1.2}_{-0.1}$	17^{+11}_{-12}	100^{+0}_{-98}	12	-21.5	-	-	†
122368	$9.5^{+0.2}_{-0.7}$	7^{+7}_{-5}	2^{+9}_{-1}	9	-21.2	-4.1 ± 1.2	-2.5 ± 2.0	
583226	$9.8^{+0.2}_{-0.7}$	7^{+160}_{-5}	$1^{+99}_{-0.7}$	12	-21.5	-2.0 ± 0.5	-1.8 ± 0.8	
82871	$9.3^{+0.3}_{-0.7}$	11^{+76}_{-5}	5^{+95}_{-3}	15	-21.7	-2.3 ± 0.4	-1.3 ± 0.7	
68240	$10.0^{+0.3}_{-0.1}$	8^{+43}_{-3}	$0.8^{+4}_{-0.3}$	25	-22.3	-2.2 ± 0.3	-2.3 ± 0.5	
271028	$9.0^{+1.4}_{-0.3}$	97^{+140}_{-94}	100^{+0}_{-100}	12	-21.5	-1.9 ± 0.4	-2.2 ± 0.6	†
30425	$9.9^{+0.2}_{-0.6}$	12^{+229}_{-4}	$2^{+98}_{-0.7}$	22	-22.2	-2.2 ± 0.4	-3.0 ± 0.7	9
234429	$8.9^{+0.6}_{-0.8}$	9^{+62}_{-6}	12^{+88}_{-10}	10	-21.3	-3.4 ± 1.1	-	
328993	$10.3^{+0.1}_{-0.4}$	28^{+76}_{-25}	2^{+10}_{-1}	11	-21.4	-2.0 ± 0.7	-1.9 ± 1.1	
35314	$8.8^{+0.7}_{-0.4}$	35^{+47}_{-27}	62^{+38}_{-59}	20	-22.0	-2.5 ± 0.4	-2.6 ± 0.4	
118717	$10.2^{+0.6}_{-0.9}$	3^{+320}_{-1}	$0.2^{+100}_{-0.1}$	16	-21.8	-1.8 ± 0.4	-1.6 ± 0.4	†
88759	$10.3^{+0.0}_{-0.3}$	17^{+2}_{-8}	$0.8^{+0.2}_{-0.7}$	21	-22.1	-1.8 ± 0.3	-1.9 ± 0.3	
87995	$10.0^{+0.5}_{-1.4}$	13^{+210}_{-10}	1^{+99}_{-1}	16	-21.8	-2.0 ± 0.5	-2.0 ± 0.5	†

5.1 Stellar populations

5.1.1 Stellar masses, SFRs and sSFRs

We present the stellar mass, SFR and sSFR for each object in the full sample in Table 4, derived from the range of τ -models we fitted to the photometry. Around half (19/34) of the galaxies in our sample have masses close to that found for fainter galaxies with $M_* \sim 10^9 M_\odot$ (McLure et al. 2011). However, as would be expected for our more luminous sample, the remaining 15 galaxies have significantly higher masses in the range $9.5 < \log(M_*/M_\odot) < 10.5$. We note however, that masses below $\log(M_*/M_\odot) = 9.5$ become increasingly uncertain, as they often arise from non-detections

in the *Spitzer*/IRAC bands that correspond to the rest-frame optical part of the galaxy SED (and hence better trace the stellar mass of the galaxy). The majority of the galaxies have SFRs in the range $2 < \text{SFR}/M_\odot\text{yr}^{-1} < 40$ from the best-fitting τ -model, which is consistent with the SFR determined directly from the UV luminosity via the Madau et al. (1998) formalism. For the 15 galaxies that have masses at $9.5 < \log(M_*/M_\odot) < 10.5$, we find a mean sSFR of $1.0 \pm 0.1 \text{ Gyr}^{-1}$ where the error quoted is the standard error on the mean. The errors on any individual measurement of the sSFR are large as a result of degeneracies in the SED fitting process, where in general a very young model with significant attenuation cannot be distinguished from an older

model with little dust. Previous studies with less massive galaxies ($M_{\star} \sim 10^9 M_{\odot}$) found higher values of sSFR than we find in our sample. For example, Smit et al. (2013) put a lower limit of sSFR $> 4 \text{ Gyr}^{-1}$ from a small sample of $z = 6.8$ lensed LBGs, which is consistent with the value of sSFR $\simeq 10 \text{ Gyr}^{-1}$ found by Stark et al. (2013). We do not include models with nebular emission lines in our SED fitting analysis, which can result in underestimated masses by a factor of $\simeq 2.4\text{--}4.4$ (Stark et al. 2013), where the larger value assumes a continued evolution in the EW distribution of the contaminating lines beyond $z = 5$. Although correcting the masses of our sample by the factors derived by Stark et al. (2013) would bring our results into line with previous results, the inclusion of nebular emission lines into the SED models not only affects the resulting stellar mass estimate, but can result in younger ages and lower dust attenuation which subsequently affects the SFR estimate (de Barros et al. 2012). However Curtis-Lake et al. (2013) found that sSFRs at $z \simeq 6$ were increased by a factor of at most two when nebular emission lines were included, and therefore it is likely that our sample shows a genuinely lower sSFR than lower-mass galaxies at $z \simeq 7$, with a sSFR $\lesssim 2 \text{ Gyr}^{-1}$. The lower value of sSFR we find for our sample of $M_{\star} \sim 10^{10} M_{\odot}$ galaxies is consistent with a galaxy formation model where star formation is most efficient in smaller galaxies (e.g. Kauffmann et al. 2003).

From the subset of galaxies in our sample that have lower and hence more uncertain masses, there are several objects that have very low masses coupled with high best-fitting SFRs $\simeq 30 - 100 M_{\odot} \text{ yr}^{-1}$ and hence extreme sSFR $\simeq 100 \text{ Gyr}^{-1}$. Closer inspection of the best-fitting SEDs for these candidates show that they tend to have very blue SEDs and non-detections in the *Spitzer*/IRAC bands, which results in the SED fitting procedure fitting the youngest possible age (10 Myr) in an attempt to reproduce the steep spectral slope. While it is plausible that the extreme sSFRs observed could be genuine, resulting from a brief burst of star-formation in the galaxy, insufficient depth in the rest-frame optical region of the spectrum is more likely the cause of the large sSFRs. The near-infrared data available in the UltraVISTA and UDS fields is substantially deeper than the *Spitzer*/IRAC imaging, and the photometric errors on many of these galaxies are sufficiently large to allow a SED fit with a more moderate SFR and mass within the errors, as illustrated by the large uncertainties shown in Table 4.

With improved photometry for the Bowler et al. (2012) sample of galaxies, we find that the derived physical properties of the objects are more extreme in the analysis presented here. The high SFR and sSFR for several of the Bowler et al. (2012) objects are also a result of extremely young best-fitting ages as was the case for fainter galaxies described above, where the close-to-flat near-infrared photometry can typically be best-fitted using a model with a young age, high SFR and a large dust-attenuation. By requiring zero dust attenuation using the Madau, Pozzetti, & Dickinson (1998) prescription we find more moderate values of the SFR, which then produce slightly lower sSFRs $\sim 10 - 60 \text{ Gyr}^{-1}$. In the case of the top candidate in Bowler et al. (2012) which now has a more moderate SFR, further tension in the fitting was introduced in Bowler et al. (2012) by including confused IRAC photometry, which is excluded from the SED fitting procedure here.

5.1.2 Rest-frame UV slope

The rest-frame UV slope β was calculated for the galaxies in our sample by fitting a power law to the *YJHK* or *JHK* photometry following the method of Rogers et al. (2013a). The error on the derived slope is reduced by including the *Y*-band. However, above $z = 6.8$ the *Y*-band is an unreliable continuum measure, due to the Lyman-break moving through the filter and potential contamination by Ly α -emission, and hence we quote the *JHK* value for comparison. The median values of $\beta = -2.1$ and $\beta = -2.0$ including the *YJHK* and *JHK* bands respectively, are consistent with the value found for fainter galaxies at $z \sim 7$ using the same fitting method ($\beta = -2.1 \pm 0.2$; Rogers et al. 2013a; Dunlop et al. 2013) and using the colours directly ($\beta = -2.33 \pm 0.33$; Bouwens et al. 2013). However, there is evidence that a colour-magnitude relation extends to $z = 7$ (Bouwens et al. 2013), and in which case we would expect the relatively luminous galaxies presented here to have redder rest-frame UV slopes on average. The relation derived by Bouwens et al. (2013) would predict an average $\beta = -1.6$ at $M_{\text{UV}} \simeq -22$ and $\beta = -1.4$ at $M_{\text{UV}} \simeq -23$. The β values for the faintest galaxies presented here at $M_{1500} > -22$ have large errors and exhibit a wide scatter, however the two brightest galaxies (304416 and 169850) that were originally selected in Bowler et al. (2012) are detected at $\sim 10\sigma$ in the near-infrared bands and have absolute magnitudes of $M_{1500} \sim -22.7$ (correcting for gravitational lensing as discussed in Section 7.3). Both candidates show bluer β values than predicted by the relation from Bouwens et al. (2013), with $\beta_{YJHK} = -1.9 \pm 0.2$ and $\beta_{YJHK} = -2.2 \pm 0.2$ respectively. However, the most recent results at $z = 5$ suggest that the bluest galaxies have a similar β at all luminosities, with an increased scatter to redder values with increasing luminosity, possibly as a result of greater dust attenuation or age spread (Rogers et al. 2013b). The bluer values of β we find for the two brightest members of our sample are coupled with best-fitting models that show low values of dust extinction ($A_V = 0 - 0.2$), and therefore these galaxies could be extreme examples of the generally redder population of galaxies at $M_{1500} \simeq -22.5$. Our results contrast with Willott et al. (2013), who found a redder rest-frame UV slope of $\beta = -1.44 \pm 0.1$ in a stack of bright $M_{1350} \sim -22$ galaxies at $z = 6$, which was attributed to dust-reddening. Further imaging within the ultra-deep UltraVISTA survey regions over the next few years will allow more accurate β measurements for a larger sample of objects, and hence a meaningful constraint on the very bright-end of the colour-magnitude relation at $z \simeq 7$.

5.2 Nebular emission

There is growing evidence for the presence of nebular emission lines in the spectra of high-redshift galaxies (Stark et al. 2013; Smit et al. 2013). Aside from the physical repercussions for the star-formation environment at high redshift, the contamination of the rest-frame optical photometry by nebular emission lines can increase the stellar mass estimate and hence artificially suppress the derived sSFR (Stark et al. 2013). In contrast to the majority of LBGs known, our candidates are sufficiently bright to be detected in the *Spitzer*/IRAC [3.6 μm] and [4.5 μm] -

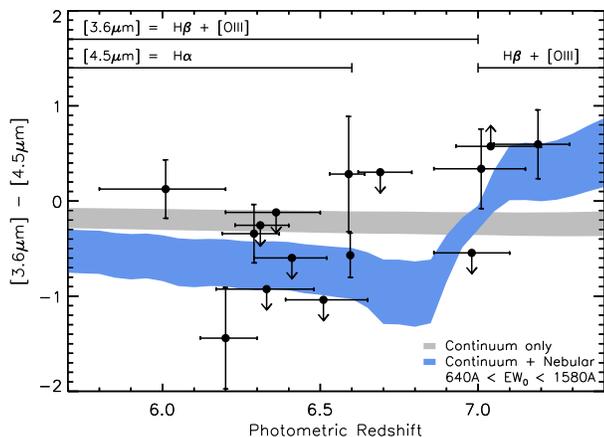


Figure 3. The $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colours plotted against photometric redshift for the UltraVISTA DR2 and UDS candidates. The grey and blue shaded regions represent the predicted $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colours as a function of redshift as described in the text, for Bruzual & Charlot (2003) models with or without strong emission lines added respectively. Candidates with confused IRAC photometry, defined as when flux from a nearby object enters the 2.8-arcsec diameter circular aperture of the candidate, are not plotted here. If an object was not detected at greater than 2σ significance in either band then it has been excluded from the plot. Limit arrows represent objects that were undetected at less than 2σ significance in the appropriate band, and here the magnitude in the undetected band has been replaced by the locally-determined 2σ limiting depth for that object. The dominant emission lines present in each IRAC filter at that particular redshift are shown by the marks at the top of the plot. The data point at $z = 6.6$ with extremely small errors on the redshift is the spectroscopically-confirmed LAE Himiko.

filters, and several galaxies within our sample show an unusual $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colour that cannot be reproduced by continuum fits for high or low-redshift galaxies. The $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colours of dwarf stars can reproduce such colours (see Figure 3 of Bowler et al. 2012), and the stellar locus in $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colour would roughly follow the predicted nebular emission curve. The dependence occurs because cooler T-dwarfs have extremely red spectra ($[3.6\mu\text{m}] - [4.5\mu\text{m}] \sim 0.0-2.0$) and hence become confused with higher redshift candidates, whereas the M and L dwarfs tend to have colours of $[3.6\mu\text{m}] - [4.5\mu\text{m}] \sim -0.5$ and best-fit galaxy redshifts of $z_{\text{phot}} < 6.7$. However, we expect the degree of stellar contamination in our sample to be very low, as a result of our SED fitting analysis in combination with FWHM measurements (see Section 4) that can aid the exclusion of point sources for the candidates with the best (but still significantly worse than the high-redshift models) stellar fits.

For comparison with our observations, we predict the $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colours for continuum only and continuum + nebular emission by adding bright nebular emission lines to stellar continuum models. The underlying continuum was taken from a Bruzual & Charlot (2003) model with a constant SFH, Chabrier IMF, ages of either 10 or 100 Myr and metallicities of either $1/5 Z_{\odot}$ or $1/50 Z_{\odot}$. The model parameters were chosen to produce a range of plausible $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colours, while including the bluest values that could be exhibited by a realistic star-forming

galaxy at high redshift in the absence of nebular emission lines. The expected colours for a given rest-frame equivalent width of H β and [OIII] combined were estimated assuming the H β to [OIII] $\lambda 4959, 5007$ and H β to [OII] $\lambda 3727$ ratios for a metallicity of $1/5 Z_{\odot}$ calculated by Anders & Alvensleben (2003), and assuming the H α to H β ratio of 2.87 from Osterbrock & Ferland (2006). The combined EW_0 of H β to [OIII] $\lambda 4959, 5007$ was chosen to be a minimum of 637 \AA at $z = 6.8$, to correspond to the lower limit for the EW_0 of the mean $z = 7$ galaxy derived by Smit et al. (2013). An upper value for the EW_0 was taken to be 1582 \AA , the value derived from the four bluest galaxies in the Smit et al. (2013) sample. The EW_0 was allowed to evolve with redshift according to $EW(\text{H}\beta + [\text{OIII}]) \propto (1+z)^{1.8}$ which has been derived from lower-redshift results (Fumagalli et al. 2012). Although we did not calculate the relative line strengths using a full recombination analysis, our results closely mimic those presented in Wilkins et al. (2013) where the full analysis was undertaken.

Our observed $[3.6\mu\text{m}] - [4.5\mu\text{m}]$ colours plotted against the best-fitting photometric redshift (without Ly α emission included), along with the predicted colours from our models are shown in Fig. 3. We excluded candidates that had confused IRAC photometry, and those that were undetected at less than 2σ significance in both IRAC filters, resulting in 15 galaxies remaining. Around half of the galaxies with isolated IRAC detections are consistent with continuum-only models, however the remaining galaxies show deviations from the approximately flat colour and follow the predictions for continuum + nebular emission line models. de Barros et al. (2012) found that 65% of $z \sim 3 - 6$ galaxies showed signs of strong nebular emission lines, which would agree with our results that a fraction of our sample show IRAC colours consistent with no nebular contamination, assuming that the galaxies that are detected in the IRAC bands are representative of the population as a whole. Our results are also consistent with Smit et al. (2013), who found evidence for strong nebular emission in the majority of seven lensed galaxies around $z_{\text{phot}} \simeq 6.7$. Future deconvolution analysis of the ‘ultra-deep’ imaging regions and spectroscopic confirmation of the candidates would allow tighter constraints on the prevalence and strength of nebular emission lines in the SEDs of bright $z \sim 7$ galaxies.

6 GALAXY SIZES

Lyman-break galaxies are known to be smaller at high-redshift, with the median half-light radius for faint ($0.1-1L^*$) galaxies being $r_{1/2} < 1$ kpc at $z = 7$ (Oesch et al. 2010; Huang et al. 2013). Confirmation of the existence of a clear size-magnitude relation (as observed at lower redshift) becomes challenging at $z > 6$, because of the restricted luminosity baseline over which sizes can be accurately measured. For example, the few tens of galaxies at $z > 7$ detected and measured in the HUDF by Oesch et al. (2010) and more recently in the UDF12 dataset by Ono et al. (2013) typically have sizes at the limit of the resolution of *HST* (< 0.1 arcsec, which corresponds to a physical size of $d < 0.5$ kpc at $z = 7$). By incorporating the wider-area CANDELS data, Grazian et al. (2012) were able to determine a size-magnitude relation at $z = 7$ extending to $M_{\text{UV}} \simeq -21$. An extrapola-

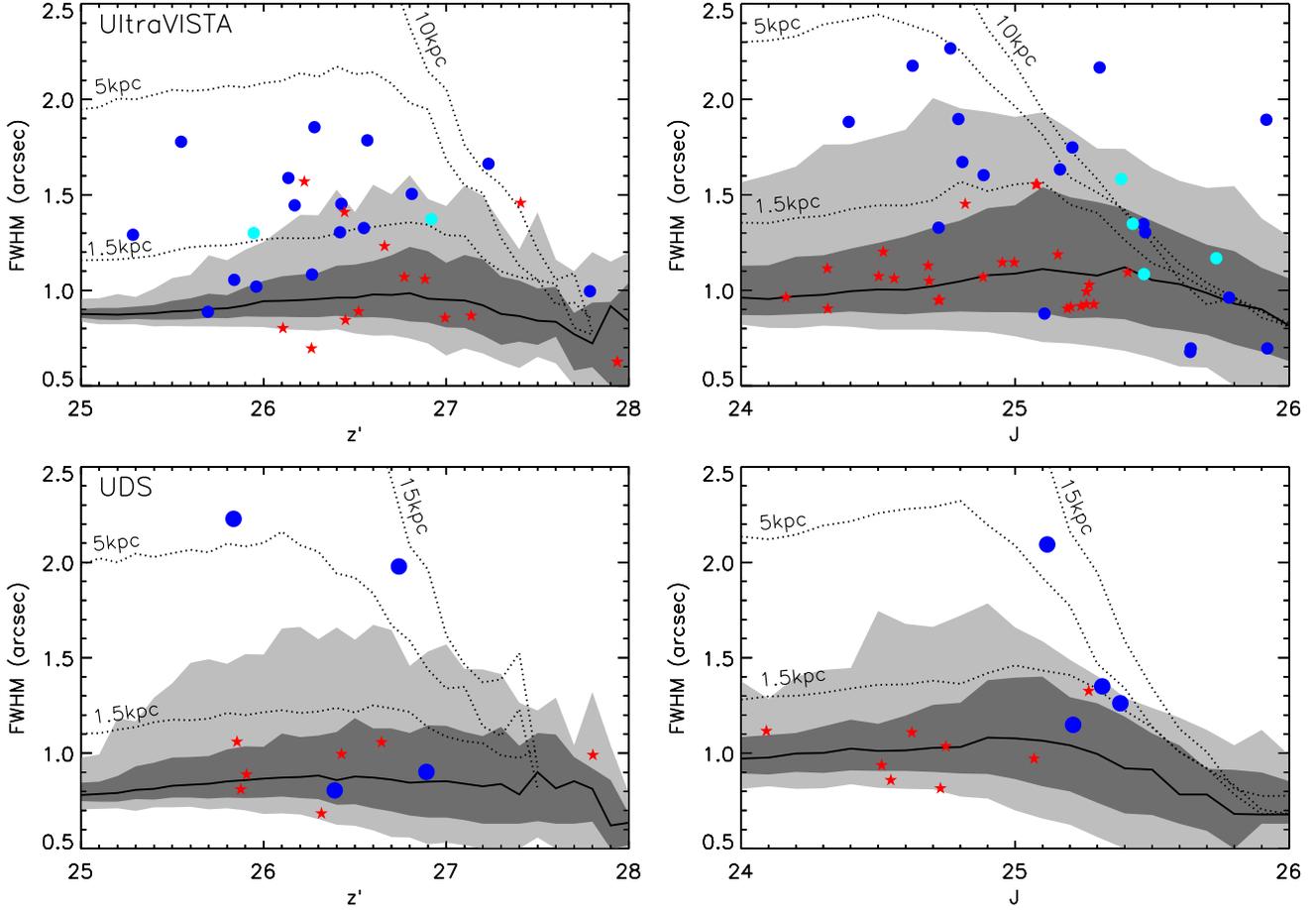


Figure 4. The measured FWHM as a function of magnitude for the UltraVISTA and UDS galaxy samples in the top and bottom row respectively. Each column shows the results for the z' and J -bands, with the galaxy candidates presented as the blue circles. The four galaxies that lie within the COSMOS CANDELS imaging are shown as cyan circles. The magnitudes presented for the real and artificial sources are raw 1.8-arcsec diameter circular aperture values. The red stars show the FWHM of faint galactic dwarf stars, that were originally selected as high-redshift galaxy candidates but later excluded as stars based on the SED fitting analysis. The black curve shows the median value of recovered FWHM as it varies with the magnitude of injected point sources, and the dark and light grey bands show the 68% and 95% confidence intervals respectively. The dotted lines show the median value of recovered FWHM for Sersic profiles with $r_{1/2} = 1.5, 5.0$ and 10.0 or 15.0 kpc, that were convolved with the PSF and injected into the images.

tion of the size-magnitude relations of Oesch et al. (2010) and Grazian et al. (2012) would predict $1.0 \lesssim r_{1/2} \lesssim 1.5$ kpc for the bright galaxies detected here ($-22 \gtrsim M_{UV} \gtrsim -23$ at $z = 7$). Converting the half-light radius into a full-width at half-maximum (FWHM) depends on the functional form of the surface-brightness profile; here we assume a Sersic profile with Sersic index $n = 1.5$ following Oesch et al. (2010).

6.1 FWHM measurements

We measured the FWHM of the galaxies in the final UltraVISTA and UDS samples shown in Table 3 using SEXTRACTOR, where the value is calculated from a Gaussian fitted to the core photometry. Note that fitting a Gaussian to an extended Sersic profile causes an overestimate of the FWHM, and can be sensitive to the deblending procedure in SEXTRACTOR, which can cause nearby objects to be grouped and the resulting FWHM measurement to be larger than for the

central object alone. The FWHM of the UltraVISTA and UDS galaxies in the z' and J -bands are shown in Fig. 4, along with the measured FWHM of simulated stars and galaxy profiles within each field. The Y -band imaging has poorer seeing than the J -band, and is of insufficient depth in the UDS image for FWHM measurements, and hence is not used. To simulate the range of sizes of recovered point spread functions (PSFs) and typical galaxy profiles, we injected and recovered simulated source profiles into the images. The PSF derived using the method described in Section 2.4, and Sersic galaxy profiles with a given $r_{1/2}$, were scaled to an input 1.8-arcsec circular aperture magnitude and injected into blank regions of the z' and J -band images. Simulated galaxy profiles were created with intrinsic $r_{1/2} = 1.5, 5.0, 10.0$ and 15.0 kpc, and were then convolved with the PSF (modelled here as a Moffat function obtained by fitting to the stacked PSF). The injected images were then analysed using SEXTRACTOR, and the FWHM and aperture magnitude measurement was

extracted for each recovered source, thus providing a consistent size measurement for comparison with our sample. In general, the recovered FWHM increases towards fainter magnitudes as a result of noise in the measurement, with the 68% and 95% confidence levels showing asymmetry as a result of the increased probability of detection of a source that has been boosted in magnitude and size. As you would expect, the FWHM measurements become increasingly unreliable to fainter magnitudes, where the median FWHM turns over and starts to decrease as a result of the bias in selection; only sources that sit on a noise spike at such faint magnitudes will be detected and included.

Of the 30 galaxies in our final UltraVISTA sample, approximately two-thirds are consistent with having $r_{1/2} \geq 1.5$ kpc assuming a Sersic profile with $n = 1.5$. The z' -band imaging shows a clearer separation of the measured FWHM of galaxies in our sample and that from PSFs, due to the increased depth and better seeing available compared to the J -band imaging. Several of our brightest galaxies, including the brightest candidates from Bowler et al. (2012), are clearly resolved in the ground-based imaging, showing FWHM values that would suggest intrinsic sizes up to $r_{1/2} \sim 5$ kpc. Although the majority of the UltraVISTA galaxies are also consistent with being unresolved in the ground-based imaging, as shown by the 95% percentile displayed in Fig. 4, the distribution of FWHM values away from the locus of recovered genuine point sources suggests otherwise. Stellar contamination of the sample is strongly ruled out by the SED fitting, and given that the current known size-magnitude relation would predict a FWHM that is only just resolvable from the ground, it is entirely plausible that we would find unresolved galaxies with the available seeing. This point is reinforced in Section 6.2, where we analyse *HST* imaging of four galaxies in the UltraVISTA sample that are consistent with being a point source in the ground-based data (highlighted in cyan on Fig. 4), and find strong evidence that they have extended low surface brightness features.

From the final UDS sample of four galaxies, two appear resolved in the z' -band imaging (see Fig. 4), including the known extended $z = 6.595$ galaxy Himiko, which is the only UDS candidate to appear clearly resolved in the J -band imaging. Himiko has an extent of ~ 9 kpc in the continuum measured from higher resolution *HST* imaging, consistent with the FWHM measured here, and a diffuse Ly α halo greater than 17 kpc across. There are no sources in the UltraVISTA sample that have FWHM suggesting that they exceed 5 kpc in size, supporting the conclusion that Himiko is an extremely rare triple merger system and not a typical $m_{AB} \sim 25$ galaxy at $z = 7$.

Our results agree with Willott et al. (2013), who found that around half of their bright $z \sim 25$ LBGs at $z = 6$ were resolved in the CFHTLS ground-based imaging which had a seeing of 0.85-arcsec. They found a median galaxy FWHM = 1.1-arcsec, which corresponded to an intrinsic FWHM = 0.35-arcsec or a $r_{1/2} = 2$ kpc. Willott et al. (2013) noted that, from high-resolution imaging of several members of their sample with *HST*, many of the galaxies appeared clumpy and extended with features up to 6 kpc from the centre. Similarly, the two brightest z_{850} -drops presented in Ono et al. (2013) are formed of multiple components. These results suggest that mergers or interactions may be prevalent

in bright galaxies at high redshift, although it is possible that these low surface brightness features are also present in fainter galaxies beyond the current detection limits.

6.2 *HST* imaging from CANDELS

Of the 30 candidates within the 1 deg² of UltraVISTA imaging, four lie within the 200 arcmin² of the CANDELS COSMOS field (Grogin et al. 2011) which lies close to the centre of the CFHTLS pointing that defines our maximal survey area, and hence in the left-central strip of the ‘ultra-deep’ UltraVISTA imaging. The CANDELS COSMOS field consists of a mosaic of 4×11 pointings of *HST*/WFC3 providing F125W (J_{125}) and F160W (H_{160}) imaging, and ACS imaging taken with the F606W (V_{606}) and F814W (I_{814}) filters in the parallel observations (Koekemoer et al. 2011). In Fig. 5 we present postage-stamps of the four objects in the J_{125} and H_{160} images, with the corresponding VISTA J and H images for comparison. The measured photometry and FWHM values from SExtractor are shown in Table 5.

Of the four candidates we find in the UDS, the LAE Himiko lies within the 145 arcmin² of CANDELS UDS imaging. The detailed analysis of Himiko from the *HST* imaging was presented by Ouchi et al. (2013), who found the galaxy to consist of three components in an apparent triple-merger system. A triple-merger system is very rare at any redshift, and the UltraVISTA galaxies that lie within the CANDELS COSMOS imaging do not show any evidence of multiple components. The four galaxies in the CANDELS imaging however are fainter than Himiko, with absolute magnitudes in the range $-21.9 \leq M_{1500} \leq -21.5$ compared to $M_{1500} = -22.1$, and the brightest candidates presented here are comparably extended from the FWHM measurements in the ground-based imaging. Our sample includes galaxies brighter or comparable to the luminosity of Himiko, and hence follow-up with *HST* could determine how many of the most luminous $z = 7$ galaxies are in merger systems and would constrain the size-magnitude relation at the very bright end.

In the UltraVISTA imaging the four candidates, highlighted in Fig. 4, show FWHM that are consistent with being unresolved at ground-based resolution, with the exception of 271028 which appears marginally resolved in the z' -band imaging. The typical resolution of the *HST*/WFC3 J_{125} and H_{160} imaging in the CANDELS fields is 0.20 ± 0.01 -arcsec (Koekemoer et al. 2011), and hence of the four candidates, three immediately appear resolved in the WFC3 data on the basis of the FWHM measurements. The brightest object in the WFC3 imaging, 211127, is close to being unresolved with a FWHM = 0.22 that only slightly exceeds the value expected for a point source. The remaining galaxies have larger FWHM in the range 0.3–0.5 arcsec, which is supported by the observed elongation of several galaxies in the images shown in Fig. 5. Note that object 268576 lies near the edge of the image and hence the size and photometry measurements are subject to large uncertainties.

6.2.1 Photometry

When comparing the measured aperture magnitudes shown in Table 5, we find an offset between the photometry measured in the UltraVISTA J and *HST*/WFC3 J_{125} imaging

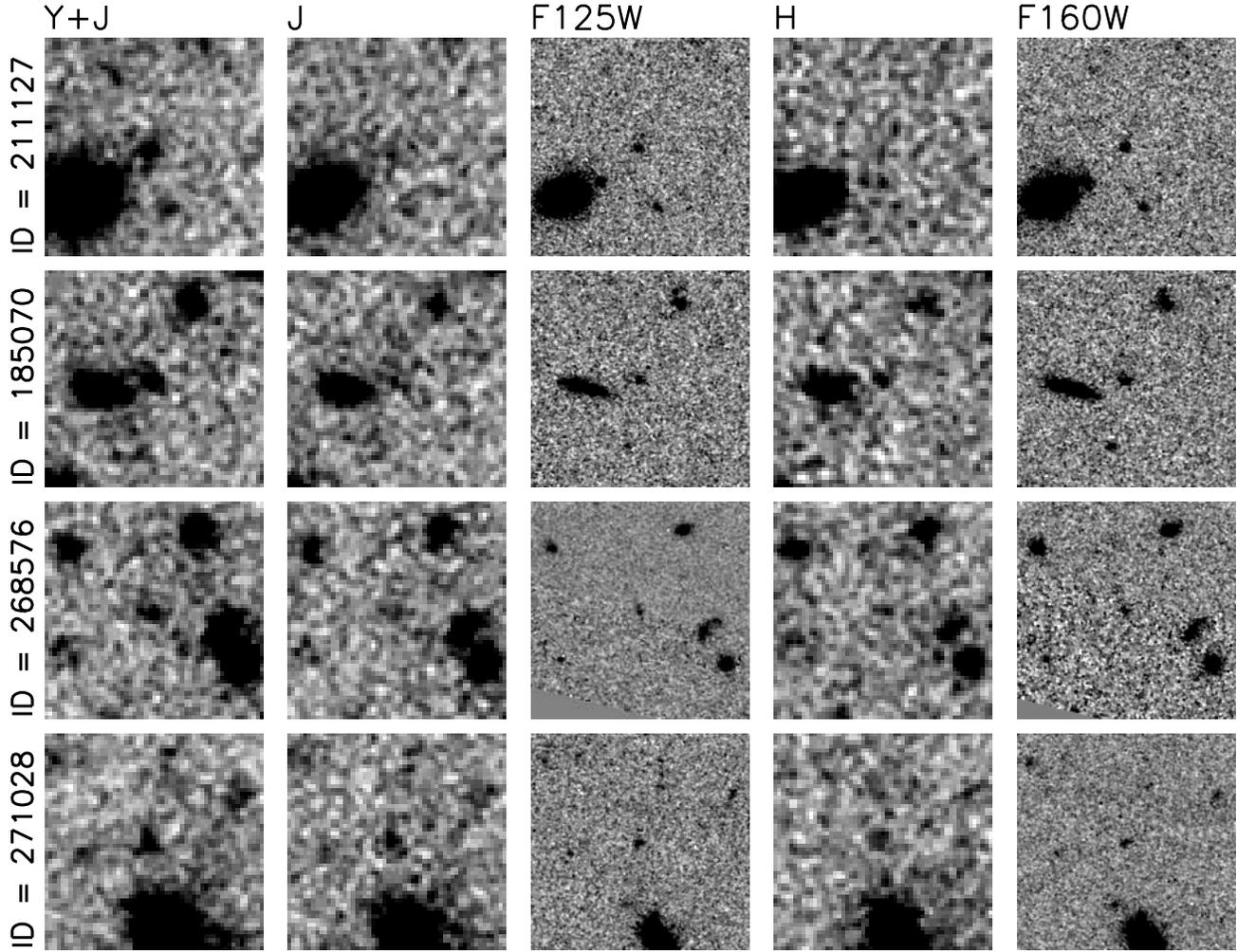


Figure 5. Ground-based and *HST* postage-stamp near-infrared images of the four galaxy candidates that lie with the CANDELS COSMOS imaging. For each object the UltraVISTA *Y + J* detection image is shown on the left, followed by the UltraVISTA *J*-band, WFC3 J_{125} , UltraVISTA *H* and WFC3 H_{160} stamps. Each stamp is 10×10 arcsec², with North to the top and East to the left. Three of the objects presented are fairly close to a low-redshift extended galaxy, however with the improved resolution provided by *HST* the objects are confirmed as clearly separate objects. Note that object 268576 lies near the edge of the WFC3 imaging, and hence the errors on the photometry and size measurements are correspondingly higher.

Table 5. The measured photometry and errors for the four $z \simeq 7$ galaxies in our sample that lie within the area of *HST* imaging provided by the CANDELS COSMOS survey. The measurements from the UltraVISTA *J* and *H* data were made in a 1.8-arcsec diameter circular aperture and corrected to the 80% enclosed flux level. The J_{125} and H_{160} photometry was measured in a 0.6-arcsec diameter circular aperture, which corresponds to 80% enclosed flux for a point source. The object ID 211127 appears barely resolved in the WFC3 imaging based on a measured FWHM close to the typical resolution of 0.20-arcsec, but as shown in Fig. 6, all four galaxies are clearly resolved in the *HST* imaging.

ID	<i>J</i>	J_{125}	<i>H</i>	H_{160}	FWHM <i>J</i> /arcsec	FWHM J_{125} /arcsec	M_{1500}
211127	$25.4^{+0.3}_{-0.2}$	$25.6^{+0.1}_{-0.1}$	> 25.8	$25.6^{+0.1}_{-0.1}$	1.1	0.21	-21.9
185070	$25.3^{+0.2}_{-0.2}$	$25.7^{+0.1}_{-0.1}$	$25.6^{+0.6}_{-0.4}$	$25.6^{+0.1}_{-0.1}$	1.3	0.31	-21.9
268576	$25.6^{+0.3}_{-0.3}$	$25.7^{+0.1}_{-0.1}$	> 25.9	$26.0^{+0.1}_{-0.1}$	1.2	0.48	-21.5
271028	$25.3^{+0.3}_{-0.3}$	$25.8^{+0.1}_{-0.1}$	$25.3^{+0.6}_{-0.4}$	$26.0^{+0.1}_{-0.1}$	1.6	0.37	-21.5

which is inconsistent with the galaxies being point sources. Excluding object 268576 near the edge of the imaging, we find a mean offset of 0.4 mag between the raw 0.6-arcsec diameter aperture photometry measured on the J_{125} image (80% enclosed flux for a point source) and the 1.8-arcsec diameter aperture photometry measured on the J -band image ($\sim 70\%$ enclosed flux) when corrected to match the enclosed flux level of the WFC3 imaging. Of course the most extended galaxies in the higher resolution HST imaging will similarly appear extended in the ground-based imaging, as we have shown the majority of the galaxies in our sample appear resolved in the z' and J -band data. However, because the circular apertures we used for photometry in the space- and ground-based data are designed to enclose around 70-80% of the flux of a point source, in the case of a resolved galaxy, more of the flux will spread beyond the aperture in the higher resolution HST imaging compared to the ground-based imaging where the seeing dominates over the intrinsic FWHM. We created model galaxy profiles to predict the magnitude offset we would expect between the VISTA and WFC3 imaging for an extended profile as compared to a point source. A magnitude offset of 0.4 mag could be reproduced if the simulated galaxy with an exponential profile (Sersic index $n = 1$) had a half-light radius of $r_{1/2} = 1$ kpc. Such a profile would result in 0.05 mag being lost from the ground-based aperture (increased to 0.1 mag for a profile with $r_{1/2} = 1.5$ kpc). The ground-based imaging provides a closer measure of the true total flux of the galaxy than the results of using small apertures on higher resolution HST imaging, when the galaxy is assumed to be unresolved in both cases.

Finally, we measured the photometry for our sample using larger apertures, to empirically determine what the effect of assuming our galaxies are point sources in the ground-based imaging has on the total magnitudes derived. We used 3-arcsec diameter circular apertures (corrected to total assuming a point source) and two further magnitudes from SEXTRACTOR; the MAG_AUTO which returns an estimate of the total magnitude using Kron apertures and MAG_ISOCOR which uses isophotal apertures corrected to total using a Gaussian approximation to the galaxy profile. For the brightest objects, where Kron and isophotal magnitudes correspond to a sufficiently accurate total magnitude, we found an offset of ~ 0.1 mag compared with the total magnitude initially measured in a 1.8-arcsec aperture and then corrected to total, which is in agreement with that found for a simulated Sersic profile with $r_{1/2} = 1$ kpc as described above. Hence, from our ground-based imaging measurements, we expect at most a 0.1 mag brightening of our candidates to correct for missing flux beyond the apertures we use. In the case of the J_{125} and H_{160} imaging however, the offset is significantly larger and can result in an underestimation of the total magnitude of the galaxy by ~ 0.4 mag for photometry measured in a 0.6-arcsec diameter circular aperture.

6.2.2 Surface-brightness profiles

The surface-brightness profiles for the four galaxies in our sample that lie within the CANDELS COSMOS imaging are shown in Fig. 6. Each galaxy was normalised within the central circular aperture of radius 0.1 arcsec, and the nor-

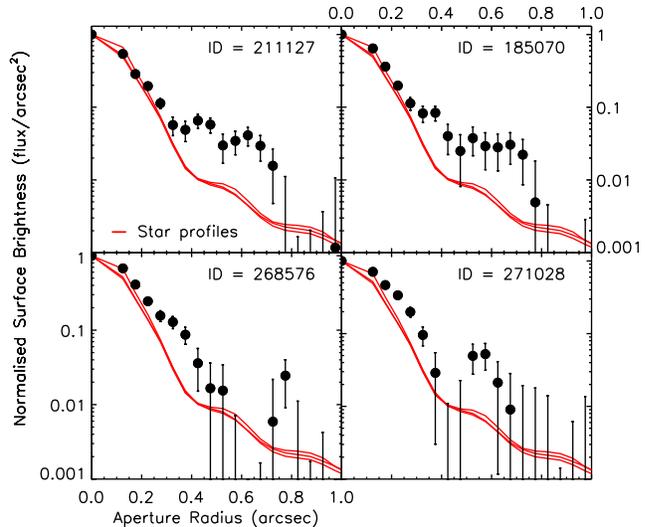


Figure 6. The J_{125} surface-brightness profiles of the four galaxies in the UltraVISTA sample that lie within the CANDELS COSMOS HST imaging. Stellar profiles are shown for comparison (red lines), and all the galaxies and stars have been normalised to a peak flux of 1.0 in a 0.2-arcsec diameter aperture. All four galaxies show extended profiles.

malised surface brightness was calculated in annuli of width 0.05 arcsec. We extracted several unsaturated stars from the imaging and computed the surface brightness for comparison with the galaxy candidates. The profiles show that the first three galaxies show compact cores, surrounded by extended emission that is observable to a maximum radius of 0.8-arcsec. For object 211127, which has a FWHM consistent with being a point source in the J_{125} image, the presence of extended emission further rules out a galactic dwarf star (which is strongly disfavoured from the SED fitting analysis). The compact cores observed for galaxies 211127 and 185070 are consistent with being unresolved in the HST imaging and could suggest a bulge component or active nucleus (further discussed in Section 8.3). The largest galaxy in our sample as measured by both the UltraVISTA J -band and J_{125} FWHM values, 271028, has a core that is significantly larger than the stellar point source surface brightness profile.

7 THE LUMINOSITY FUNCTION

The result of the analysis presented above is a sample of 34 $z \simeq 7$ galaxies with $M_{1500} \leq -21.2$, selected from the 1.65 deg^2 of imaging within the UltraVISTA and UDS fields. In this section we use our sample of galaxies to calculate the very bright end ($M_{1500} \simeq -22.5$) of the luminosity function at $z = 7$. Before we proceed to calculate the binned luminosity function points however, we must calculate how complete our selection process is and hence ascertain the true number density of sources our sample implies. At the bright end of the LF the use of realistic simulations becomes even more important, because the steeply-declining function can cause flux boosting of a significant number of galaxies into the sample from below the flux density limit of the data.

7.1 Completeness simulations

The sample of galaxies presented here will suffer incompleteness due to a combination of blending with foreground/background objects and photometric scattering which can shift galaxies over/below the detection threshold of the survey, as well as potentially changing the colours in such a way as to cause rejection as a contaminant. Hence, to calculate the LF accurately, the numbers of galaxies in each magnitude bin must be corrected for these effects.

We calculated the completeness of our two survey fields by injecting sources, assumed to be unresolved in ground-based imaging (see Section 6), that mimic high-redshift LBGs, and recovering them using the same method used for the selection of galaxies in this paper. We first populated a grid of absolute magnitude vs. redshift space with steps of $\Delta M = 0.1$ and $\Delta z = 0.05$, assuming a linearly evolving Schechter function with redshift as described in McLure et al. (2009) and McLure et al. (2013). To determine the effect of a less steep functional form on the completeness corrections, we also ran simulations where a double power law was used to populate the input $M_{UV} - z$ plane (see Section 7.2). When populating the input $M_{UV} - z$ grid, the LF was integrated down to one magnitude fainter than the median 5σ -limit of each field (e.g. an apparent magnitude limit of $Y = 27$ for the UltraVISTA DR2 strips), to account for scattering into the sample from below the magnitude limit of the survey. For each galaxy in the grid we randomly assigned a rest-frame UV slope β , drawing the value from a Gaussian distribution centred on $\beta = -2.0$ with a standard deviation of $\sigma = 0.2$. We then calculated the observed total magnitude in each band for a galaxy model with that β , M_{UV} , and z combination, taking models from the Bruzual & Charlot (2003) library. To inject the galaxy into each band, the PSF (as determined using the method described in Section 2.4) was scaled to the apparent magnitude in each band and added into the images at a random position. For each survey field, we performed the simulation on 4 to 5 subsections to provide a representative result given the different depths of the individual mosaic panels and strips in the UltraVISTA and UDS imaging, while keeping computing time reasonable. We then ran SExtractor on the full optical and near-infrared images with the artificial sources injected, and selected objects from the resulting catalogues that passed the Y , J , or $Y + J$ cuts imposed on the genuine galaxy catalogues described in Section 3. A non-detection in the i -band was required, using local depths for the UltraVISTA ultra-deep data and global depths for candidates in the UDS to match our selection criterion used for the selection of real sources. Finally, SED fitting was performed and candidates were retained if the χ^2 value was acceptable, and no low-redshift solution had a χ^2 within $\Delta\chi^2 < 4$ from the minimum. In total we added a total of 750000 artificial sources to each field through multiple runs of the simulation, with roughly 10000 – 50000 candidates being retrieved depending on the functional form assumed for the input LF.

7.2 The binned luminosity function

The luminosity function in a given absolute magnitude bin, $\Phi(M)$, derived from our data over the full redshift range of our survey ($6.5 < z < 7.5$) was estimated using the classic

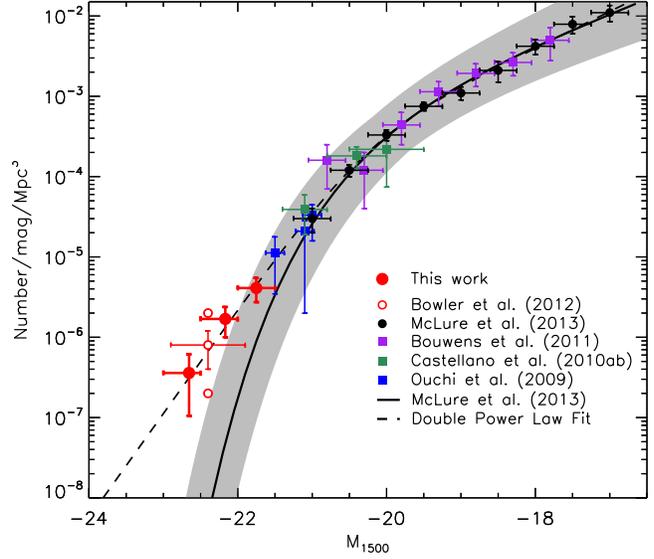


Figure 7. The $z = 7$ UV ($\sim 1500\text{\AA}$) luminosity function. The results from our sample of galaxies from the UltraVISTA DR2 and UDS fields are shown as the red filled circles. The previous estimate calculated in Bowler et al. (2012) is shown as the open red circle, where the upper and lower circle represent the prediction if ten or one of the candidates were confirmed to be at $z > 6.5$. The best-fitting Schechter function at $z = 7$ from McLure et al. (2013) is plotted as the black line, and the best-fitting double power law is shown as the dashed line. By varying the Schechter function parameters (M^* , ϕ^* and α), a one-sigma confidence limit on the best-fitting LF can be obtained, and is shown as the grey shaded region. Data points determined by McLure et al. (2013) and Bouwens et al. (2011) are shown extending to $M_{UV} = -17$. The remaining data points were obtained from wider-area ground-based surveys by Ouchi et al. (2009a) and Castellano et al. (2010a,b).

$1/V_{\max}$ method of Schmidt (1968):

$$\Phi(M) = \sum_{i=1}^N \frac{C(M_i, z_i)}{V_{\max, i}} \quad (1)$$

where the sum is over the N objects in the sample within the magnitude bin. In the simplest form, the luminosity function is found by summing $1/V_{\max}$ for all N galaxies in the bin, where V_{\max} is the maximum volume the galaxy could occupy and still be included in the sample. The V_{\max} was calculated by redshifting the best-fitting SED for each galaxy in the sample until it was fainter than the required Y and/or J -limit for detection in each field (with $z_{\max} \leq 7.5$), and summing the resulting volumes. The incompleteness is taken into account via the correction factor $C(M_i, z_i)$, which depends on the absolute magnitude and redshift of the galaxy.

When calculating the binned LF from our sample, we exclude the 0.29 deg^2 of shallower data within the UltraVISTA field not covered by the ultra-deep strips, as the volume is small compared with that from the UltraVISTA ultra-deep and UDS fields combined. The inclusion of the deep UltraVISTA field in the LF calculation leaves the fainter bins unchanged, as none of these objects could have been detected in the shallower data, and reduces the brightest bin by 18%. To be conservative in our LF esti-

mation, galaxies were only included if the best-fitting photometric redshift without Ly α emission was in the range $6.5 < z < 7.5$. We chose three 0.5 magnitude wide bins centred on $M_{1500} = -22.75$, -22.25 and -21.75 to span the range of absolute magnitudes within our sample. The brighter bin at $M_{1500} = -22.75$ is occupied by two galaxies, the top two candidates presented in Bowler et al. (2012), and the fainter bins contain 6 and 9 galaxies respectively. We note that if all the galaxies listed in Table 3 were included (i.e. also those which require Ly α -emission in the SED to lie at $z > 6.5$) the faintest bins would rise by $\simeq 0.1$ dex.

To calculate the correction required for each galaxy due to the incompleteness of the survey, we compared the results when using an evolving Schechter function and several double power law functions. The exponential cut-off at the bright end of the Schechter function results in a significant number of galaxies being scattered into the sample from below the limit of the survey. The number of these scattered galaxies dominates over the intrinsic number of galaxies in the bin, resulting in a completeness value that exceeds one and a reduction in the final number density derived. However, even correcting the number density down as a result of scattering, we find an excess of galaxies above the Schechter function prediction, implying that the density of galaxies we find cannot be accounted for by such a function. One way to reconcile our results with a Schechter functional form, given that we find an excess of galaxies at $M_{1500} \lesssim -22.0$, would be to assume that the characteristic magnitude, M^* , is brighter than the current determination of the $z = 7$ parameters (McLure et al. 2013; Schenker et al. 2013). However, the result of shifting M^* to a brighter value would bring the fit into conflict with the data points around the knee of the function, in particular the points from Ouchi et al. (2009a) and the brightest bin from McLure et al. (2013). Hence we proceed to calculate the completeness corrections assuming a shallower decline at the bright end of the LF as implied by our data, using a double power law (DPL). Our DPL function, which is the parameterisation commonly used to fit the quasar luminosity function (see Section 8.3), has the following functional form:

$$\phi(M) = \frac{\phi^*}{10^{0.4(\alpha+1)(M-M^*)} + 10^{0.4(\beta+1)(M-M^*)}}. \quad (2)$$

where ϕ^* , M^* and α are the normalisation, the characteristic magnitude and the faint-end slope in common with the Schechter function, and the bright-end slope is described by the power β . We carried out completeness simulations for DPL functions for bright-end slopes in the range $-4.6 \leq \beta \leq -4.0$. The results of the simulations are relatively insensitive to the steepness of the DPL function, with the results changing by less than 4% in the faintest bins and the brightest bin remaining unchanged. The data suggest a more moderate value of $\beta = -4.3$, which we use for our final results presented in Fig. 7 and Table 6.

Finally, if the underlying LF was a Schechter function, the absolute magnitude distribution should be dominated by galaxies at the limiting depth of the survey, however we find a more uniform distribution in the range $-22.5 < M_{UV} < -21.5$. We calculate the probability of obtaining the absolute magnitude distribution we find by drawing a sample at random from the output of our simulations, and determining the number of times the distribution

has the same number or more galaxies in the brighter 0.5 mag section of the $M_{1500} = -22.0$ bin as compared to the fainter section. For a Schechter function we find this distribution in only 3% of cases, whereas for the DPL we find this distribution in 18% of the samples drawn. The detection of two extremely bright $M_{1500} \simeq -22.7$ galaxies further supports our claim of a shallower functional form for the LF.

7.3 Gravitational lensing

Given that the galaxies in our sample are the brightest known to date at $z = 7$, we must consider the possibility that they are gravitationally lensed. There are two scenarios to consider, strong lensing by a lower-redshift galaxy directly along the line-of-sight of our candidate and moderate lensing by galaxies close to the line-of-sight. The first case is ruled out in our sample by the deep optical non-detections of our galaxies. The second case, of galaxies close to the line-of-sight causing moderate magnifications has been discussed before at $z = 5 - 6$ by McLure et al. (2006), who found a moderate magnification of < 0.1 mag for several candidates in their sample, and Willott et al. (2013) who concluded that the brightest candidate in their sample was lensed by ~ 0.25 mag. Following the method detailed in McLure et al. (2006), we calculated the magnification due to gravitational lensing from galaxies within 10-arcsec of each high-redshift galaxy. The velocity dispersion, σ_V , of each potential lensing galaxy was estimated using the Faber-Jackson relation for early-type galaxies from Bernardi et al. (2003), which gives a correlation between the absolute magnitude in the rest-frame i -band and σ_V at $z = 0$. The magnification can then be calculated by assuming the gravitational potential is an isothermal sphere. The absolute magnitude in the i -band was estimated from the nearest band to the redshifted i -band from a K -band selected SEXTRACTOR catalogue, where the total magnitude was estimated using MAG_AUTO. We assumed the redshift of the lensing galaxy from the COSMOS Photometric Redshift Catalogue (Ilbert et al. 2008) for the UltraVISTA sample, and the best-fitting photometric redshift for the UDS sample, where we found the z_{phot} using our photometric redshift fitting routine described in Section 3.5 using the K -band selected catalogue.

All of the 34 candidates within our sample have at least one low-redshift galaxy within 10 arcsec on the sky, and several galaxies have up to seven nearby low-redshift galaxies. The largest magnification comes from the closest galaxies as you would expect, with galaxies around 2.5-arcsec separation typically providing a magnification of 0.1 mag. Our selection procedure naturally removed genuine high-redshift galaxies in the wings of low-redshift galaxies, with potentially high-magnification factors, as the foreground galaxy will contaminate the photometry of the candidate resulting in rejection. If the magnification factors from all of the galaxies within 10 arcsec of the high-redshift galaxy are combined, we find that galaxies in our sample can appear up to 0.3 mag brighter as a result of gravitational lensing. For example, the brightest galaxy in our sample at a measured $M_{1500} = -23.0$, has an estimated lensing boost of 0.3 mag as a result of multiple objects close to the line-of-sight. To present the most conservative estimate of the bright-end of the $z = 7$ LF, we demagnify all of the galaxies in our sample according to our estimate of the magnification, and use

Table 6. The binned LF points at $z = 7$ derived from our combined UltraVISTA and UDS analysis. We use three 0.5 mag wide bins centred on $M_{1500} = -21.75, -22.25, -22.75$ and plot the point at the position of the mean completeness corrected absolute magnitude in that bin, which is displayed in the second column. The errors derived on the number density are the Poisson errors on the number counts.

M_{1500} range /mag	M_{1500} weighted /mag	ϕ /mag $^{-1}$ Mpc $^{-3}$	ϕ_{upper} /mag $^{-1}$ Mpc $^{-3}$	ϕ_{lower} /mag $^{-1}$ Mpc $^{-3}$
$-21.5 < M < -22.0$	-21.75	4.10×10^{-6}	2.73×10^{-6}	2.73×10^{-6}
$-22.0 < M < -22.5$	-22.17	1.69×10^{-6}	1.00×10^{-6}	1.00×10^{-6}
$-22.5 < M < -23.0$	-22.66	3.59×10^{-7}	1.05×10^{-7}	1.05×10^{-7}

the corrected magnitudes in our derivation of the LF points. Furthermore, we present the LF points with a weighted central magnitude for each bin shown in Table 6, calculated from the mean completeness corrected absolute magnitude after demagnification.

8 DISCUSSION

The results of our LF analysis are shown in Fig. 7. Our results at bright magnitudes $M_{1500} < -21.5$ are clearly in tension with the best-fitting Schechter function fit to fainter galaxies, and therefore we also fit a DPL to our derived points and the fainter bins from McLure et al. (2013), with the parameterisation presented in Table 7. Early work presented in Bowler et al. (2012) indicated that there may be an excess of galaxies at the bright end of the $z = 7$ LF, which has been reproduced here with a confirmation of the four brightest galaxies and the majority of the fainter candidates (see the discussion in Section 4.3 for more details). In particular, our two brightest bins contain our most secure candidates (with the two galaxies at $M_{1500} \sim -22.5$ now detected at $> 10\sigma$ significance) and therefore provide a strong challenge to the Schechter LF fitted to the fainter datapoints. The tension between the ground- and space-based observations could be compounded by the effect of missing flux from the use of small apertures and a point-source correction in the *HST* images. As discussed in Section 6, we find an offset of ~ 0.4 mag between the total magnitudes (and therefore absolute magnitude) derived from the UltraVISTA imaging and *HST* CANDELS imaging, when the galaxies are treated as point sources. The size-magnitude relation would suggest that this effect is not a problem for the faintest and therefore smallest galaxies that population the LF, however around the knee of the function at $M_{1500} \simeq -21$, the absolute magnitudes of the galaxies could be underestimated. Other ground-based analyses support a shallower decline at the bright-end of the LF, with the brightest bin of Ouchi et al. (2009a) higher (but still consistent with) the best-fitting Schechter function.

Our derivation of the bright-end of the LF at $z = 7$ is not in conflict with the current sample of galaxies detected in the CANDELS survey, which in total provides ~ 800 arcmin 2 of imaging. Using the DPL fit to our data points at $z = 7$, we would predict ~ 4 galaxies at $M_{1500} < -21.5$ in the CANDELS wide survey. Finkelstein et al. (2013) has suggested that there may be an excess of $z \simeq 7$ high-SFR galaxies in the CANDELS Great Observatories Origins Deep Survey North (GOODS-N), where they found an unusually bright $z = 7.51$ LBG with $H_{160} = 25.6$ ($M_{\text{UV}} = -21.2$). Ono

Table 7. The best-fitting double power law parameters for the fits shown in Fig. 8 at $z = 5, 6$ and 7. The upper part of the table shows the results at $z = 7$ when all the DPL parameters are allowed to vary (shown in Fig. 7). The central section shows the results when fixing the value of $\beta = -4.4$ to the best-fitting value at $z = 5$, and the lower section shows the results with ϕ^* also fixed.

z	ϕ^* /mag $^{-1}$ Mpc $^{-3}$	M^* /mag	α	β
7.0	3.6×10^{-4}	-20.3	-2.1	-4.2
7.0	3.1×10^{-4}	-20.4	-2.2	-4.4
6.0	6.1×10^{-4}	-20.3	-1.9	-4.4
5.0	3.9×10^{-4}	-21.0	-1.9	-4.4
7.0	3.9×10^{-4}	-20.3	-2.1	-4.4
6.0	3.9×10^{-4}	-20.5	-2.1	-4.4
5.0	3.9×10^{-4}	-21.0	-1.9	-4.4

et al. (2012) found one galaxy at $M_{\text{UV}} = -21.8$ that lies within the CANDELS GOODS-N wide field, that was spectroscopically confirmed to be at $z = 7.21$. The LAE Himiko lies within the CANDELS UDS field ($M_{1500} = -22.1$) and of the four new galaxies we find in the CANDELS COSMOS field, two have photometric redshift at $z > 6.5$ and $M_{1500} < -21.5$. The above examples shows that there exist secure detections of $M_{1500} < -21.5$ LBGs at $z \simeq 7$ within the CANDELS fields, with a number in agreement with our predictions, and also illustrates the potential for follow-up observations of the brightest high-redshift galaxies as all the previously known $M_{1500} < -21.5$ examples have been spectroscopically confirmed.

We note here that when finished, the VISTA VIDEO survey (Jarvis et al. 2013) should provide the necessary Z and Y -band depth to search for extremely bright $z = 7$ LBGs over a total area of 12 deg 2 , and therefore determine whether our DPL extrapolation holds to brighter magnitudes.

8.1 Comparison with $z = 5$ and $z = 6$ results

We present the best constraints available on the $z = 5$ and $z = 6$ LF in Fig. 8. In particular, the number density at the bright end has been determined by McLure et al. (2009) who exploited the DR8 release of near-infrared imaging within the UDS field to select a sample of LBGs at $z = 5$ and $z = 6$. For comparison with our results at $z = 7$, we fit DPLs to the binned LF points at $z = 5, 6$ and 7 from McLure et al. (2009, 2013) and this work using a simple χ^2 -minimising

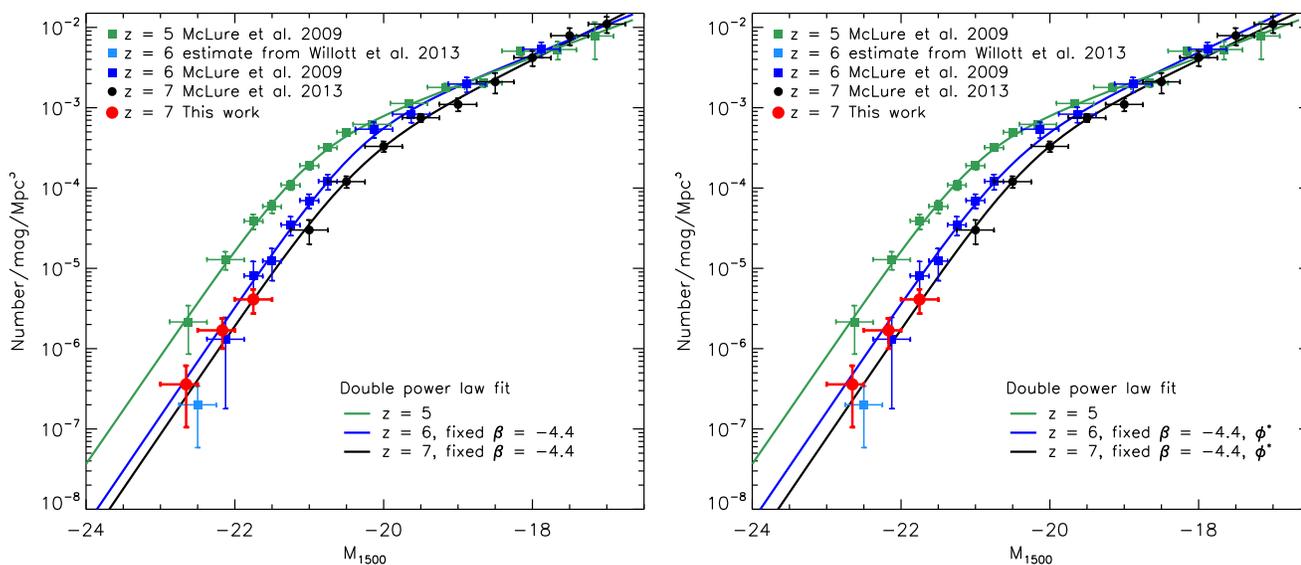


Figure 8. The $z = 5$, $z = 6$ and $z = 7$ UV ($\sim 1500\text{\AA}$) luminosity function points with the best-fitting double power law fits. The filled red circles show our results for bright galaxies in UltraVISTA and the UDS at $z = 7$, and the black points show the results at $z = 7$ from McLure et al. (2013). At $z = 5$ and $z = 6$ we plot the results of McLure et al. (2009) that were obtained from a similar photometric redshift fitting approach to the work presented here. We estimate the number density of $M_{1500} = -22.5$ galaxies from the work of Willott et al. (2013) and plot this point in light blue. The left-hand panel shows DPL fits where the bright-end slope β has been fixed to the best-fitting value at $z = 5$, and the right-hand panel shows the results when both β and ϕ^* are fixed to the value at $z = 5$. The best-fitting parameters in each case are presented in Table 7.

routine and present the results in Table 7 and Fig. 8. At $z = 5$ we allow all the DPL parameters to be varied in the fitting procedure (ϕ^* , M^* , α and β), and for $z = 6$ and $z = 7$ we fix the value of the bright-end slope to the best-fitting value at $z = 5$ ($\beta = -4.4$). Fixing β ensures that we are not overly-fitting to the bright points that are affected by large Poisson errors and cosmic variance. We find that the reduced χ^2 values for the double power law fits are improved compared to a Schechter function, providing additional motivation for the choice of functional form. By alternately fixing ϕ^* and M^* we find that M^* -evolution is preferred in the fitting procedure, supporting previous conclusions in favour of pure ‘luminosity evolution’ at $z = 5 - 7$, and we display the fits with ϕ^* fixed in Fig. 8. The results of McLure et al. (2009) around $M_{1500} \simeq -22$ are not in conflict with our high derived number density, and would support very little evolution in the bright-end of the LF between $z = 6$ and $z = 7$.

Our results however, appear in conflict with the brightest bin at $M_{1350} = -22.5$ from Willott et al. (2013), who used a step-wise maximum-likelihood analysis to conclude that there was an exponential cut-off at the bright-end of the $z = 6$ LF. From a sample of 40 i -dropout LBGs within a total area of 4 deg^2 from the four separate CFHT Legacy Survey fields, Willott et al. (2013) calculated a number density of $2.66^{+5.12}_{-1.75} \times 10^{-8} / \text{mag}/\text{Mpc}^3$ at $M_{1350} = -22.5$, which, if unchanged at $z = 7$, would predict an order of magnitude fewer objects than we find in the UltraVISTA and UDS fields. On closer inspection of the sample obtained from the CFHTLS field however, there exist at least two candidates at $M_{1350} < -22.25$ that should be represented

in the LF derivation, which would appear to contradict the space density calculated by Willott et al. (2013). Willott et al. (2013) note that they are unable to actually measure a volume density as low as $\sim 10^{-8} / \text{mag}/\text{Mpc}^3$ from their dataset, and the anomalously low position of their brightest data-point could be an artefact of their chosen method of fitting the LF. Simply assuming the two brightest galaxies (WHM5 and WHM29) that occupy the $M_{1350} = -22.5$ bin could have been selected in the full survey volume of 10^7 Mpc^3 , we estimate the measured value of the number density from the CFHTLS analysis and present this point in Fig. 8. The brightest bin from this work and Willott et al. (2013) both contain a small number of objects and hence are sensitive to Poisson errors and cosmic variance, however within the errors, our determination of the bright-end of the $z = 7$ LF is not in conflict with that found at $z = 6$ by Willott et al. (2013), assuming our simple binned estimate for the bright bin at $M_{1350} = -22.5$. Finally, the best-fitting DPL we find at $z = 6$ would predict 4 galaxies in the brightest bin of the 4 deg^2 of data utilised by Willott et al. (2013), and hence our results are consistent within Poisson errors.

8.2 Cosmic variance

The derived LF can vary depending on the dimensions of the survey field observed, as a result of the underlying large-scale density fluctuations within the Universe. We find more galaxies within the UltraVISTA ultra-deep data than in the UDS field, which covers a similar area of sky. The interpretation is complicated, however, by the lower selection efficiency of the UDS dataset compared to UltraVISTA, as a

result of the substantially shallower Y -band imaging in the field. Moreover, because we require that a candidate galaxy be detected at $> 2\sigma$ significance in the Y -band, our sample from the UDS field is significantly less complete for galaxies at $z > 6.8$ where the break starts to eat into the Y -band filter. We note that although all of the UltraVISTA candidates would be detected at the 2σ -level in the UDS Y -band, only the two brightest candidates would be detected at the 5σ level and hence likely included in the sample.

Cosmic variance affects the number counts of more massive and hence rarer galaxies more severely, however for small number counts the Poisson error can be more significant. We explore the sources of error on our observed number counts using the Cosmic Variance Calculator v1.02⁶ (Trenti & Stiavelli 2008), assuming a Sheth-Tormen halo mass function, $\sigma_8 = 0.9$ and a halo filling factor of 1.0. From the LF points derived from our sample, we estimated the true number of galaxies present in the UltraVISTA and UDS survey volumes for each 0.5 mag bin. Then, by inputting the completeness and survey dimensions, taking the UltraVISTA field to be a single rectangle with $0.65 \times 1.0 \text{ deg}^2$ for simplicity, we can retrieve the error on the number counts and the relative contribution of Poisson noise and cosmic variance. For the brightest bin centred at $M_{1500} = -22.75$, we find the predicted number counts in each field to be $N = 1 \pm 1$ as you would expect from simple Poisson errors. The cosmic variance for such a small number of objects is dwarfed by the Poisson uncertainty, however it still contributes $\sim 10\%$ of the total error. Similarly, the central bin we calculate at $M = -22.25$ has a predicted number of 4 ± 3 galaxies in each field where 30% of the error is a result of cosmic variance. This prediction is consistent with the 5 galaxies we find in this magnitude range in the UltraVISTA ultra-deep imaging and 1 galaxy in the UDS. The faintest bin we calculate is very incomplete for the UDS field and so we do not compare the number counts between the two fields here. Therefore, we conclude that for our sample, the errors due to Poisson noise dominate over the cosmic variance, which contributes at most 30% of the total error on the number counts.

8.3 Contribution of faint $z = 7$ quasars

Around the peak of the quasar number density at $z = 3$, there is evidence that the very brightest end of the galaxy LF is contaminated by quasars (Bian et al. 2013). There are large uncertainties in the faint-end of the $z > 3$ quasar LF, and the faint-end of the $z = 7$ QLF is completely unknown due to a lack of datasets with the required depth over an adequate area on the sky for detection (see Willott et al. 2010). In Fig. 9, we compare the galaxy LF at $z = 7$ to the known QLF at $z = 5$ and $z = 6$ and the extrapolation of the $z = 7$ QLF beyond the absolute magnitude of the faintest quasar known at $z = 7$ at $M_{1700} = -25.5$ (Venemans et al. 2013). The bright-end of the QLF can be estimated from the four $z > 6.5$ quasars known, which includes one from the UKIDSS-LAS found by Mortlock et al. (2011) and three from the VISTA VIKING survey (Venemans et al. 2013). At $z = 6$, around ten quasars are known (Willott et al. 2010), however again the constraints on the faint-end slope

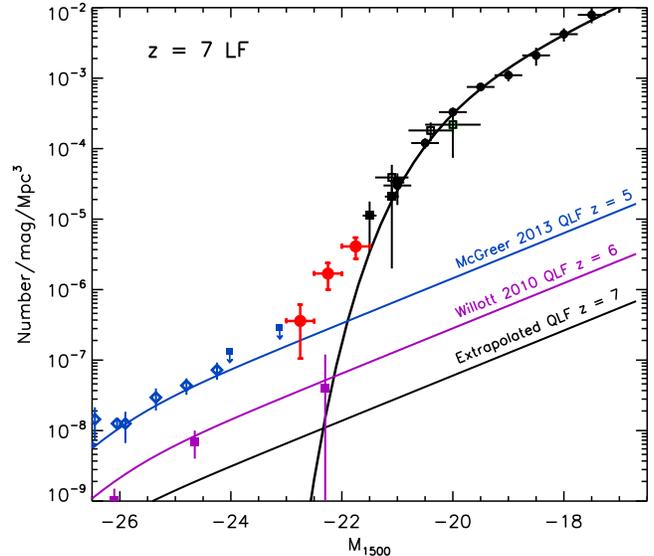


Figure 9. The $z = 7$ galaxy LF from McLure et al. (2013) and the data points from Fig. 7 are shown in black and the data points determined from our analysis are shown in red. The $z = 5$ and $z = 6$ quasar LFs are shown in blue and purple respectively. The $z = 5$ QLF points and best fitting double power law (with fixed $\alpha = -1.8$) are taken from McGreer et al. (2013). Additional upper limits at fainter magnitudes are from Ikeda et al. (2012), where we have shifted the data point at $M_{1500} = -23$ by 0.1 mag for clarity. The $z = 6$ QLF fit and points are taken from Willott et al. (2010), where we have also plotted the best fit with fixed $\alpha = -1.8$. The QLF at $z = 7$ is estimated from the $z = 5$ QLF by using a modified LEDE evolution model for $\log(\phi^*)$ taken from McGreer et al. (2013), with $\alpha = -1.8$ and $M^* = -27.0$.

are weak, illustrated by the large error bars on the faintest bin shown in Fig. 9 (with $M_{1500} \sim -22$), which contains only one quasar.

To ascertain the level of contamination of our sample by quasars, we compare the number densities of bright LBGs at $z = 7$ to the predicted QLF at $z = 7$ by evolving the $z = 5$ QLF determined by McGreer et al. (2013) using the evolution model presented in their paper. A double power law form is typically used when fitting the QLF, as described in Section 7.2. When fitting a DPL to the data points, both McGreer et al. (2013) and Willott et al. (2010) fix the value of the faint-end slope, α , as a response to the large uncertainties in the faint-end determination. Lower redshift $z < 3$ results tend to favour $\alpha = -1.5$ (Croon et al. 2009), with a tenuous steepening observed with $\alpha \simeq -1.7$ to higher redshifts (Masters et al. 2012). McGreer et al. (2013) confirm that a steeper slope of $\alpha = -1.7$ results in a marginally better fit to the data, however, on closer inspection the best-fitting DPL still under-predicts the number of faint quasars that they found. Therefore, in Fig. 9 we plot the best fitting function with $\alpha = -1.8$ from McGreer et al. (2013) to provide an upper limit on the number of faint $z = 5$ quasars. We display the data at $z = 6$ from Willott et al. (2010), overlaid with the best-fitting model with $\alpha = -1.8$, again to show an upper limit on the expected number density of quasars here. The $z = 7$ LF parameters are predicted using the modified form of the Luminosity Evolu-

⁶ <http://casa.colorado.edu/~trenti/CosmicVariance.html>

tion Density Evolution (LEDE) model described in McGreer et al. (2013), with a fixed $\alpha = -1.8$ and $M_* = -27.0$. The model predicts that $\log(\phi^*)$ evolves linearly with redshift with $k = \delta \log(\phi^*)/\delta z = -0.47$, from fitting to the measured parameters from $z = 2.2-4.9$. The strength of the evolution is supported by Venemans et al. (2013) at $z = 7$ who found $k = -0.49^{+0.28}_{-0.74}$. As can be seen in Fig. 9, the LEDE model is consistent with the $z = 6$ LF from Willott et al. (2013) at least given the uncertainties at the faint end, although see the detailed discussions in McGreer et al. (2013). To estimate the number of quasars that could contaminate our sample, and assuming that the quasar SEDs are indistinguishable from the LBGs using the selection here, we integrate the QLF within the three LF points we calculated. We predict 0.10 quasars in the fainter bin at $M_{1500} = -21.75$, and 0.07 and 0.05 in the brighter bins at $M_{1500} = -22.25$ and $M_{1500} = -22.75$ respectively. Hence, from the analysis of the LF and QLF we conclude that contamination of our sample by quasars is minimal, a results obviously consistent with our finding that the vast majority of our $z \simeq 7$ objects are spatially resolved (see Section 6). However, given the large uncertainties in the faint-end slope of the QLF, the possibility of some low-level contamination is not completely ruled out. For example Willott et al. (2010) calculated an upper limit of two quasars per deg^2 in the UltraVISTA survey, when assuming the most extreme LF parameters from the range of acceptable fits to the QLF at $z = 6$. There is evidence that quasars may contaminate bright $z = 6$ LBG samples on the order of $\sim 10\%$, for example the sample of ten LBGs presented in by McLure et al. (2009) and spectroscopically confirmed by (Curtis-Lake et al. 2012), included one Type I quasar identified by the broadened $\text{Ly}\alpha$ line (Willott et al. 2010).

Note that at high redshift, the observed QLFs only account for Type I unobscured quasars with both broad-line and narrow-line components, and there is evidence from X-ray surveys that unobscured quasars only account for 25% of the total number at $z = 4$ (Masters et al. 2012). If the ratio of Type I to Type II quasars persists to high-redshift, the results here are likely a lower limit on the number densities of quasars at $z = 7$.

8.3.1 Radio and X-ray signatures of high-redshift quasars

The rest-frame UV colours of LBGs and quasars at high-redshift are impossible to distinguish with the current photometric accuracy (Bian et al. 2013), and identical colour-colour cut selection criterion are often used for the selection of galaxies and quasars at $z > 5$ (Willott et al. 2010). Here, for completeness, we consider whether radio or X-ray emission could be detected from a quasar selected as a galaxy and included in our sample, with the data available in the COSMOS and UDS fields. The COSMOS field is imaged by the *Chandra*-COSMOS survey (Elvis et al. 2009), which has a limiting depth of 1.9×10^{-16} ergs/s/cm²/Hz in the 0.5-2.0 keV channel and 7.3×10^{-16} ergs/s/cm²/Hz in the 2-10 keV channel. In the radio, the COSMOS field is covered by the Very Large Array (VLA)-COSMOS survey (Schinnerer et al. 2010), with a sensitivity of $12 \mu\text{Jy}$ per beam. The UDS field has X-ray imaging from the Subaru/XMM-Newton deep survey Ueda et al. (2008), to depths of 6×10^{-16} ergs/s/cm²/Hz

in the 0.5-2.0 keV channel and 5×10^{-15} ergs/s/cm²/Hz in the 2-10 keV channel.

We find no radio or X-ray counterparts for any of the galaxies within our sample, when comparing to the publicly available catalogues within each field derived from the datasets described above. We also perform a stack of the objects in the VLA-COSMOS imaging and again find no detection to a limit of $\sim 12 \mu\text{Jy}$ per beam. The average quasar SEDs from Shang et al. (2011) suggests that if one of our candidates was a typical radio-loud quasar, it would just be detectable in the X-ray and radio imaging (see Figure 6 of Ouchi et al. 2009b). Furthermore, a bright radio-loud quasar such as J1429+5447 at $z = 6.21$ (Willott et al. 2010; Frey et al. 2011), would be detected at high-significance in the available radio data. Therefore, although a non-detection in the radio and X-ray for the objects in our sample rules out the possibility that the majority of the objects have strong active nuclei, our sample could still contain $\simeq 1$ radio-quiet quasar. For comparison, the $z = 7.1$ quasar discovered by Mortlock et al. (2011), which is substantially brighter than our LBGs with $M_{\text{UV}} = -26.6$, has been detected in the X-ray (Page et al. 2013) with a flux of $5.7 \pm 1.2 \times 10^{-16}$ ergs/s/cm²/Hz in the 0.5-2.0 keV channel, but not in the radio with a 3σ upper limit of $23.1 \mu\text{Jy}$ per beam (Momjian et al. 2013).

9 ASTROPHYSICAL IMPLICATIONS

As discussed in the introduction to this paper, the Schechter functional form, with its steep exponential decline at high luminosity/mass, undoubtedly provides an excellent description of the galaxy LF and mass function (MF) observed at low-redshift (Montero-Dorta & Prada 2009). Moreover, recent work extending the study of the galaxy luminosity and stellar mass function out to higher redshift, indicate that a Schechter function (or double Schechter function) still provides a good description of the data out to at least $z \simeq 3$ (Ilbert et al. 2013; Muzzin et al. 2013). However, the results we have presented here strongly suggest that this is not the case at $z \simeq 7$.

It is thus worth briefly considering whether our derived $z = 7$ galaxy LF is physically reasonable, and what the failure of a Schechter function to reproduce the bright end of the LF might mean. As already discussed in Section 7, and shown in Fig. 7 one way to describe the apparent lack of a steep exponential decline at the bright end is to parameterise the LF as a double power-law fit, which well describes the full range of available data at $z \simeq 7$. However, the physical meaning of such a double power law is unclear, and moreover it is important to check that the number density of bright $z \simeq 7$ galaxies inferred from our study is not physically unreasonable given the expected number density of appropriate dark matter halos expected to exist at these early times.

We therefore conclude by showing, in Fig. 10, a comparison of the $z \simeq 7$ ΛCDM dark-matter halo mass function, scaled via a constant mass-to-light ratio, with our new observational determination of the $z \simeq 7$ galaxy LF. We produced the halo mass function using the code provided by Reed et al. (2007) using our chosen cosmology and $\sigma_8 = 0.9$, but the basic results are not strongly influenced by the precise choice of code or parameters within current uncertainties. We then

simply scaled the halo mass function into a UV luminosity function using a constant mass-to-light ratio, set by assuming that a galaxy with $M_{1500} = -22.4$ has a stellar mass of $M_* \simeq 10^{10} M_\odot$ (as supported by our data) and a dark matter halo mass to stellar mass ratio of 30 (e.g. Behroozi et al. 2013). As can be seen from Fig. 10, the result is a predicted LF which, without any additional shifting or fitting, does an excellent job of reproducing our new $z \simeq 7$ LF from the previously inferred break luminosity at $M_{1500} \simeq -20$ out to our brightest luminosity bin. Interestingly, over this range, it is evidently indistinguishable from our double power-law fit, confirming that it provides an excellent representation of the data. It is significantly shallower than the exponential decline shown by the pre-existing Schechter function fit, and only starts to deviate from the bright-end power law at very bright magnitudes (thus suggesting that extrapolation of the double power law brightward of $M_{1500} \simeq -23$ will over-predict the number of extremely bright galaxies to be found in future wider area surveys).

This interesting result has a number of potentially important implications. First, it confirms that the number density of bright galaxies revealed in this study is not unreasonable. Neither is the inferred bright-end slope, as this essentially parallels the decline in the number density of appropriate-mass dark-matter halos. Second, it suggests that while the process (e.g. supervovae feedback) which limits star-formation in faint galaxies appears to be in operation at these early times (as evidenced from the difference between the slope of the halo mass function and the UV LF at faint magnitudes), the mechanism that limits high-mass galaxy growth may have yet to impact on the form of the LF at $z \simeq 7$, at least over the luminosity/mass range probed here. Perhaps this is because AGN have yet to grow to the masses and hence luminosities required to eject gas available for future star formation, and certainly there is little evidence for AGN within our galaxy sample (see Section 8). However, without over-speculating we can at least say that, whatever the physical mechanism which ultimately limits the masses of star-forming galaxies, our results are certainly consistent with the redshift invariant ‘mass quenching’ argument proposed by Peng et al. (2010). Since the estimated masses of our brightest galaxies have only just reached a mass comparable to the proposed critical ‘quenching mass’ of $M_* = 10^{10.2} M_\odot$, it is perhaps to be expected that the quenching of star-formation activity in galaxies which causes them to leave the ‘main sequence’ will only be revealed at lower redshifts and/or higher stellar masses than probed by the sample presented here.

10 CONCLUSIONS

We present the results of a new search for bright star-forming galaxies at $z > 6.5$ utilising the very latest data within the UltraVISTA and UKIDSS Ultra Deep Survey (UDS) fields. The 1.65 deg^2 of available overlapping optical, near- and mid-infrared data was analysed using a photometric redshift fitting method, which enables the identification of high-redshift galaxies and the rejection of contaminants such as low-redshift galaxies and galactic dwarf stars. In total we present 34 galaxies, 29 from within the ultra-deep UltraVISTA imaging, one from the deep UltraVISTA region, and

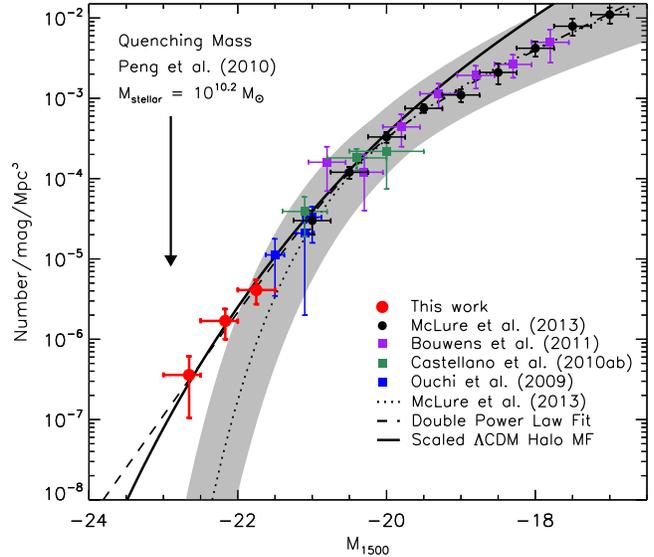


Figure 10. The $z = 7$ UV ($\sim 1500\text{\AA}$) luminosity function showing a scaled Λ CDM halo mass function as described in the text (solid black line). The results from our sample of galaxies from the UltraVISTA DR2 and UDS fields are shown as the red filled circles. Data points from other studies are as described in the caption for Fig. 7. The best-fitting Schechter function at $z = 7$ from McLure et al. (2013) is plotted as the dotted black line, and the best-fitting double power law to our data points and those from McLure et al. (2013) is shown as the dashed line. The one-sigma confidence limit on the Schechter function parameters (M^* , ϕ^* and α) is shown as the grey shaded region. We highlight the quenching mass derived by Peng et al. (2010) with an arrow, after converting to a UV luminosity using the typical mass-to-light ratio displayed by our sample.

four from the UDS field. With the improved photometry available, we reselect seven of the ten candidates presented in Bowler et al. (2012) as $6.5 < z < 7.5$ galaxies. Of the remaining original candidates, two are confirmed as galaxies at a slightly lower redshift of $6.0 < z < 6.5$ and the final candidate is now best fitted as a T-dwarf.

From the best-fitting model to the photometry of each galaxy we calculated the stellar mass, SFR and sSFR. Our sample contains some of the most massive galaxies at $z = 7$, with $M_* \simeq 10^{10} M_\odot$, which show a low sSFR compared to lower mass galaxies at $z = 7$, with an upper limit of $\text{sSFR} \lesssim 2 \text{ Gyr}^{-1}$. We find no evidence for a redder rest-frame UV slope β for our sample (median $\beta_{JHK} = -2.0$), as would be expected by an extrapolation of the colour-magnitude relation at lower redshift to $z = 7$. We measure the sizes of the galaxies in our sample and find that although the majority are consistent with being unresolved in the ground-based imaging, a significant number have a larger FWHM suggesting $r_{1/2} \gtrsim 1.5 \text{ kpc}$. For four galaxies that lie within *HST* imaging from the CANDELS COSMOS survey, we find an offset ($\sim 0.4 \text{ mag}$) between the total magnitudes when the object is assumed to be a point source in both the ground and space-based imaging. Inspection of surface-brightness profiles shows that the galaxies have an extended profile, which can lead to an underestimate of the galaxies total and therefore absolute magnitude when small apertures are

used with the assumption of a point-source profile in *HST* imaging.

From our final sample we determine the form of the bright end of the rest-frame UV galaxy luminosity function (LF) at $z \simeq 7$. We use a $1/V_{\max}$ estimator to determine the binned LF points at $M_{1500} = -22.75, -22.25$ and -21.75 , folding in the completeness of our selection methodology using injection and recovery simulations. In our determination of the LF, we take into account that some of our galaxies are gravitationally lensed by low-redshift galaxies along the line-of-sight, with a typical brightening of $0.1 - 0.3$ mag. We find that the bright end of the $z = 7$ LF does not decline as steeply as predicted by the Schechter function fitted to fainter data, and can be better described by a double power law. The possibility of significant contamination of our sample by high-redshift quasars can be excluded, with a predicted number of < 1 in the UltraVISTA and UDS imaging, calculated from an extrapolation of the $z = 5$ QLF to $z = 7$. From the observed UV LF at $z = 5$ and $z = 6$, we show that a DPL fit can provide a good description of the data and that the bright end of the LF at $z = 6$ and $z = 7$ shows little evolution. Our results at the bright end of the LF mimic the prediction from the scaled dark matter halo MF, suggesting that the physical mechanism which inhibits star-formation activity in massive galaxies (for example AGN feedback) has yet to become efficient at $z \simeq 7$. The interpretation of our results agrees with the phenomenological model of Peng et al. (2010), which would suggest that the most massive $z \simeq 7$ galaxies in our sample have only just reached the critical “quenching mass” of $M_{\star} = 10^{10.2} M_{\odot}$, above which star-formation activity is strongly suppressed.

ACKNOWLEDGEMENTS

RAAB and JSD acknowledge the support of the European Research Council via the award of an Advanced Grant. JSD acknowledges the support of the Royal Society via a Wolfson Research Merit Award. JSD acknowledges the contribution of the EC FP7 SPACE project ASTRODEEP (Ref.No: 312725). RJM acknowledges the support of the Leverhulme Trust via the award of a Philip Leverhulme research prize. RJM acknowledges the support of the European Research Council via the award of a Consolidator Grant (PI McLure). ABR acknowledges the support of the UK Science & Technology Facilities Council. HM acknowledges support of the PNCG. BMJ and JPUF acknowledge support from the ERC-StG grant EGG-278202. The Dark Cosmology Centre is funded by the Danish National Research Foundation. JA gratefully acknowledges support from the Science and Technology Foundation (FCT, Portugal) through the research grant PTDC/FIS-AST/2194/2012 and PEst-OE/FIS/UI2751/2011.

This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatories under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy survey Unit on behalf of the UltraVISTA consortium. This study was based in part on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Re-

search Council (NRC) of Canada, the Institut National des Science de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. This work is based in part on observations made with the NASA/ESA *Hubble Space Telescope*, 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 NASA contract 1407. We thank the staff of the Subaru telescope and Yuko Ideue and Yutaka Ihara for their assistance with the z' -band imaging utilised here.

REFERENCES

- Anders P., Alvensleben U. F., 2003, *A&A*, 401, 1063
 Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
 Ashby M. L. N. et al., 2013, *ApJ*, 769, 80
 Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
 Bernardi M. et al., 2003, *AJ*, 125, 1849
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, *ASP Conf. Proc.*, 281
 Bian F. et al., 2013, preprint (arXiv:1307.4835)
 Bouwens R. J. et al., 2011, *ApJ*, 737, 90
 Bouwens R. J. et al., 2013, preprint (arXiv:1306.2950)
 Bowler R. A. A. et al., 2012, *MNRAS*, 426, 2772
 Bradley L. D. et al., 2012, *ApJ*, 760, 108
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
 Castellano M. et al., 2010a, *A&A*, 511, A20
 Castellano M. et al., 2010b, *A&A*, 524, A28
 Chabrier G., 2003, *PASP*, 115, 763
 Coe D. et al., 2013, *ApJ*, 762, 32
 Croom S. M. et al., 2009, *MNRAS*, 399, 1755
 Curtis-Lake E. et al., 2013, *MNRAS*, 429, 302
 Curtis-Lake E. et al., 2012, *MNRAS*, 422, 1425
 Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, *ApJ*, 617, 746
 de Barros S., Schaerer D., Stark D. P., 2012, preprint (arXiv:1207.3663)
 Dunlop J. S., 2013, *ASSL*, 396, 223
 Dunlop J. S. et al., 2013, *MNRAS*, 432, 3520
 Ellis R. S. et al., 2013, *ApJ*, 763, L2
 Elvis M. et al., 2009, *ApJS*, 184, 158
 Finkelstein S. L. et al., 2013, preprint (arXiv:1310.6031)
 Frey S., Paragi Z., Gurvits L. I., Gabányi K. E., Cseh D., 2011, *A&A*, 531, L5
 Fumagalli M. et al., 2012, *ApJ*, 757, L22
 Furusawa H. et al., 2008, *ApJS*, 176, 1
 Grazian A. et al., 2012, *A&A*, 547, A51
 Grogin N. A. et al., 2011, *ApJS*, 197, 35

- Huang K.-H., Ferguson H. C., Ravindranath S., Su J., 2013, *ApJ*, 765, 68
- Ikeda H. et al., 2012, *ApJ*, 756, 160
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Ilbert O. et al., 2013, *A&A*, 556, A55
- Ilbert O. et al., 2008, *ASP Conf. Ser.*, 399
- Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
- Kauffmann G. et al., 2003, *MNRAS*, 341, 54
- Koekemoer A. M. et al., 2007, *ApJS*, 172, 196
- Koekemoer A. M. et al., 2013, *ApJS*, 209, 16
- Koekemoer A. M. et al., 2011, *ApJS*, 197, 36
- Lawrence A. et al., 2007, *MNRAS*, 379, 1599
- Madau P., 1995, *ApJ*, 441, 18
- Madau P., Pozzetti L., Dickinson M., 1998, *ApJ*, 498, 106
- Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulain E., 2010, *MNRAS*, 401, 371
- Masters D. et al., 2012, *ApJ*, 752, L14
- McCracken H. J. et al., 2012, *A&A*, 544
- McGreer I. D. et al., 2013, *ApJ*, 768, 105
- McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, *MNRAS*, 395, 2196
- McLure R. J. et al., 2006, *MNRAS*, 372, 357
- McLure R. J. et al., 2013, *MNRAS*, 432, 2696
- McLure R. J. et al., 2011, *MNRAS*, 418, 2074
- Momjian E., Carilli C. L., Walter F., Venemans B., 2013, preprint (arXiv:1310.7960)
- Montero-Dorta A. D., Prada F., 2009, *MNRAS*, 399, 1106
- Mortlock D. J. et al., 2011, *Nat*, 474, 616
- Muzzin A. et al., 2013, *ApJ*, 777, 18
- Oesch P. A. et al., 2010, *ApJ*, 709, L21
- Oesch P. A. et al., 2013, preprint (arXiv:1301.6162)
- Oke J. B., 1974, *ApJS*, 27, 21
- Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713
- Ono Y. et al., 2013, *ApJ*, 777, 155
- Ono Y. et al., 2012, *ApJ*, 744, 83
- Osterbrock ., Ferland ., 2006, *Astrophysics of gaseous nebulae and active galactic nuclei*
- Ouchi M. et al., 2013, *ApJ*, 778, 102
- Ouchi M. et al., 2009a, *ApJ*, 706, 1136
- Ouchi M. et al., 2009b, *ApJ*, 696, 1164
- Page M. J., Simpson C., Mortlock D. J., Warren S. J., Hewett P. C., Venemans B. P., McMahon R. G., 2013, preprint (arXiv:1311.1686)
- Peng Y. et al., 2010, *ApJ*, 721, 193
- Postman M. et al., 2012, *ApJS*, 199, 25
- Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, *MNRAS*, 374, 2
- Robertson B. E. et al., 2013, *ApJ*, 768, 71
- Rogers A. B., McLure R. J., Dunlop J. S., 2013a, *MNRAS*, 429, 2456
- Rogers A. B. et al., 2013b, preprint (arXiv:1312.4975)
- Sanders D. B. et al., 2007, *ApJS*, 172, 86
- Schenker M. A. et al., 2013, *ApJ*, 768, 196
- Schinnerer E. et al., 2010, *ApJS*, 188, 384
- Schmidt M., 1968, *ApJ*, 151, 393
- Scoville N. et al., 2007a, *ApJS*, 172, 38
- Scoville N. et al., 2007b, *ApJS*, 172, 1
- Shang Z. et al., 2011, *ApJS*, 196, 2
- Smit R. et al., 2013, preprint (arXiv:1307.5847)
- Sobral D., Best P. N., Matsuda Y., Smail I., Geach J. E., Cirasuolo M., 2012, *MNRAS*, 420, 1926
- Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, *ApJ*, 763, 129
- Trenti M. et al., 2011, *ApJ*, 727, L39
- Trenti M., Stiavelli M., 2008, *ApJ*, 676, 767
- Ueda Y. et al., 2008, *ApJS*, 179, 124
- Venemans B. P. et al., 2013, preprint (arXiv:1311.3666)
- Wilkins S. M. et al., 2013, preprint (arXiv:1308.6146)
- Willott C. J. et al., 2010, *AJ*, 139, 906
- Willott C. J. et al., 2013, *AJ*, 145, 4
- Zheng W. et al., 2012, *Nat*, 489, 406

APPENDIX A: IMAGES AND SED FITS

In this appendix we present postage-stamp images and the best-fitting galaxy and star SED fits for the galaxies in our sample. The 30 galaxies from the UltraVISTA field are shown in Fig. A1, followed by the four UDS galaxies in Fig. A2.

APPENDIX B: BOWLER ET AL. 2012 IMPROVED PHOTOMETRY AND SED PLOTS

Here we present improved photometry and SED fitting results for the 10 high-redshift galaxy candidates from Bowler et al. (2012), and B1 shows postage-stamp images and SED fits to the three Bowler et al. (2012) candidates that are not included in our new sample.

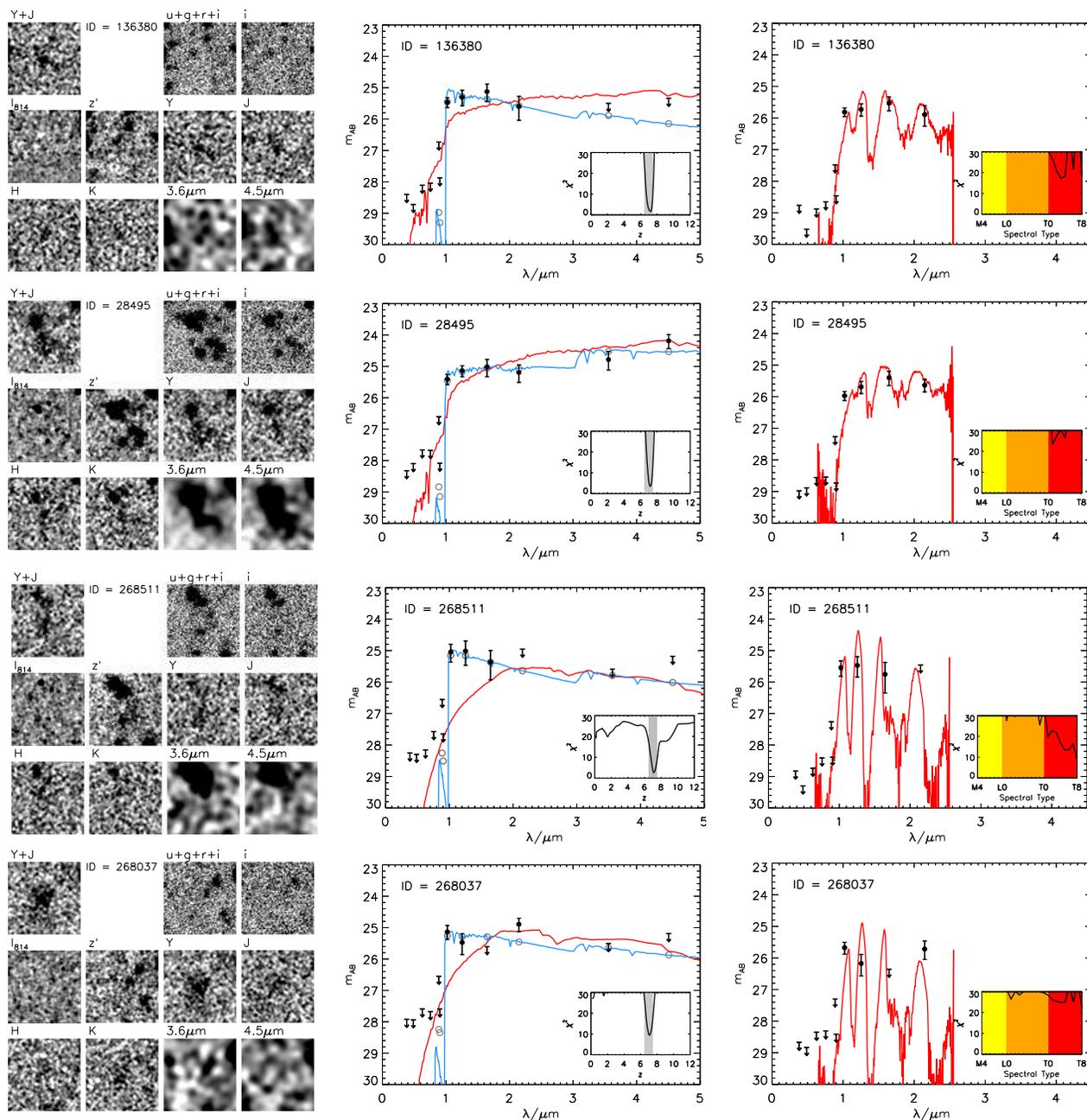


Figure A1. Postage-stamp images and galaxy and star SED fits to the 30 galaxies from the UltraVISTA field. The stamps shown to the left are 10×10 -arcsec, with the grey scale determined from saturating all pixels that exceed 3σ from the background. The measured photometry and errors are shown as the black points, with the grey circles showing the predicted photometry from the best-fitting high-redshift model. In the central plots, the best fitting low-redshift ($z < 4.5$) and high-redshift solutions are shown as the red and blue lines respectively. On the right, the best-fitting stellar templates are presented, where here the photometry was measured in smaller 1.2-arcsec diameter apertures. The insets on each plot show the chi-squared distribution as a function of redshift or stellar type, with the grey band on the redshift- χ^2 plot showing the range of redshifts covered by our sample ($6.5 < z < 7.5$).

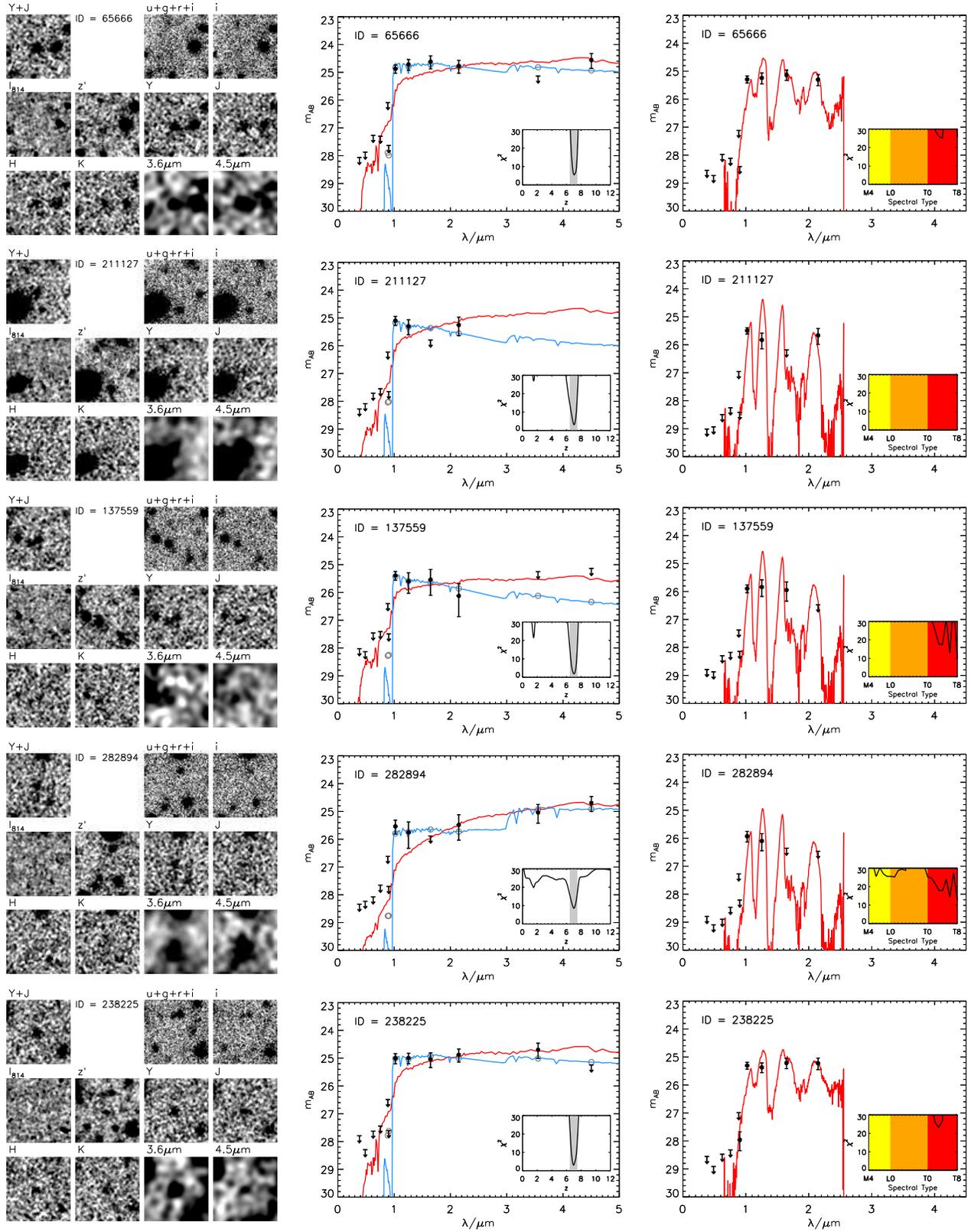


Figure A1 – *continued*

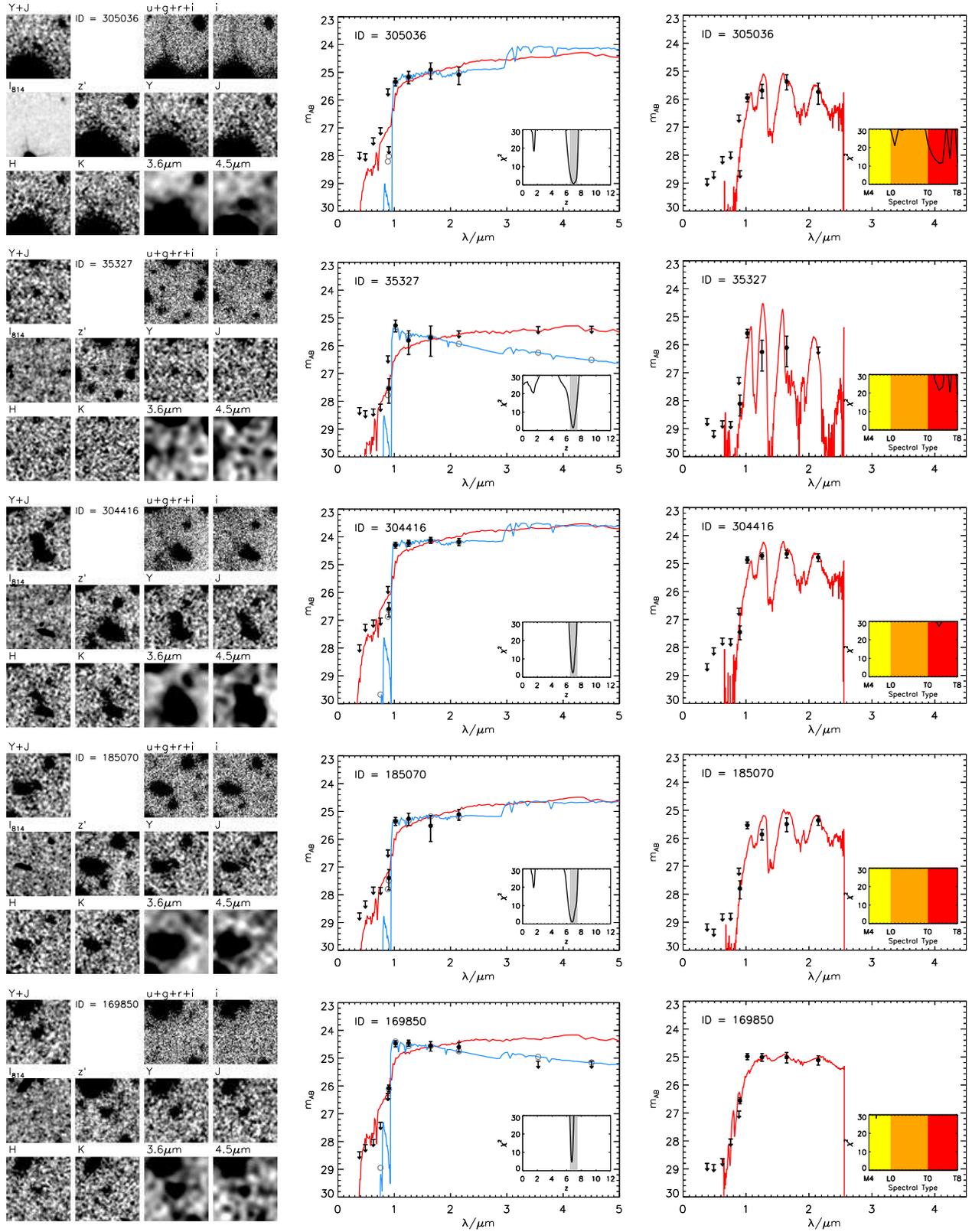


Figure A1 – continued

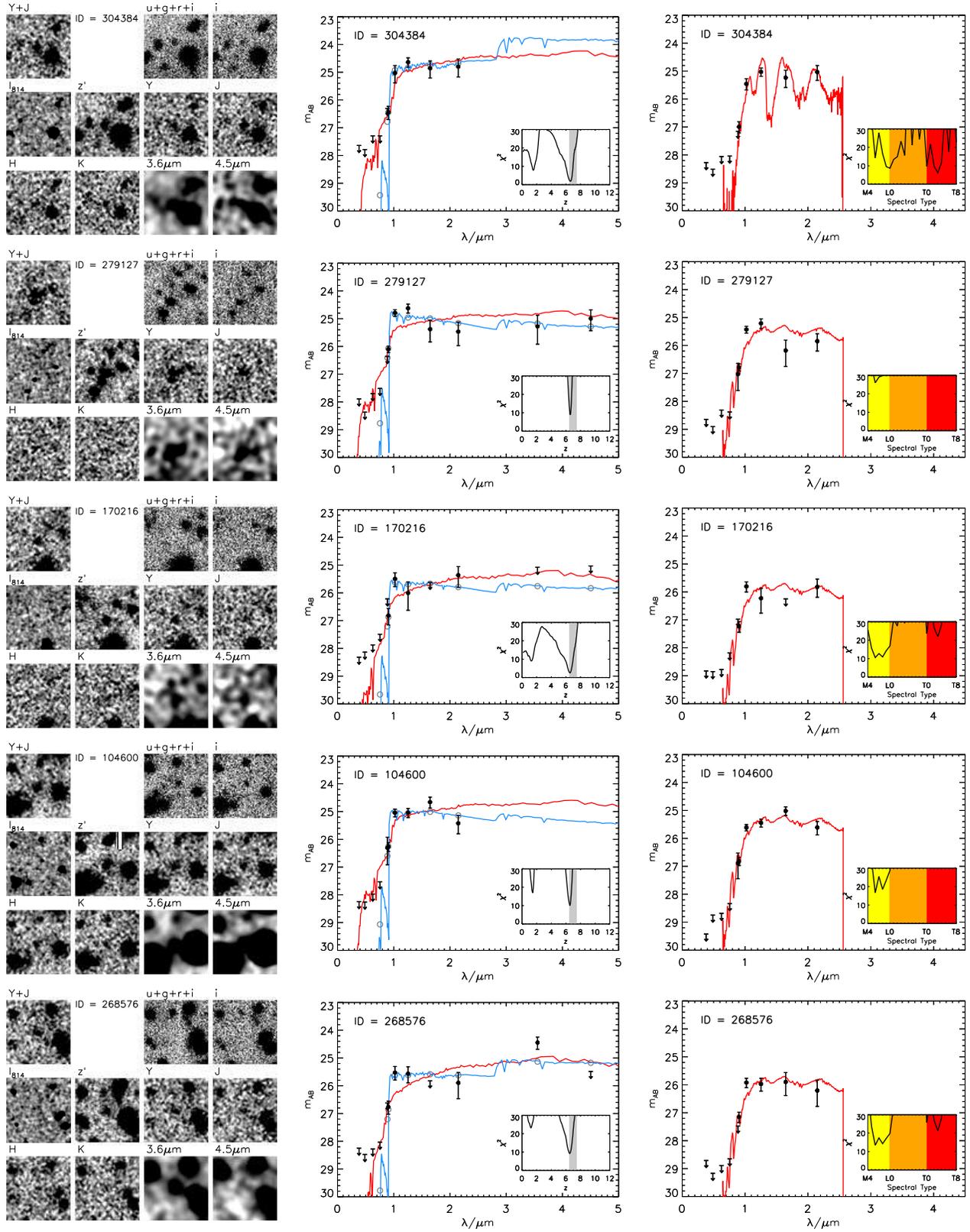


Figure A1 – *continued*

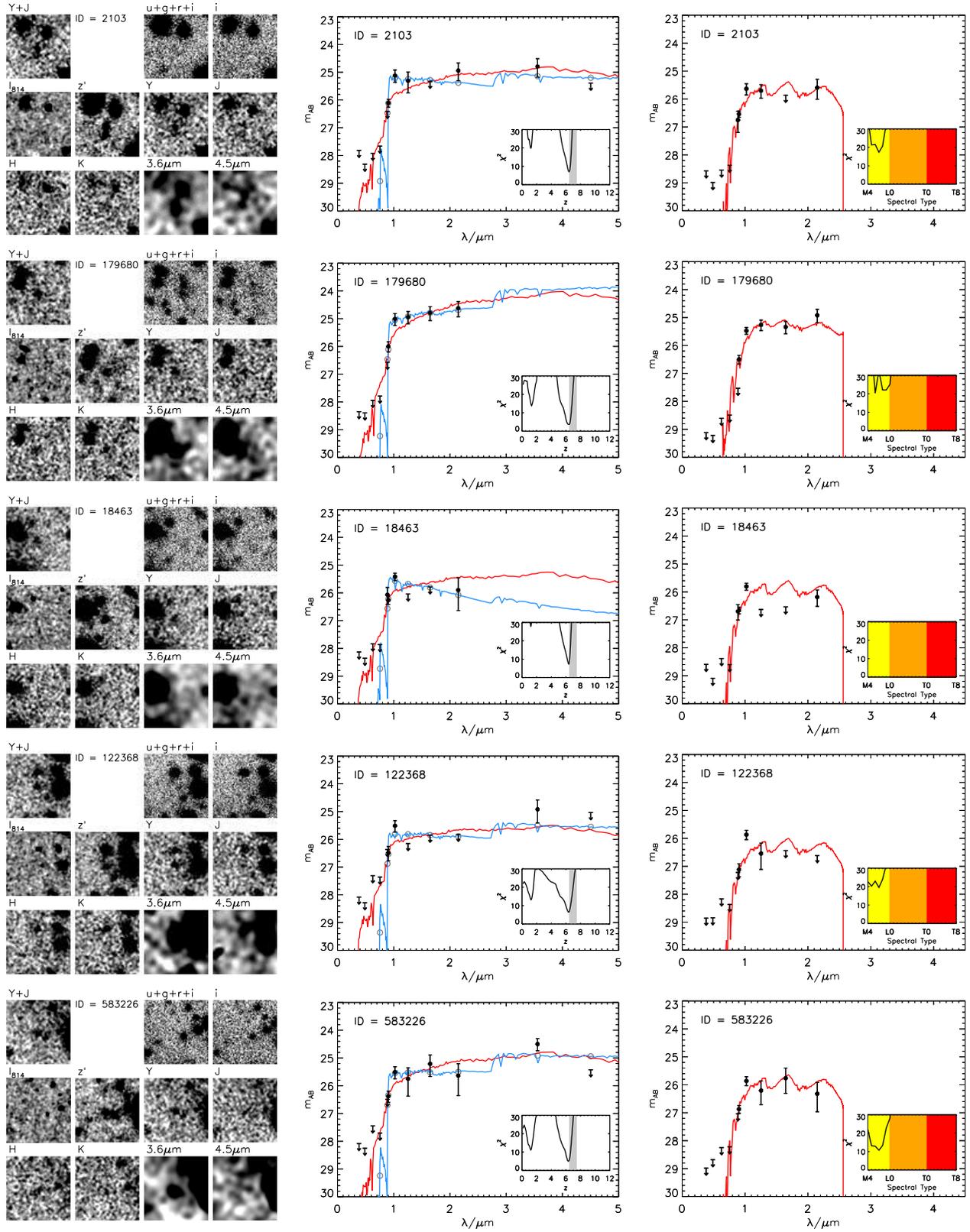


Figure A1 – continued

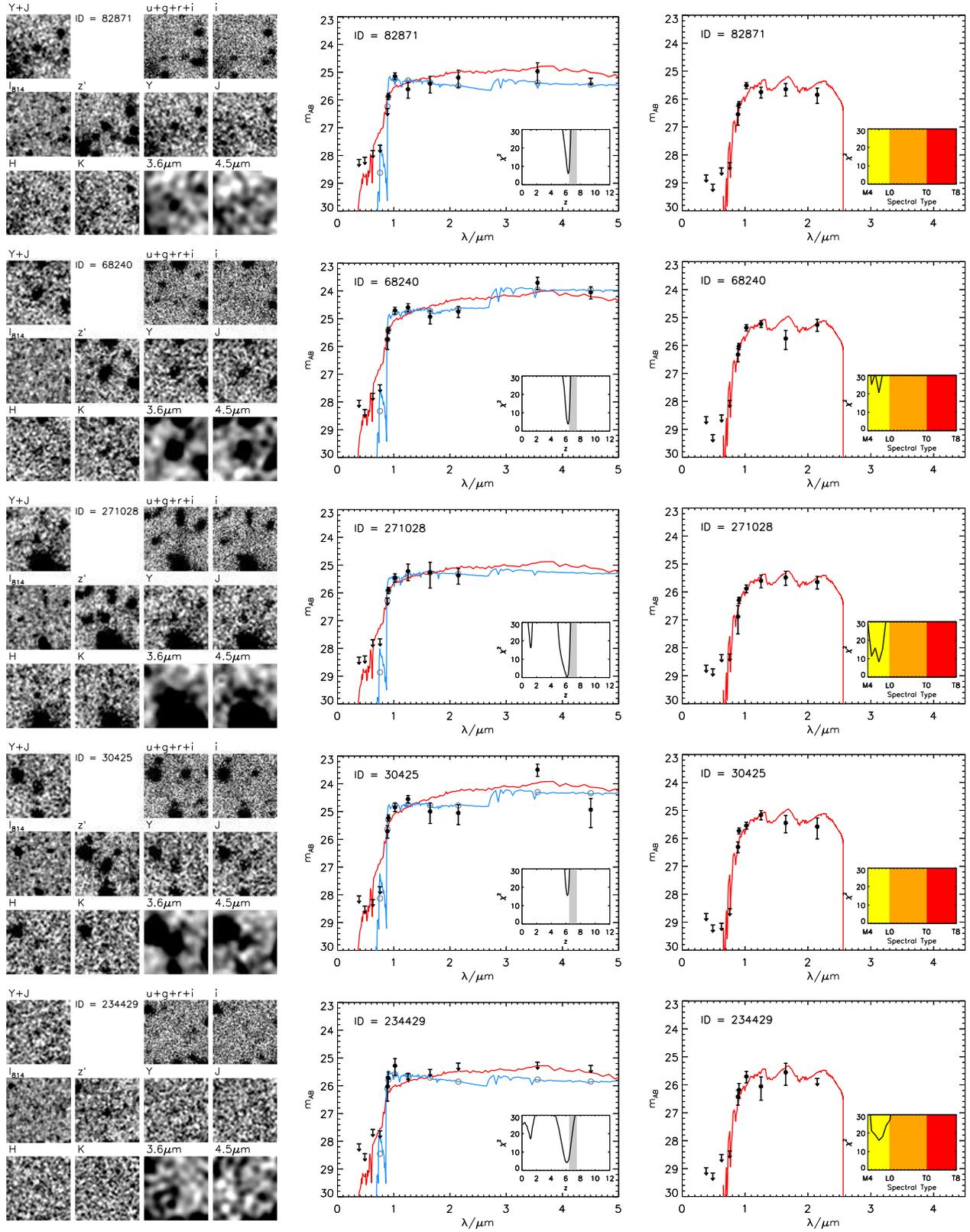


Figure A1 – continued

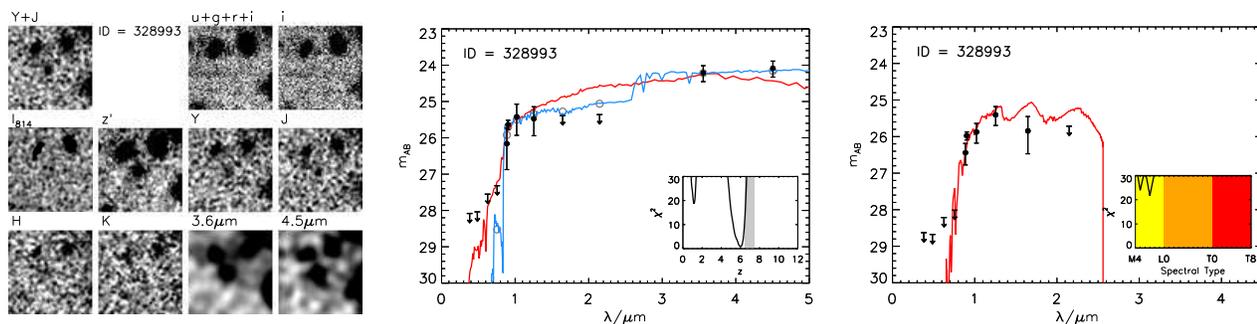


Figure A1 – continued

Table B1. The DR2 UltraVISTA magnitudes for the ten galaxy candidates presented in Bowler et al. (2012). The magnitudes were measured in a 1.8-arcsec diameter circular aperture in all cases except the IRAC magnitudes which were measured in a 2.8-arcsec diameter circular aperture. All magnitudes have been corrected to the 84% enclosed flux level using appropriate point-source corrections, and the errors presented are determined from the local error method detailed in Section 2.3. The signal-to-noise of the detection is presented in brackets after each magnitude. Where an object was detected at less than 2σ significance, the magnitude is replaced with the 2σ local depth as an upper limit.

ID	z'	Y	J	H	Ks	3.6 μ m	4.5 μ m
277912	$26.6^{+0.3}_{-0.2}$ (4)	$24.3^{+0.1}_{-0.1}$ (16)	$24.2^{+0.1}_{-0.1}$ (10)	$24.1^{+0.1}_{-0.1}$ (10)	$24.2^{+0.1}_{-0.1}$ (8)	$23.4^{+0.2}_{-0.2}$ (5)	$23.4^{+0.2}_{-0.2}$ (4)
155880	$26.1^{+0.1}_{-0.1}$ (9)	$24.5^{+0.1}_{-0.1}$ (11)	$24.5^{+0.1}_{-0.1}$ (10)	$24.6^{+0.2}_{-0.2}$ (6)	$24.6^{+0.2}_{-0.2}$ (6)	> 25.1 (1)	> 25.1 (0)
218467	> 27.6 (1)	$25.0^{+0.2}_{-0.2}$ (5)	$25.0^{+0.2}_{-0.2}$ (5)	$25.0^{+0.3}_{-0.2}$ (4)	$24.9^{+0.3}_{-0.2}$ (4)	$24.7^{+0.3}_{-0.2}$ (4)	> 25.2 (1)
61432	> 27.6 (1)	$24.9^{+0.2}_{-0.1}$ (7)	$24.7^{+0.2}_{-0.2}$ (5)	$24.6^{+0.3}_{-0.2}$ (4)	$24.8^{+0.3}_{-0.2}$ (4)	> 25.1 (0)	$24.6^{+0.3}_{-0.2}$ (4)
277880	$26.4^{+0.3}_{-0.3}$ (4)	$25.0^{+0.4}_{-0.3}$ (3)	$24.6^{+0.2}_{-0.2}$ (6)	$24.9^{+0.4}_{-0.3}$ (3)	$24.8^{+0.4}_{-0.3}$ (3)	$24.6^{+0.3}_{-0.2}$ (4)	$24.7^{+0.3}_{-0.3}$ (3)
268511	> 27.6 (0)	$25.0^{+0.3}_{-0.2}$ (3)	$25.0^{+0.5}_{-0.3}$ (2)	$25.4^{+0.6}_{-0.4}$ (2)	> 25.0 (0)	> 25.6 (1)	> 25.2 (1)
271105	$26.1^{+0.1}_{-0.1}$ (7)	$25.0^{+0.1}_{-0.1}$ (9)	$24.1^{+0.1}_{-0.1}$ (12)	$23.9^{+0.1}_{-0.1}$ (11)	$24.0^{+0.1}_{-0.1}$ (13)	$23.5^{+0.2}_{-0.2}$ (5)	$23.4^{+0.2}_{-0.2}$ (5)
95661	$25.4^{+0.1}_{-0.1}$ (15)	$24.8^{+0.2}_{-0.2}$ (5)	$25.0^{+0.4}_{-0.3}$ (2)	> 25.2 (1)	$25.0^{+0.7}_{-0.4}$ (2)	$24.4^{+0.2}_{-0.2}$ (5)	$23.6^{+0.2}_{-0.2}$ (5)
28400	$25.2^{+0.1}_{-0.1}$ (16)	$24.8^{+0.2}_{-0.1}$ (7)	$24.6^{+0.1}_{-0.1}$ (8)	$25.0^{+0.4}_{-0.3}$ (3)	$25.1^{+0.4}_{-0.3}$ (3)	$23.5^{+0.2}_{-0.2}$ (5)	$24.9^{+0.6}_{-0.4}$ (2)
2233	$25.9^{+0.1}_{-0.1}$ (9)	$25.5^{+0.4}_{-0.3}$ (3)	$25.3^{+0.4}_{-0.3}$ (2)	> 25.3 (0)	> 25.5 (1)	$25.1^{+0.5}_{-0.4}$ (2)	> 25.4 (0)

Table B2. The best-fitting photometric redshift parameters and galaxies sizes derived from the improved UltraVISTA DR2 imaging of the ten galaxy candidates presented in Bowler et al. (2012). The photometric redshift is calculated by fitting to all available photometric bands including the IRAC 3.6 μ m and 4.5 μ m filters. The object 28400 has an unusually blue [3.6 – 4.5] colour and hence cannot be fitted well with our templates (that do not include potential nebular emission), and so this object has a large χ^2 -value. Best-fitting redshifts with Ly α emission included are shown in the centre of the table; note that here we do not include the IRAC photometry. The FWHM values presented on the right-hand-side were calculated using SExtractor; missing values indicate that the object was not significantly detected in that band.

ID	No Ly α			With Ly α				Star		FWHM				
	z	χ^2	A_V /mag	Z	z	χ^2	EW $_0$ / \AA	A_V / Z_\odot	Z	Stellar Type	χ^2	z'	Y	J
277912	$6.85^{+0.08}_{-0.08}$	2.4	0.0	1.0	6.84	2.4	0	0.0	1.0	T3	27.2	-	1.7	1.9
155880	$6.70^{+0.05}_{-0.06}$	4.6	0.2	1.0	6.86	4.4	50	0.2	1.0	M6	28.0	1.4	1.7	2.2
218467	$6.98^{+0.12}_{-0.12}$	3.1	0.5	1.0	7.01	3.0	10	0.5	1.0	T3	23.2	1.0	1.0	1.6
61432	$7.04^{+0.16}_{-0.11}$	5.5	0.4	1.0	7.04	5.5	0	0.4	1.0	T4	25.0	1.7	1.1	1.6
277880	$6.67^{+0.11}_{-0.12}$	2.2	0.7	0.2	6.66	2.2	0	0.7	0.2	T3	6.3	1.3	1.1	1.7
268511	$7.12^{+0.14}_{-0.11}$	2.6	0.0	0.2	7.25	2.2	80	0.0	0.2	T8	9.2	1.7	1.4	0.5
271105	$6.51^{+0.05}_{-0.04}$	15.5	1.4	1.0	6.55	22.5	0	1.0	1.0	T0	2.4	1.6	1.1	0.9
95661	$6.25^{+0.10}_{-0.13}$	5.1	0.1	0.2	6.31	4.9	20	0.0	0.2	M7	17.4	2.7	0.8	1.8
28400	$6.20^{+0.10}_{-0.08}$	15.3	0.0	1.0	6.59	13.6	150	0.0	0.2	M7	31.9	1.3	2.1	1.3
2233	$6.24^{+0.12}_{-0.20}$	2.6	0.0	0.2	6.30	2.6	20	0.0	0.2	M5	13.9	1.4	2.8	-

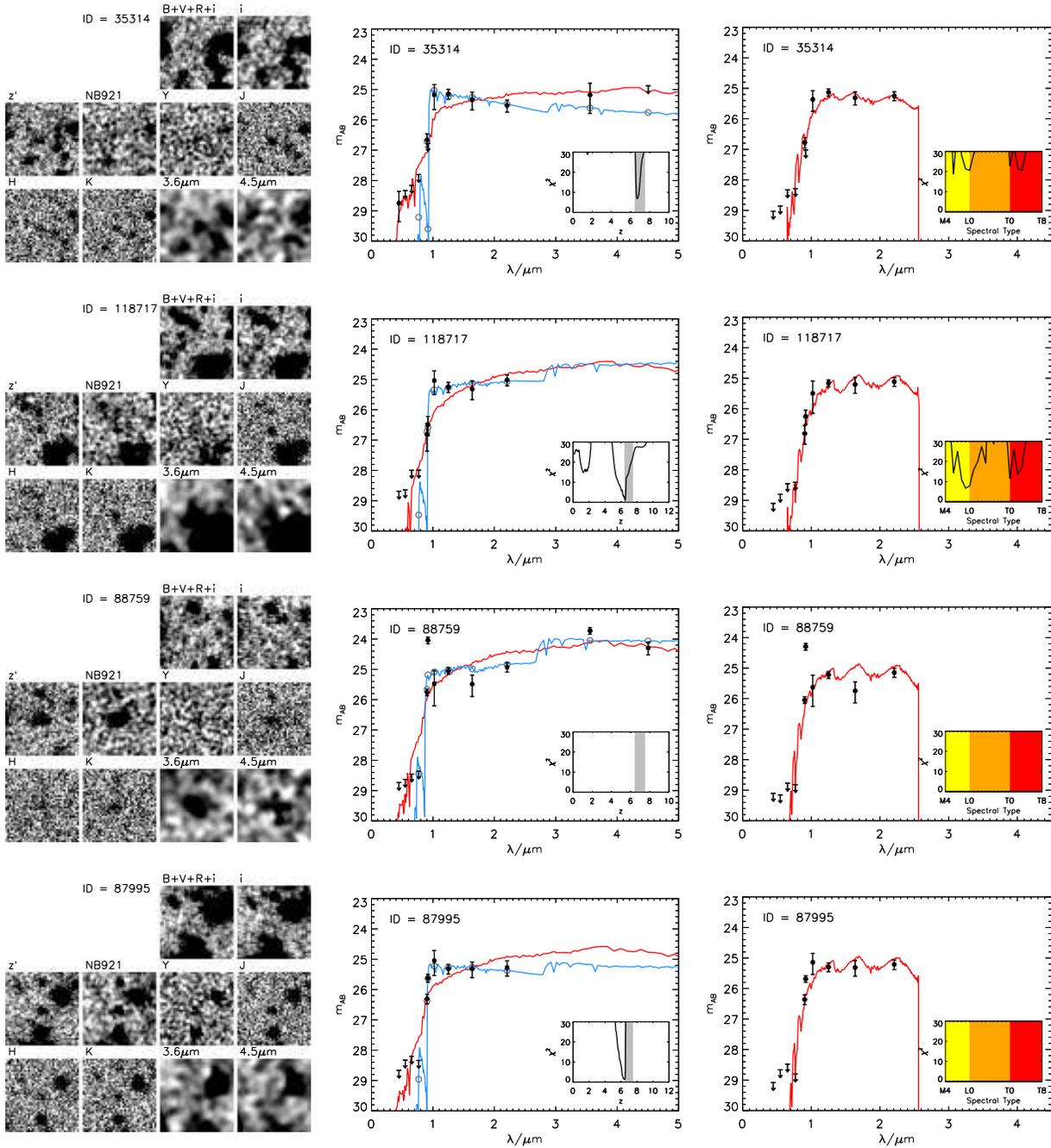


Figure A2. Postage-stamp images and galaxy and star SED fits to the four galaxies from the UDS field. See the caption of Fig. A1 for details.

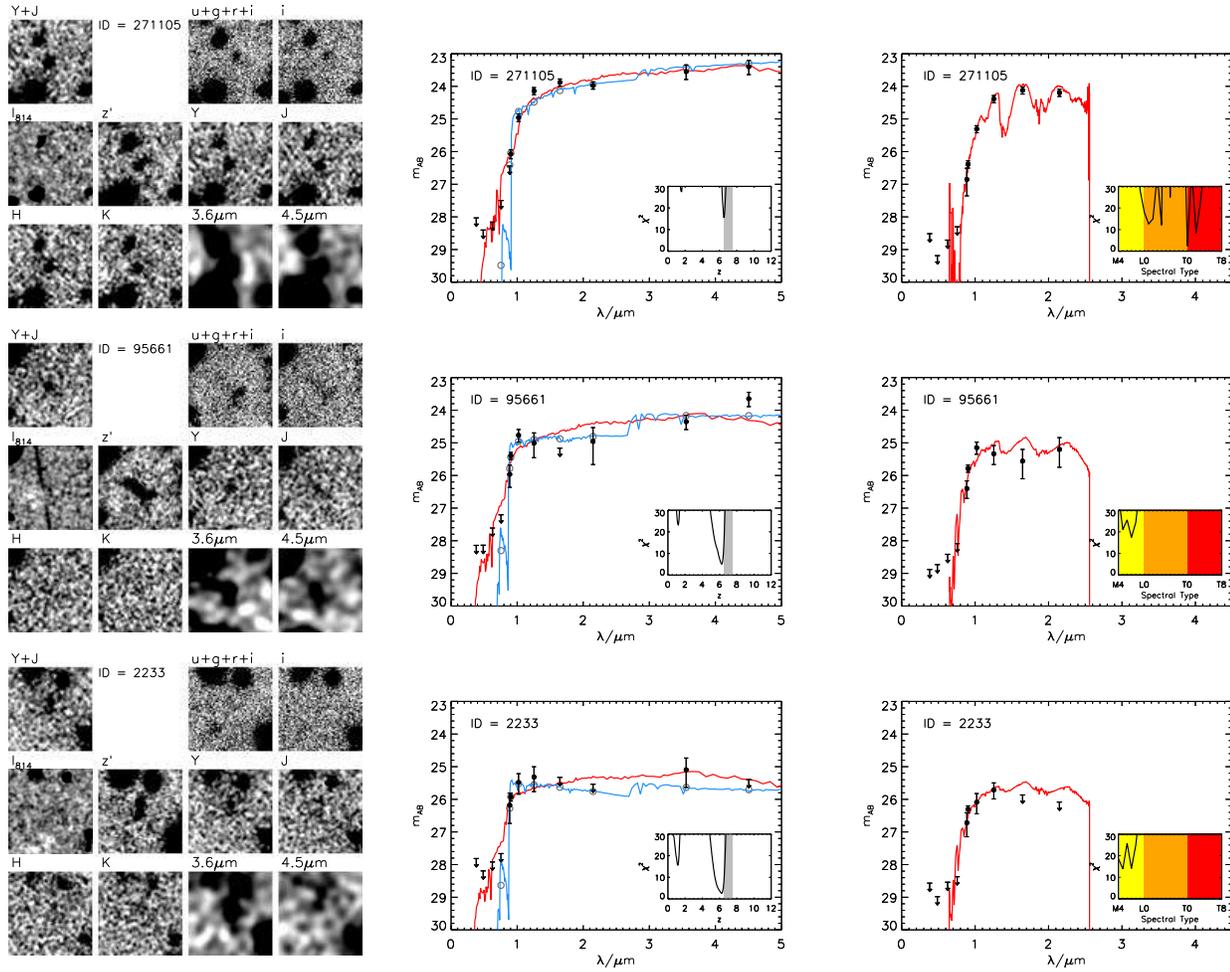


Figure B1. Postage-stamp images, and galaxy and star SED fits for the three candidates from the Bowler et al. (2012) sample that are not present in our final sample. The details of the images and plots are described in the caption for Fig. A1. With the improved photometry, candidate 271105 is now best-fitted as a type-T0 dwarf star, showing the characteristic hook-like spectrum. The other candidates now have best-fitting photometric redshifts in the range $6.0 < z < 6.5$.

The colour distribution of galaxies at redshift five

A. B. Rogers^{1*}, R. J. McLure¹, J. S. Dunlop¹, R. A. A. Bowler¹, E. F. Curtis-Lake¹, P. Dayal¹, S. M. Faber², H. C. Ferguson³, S. L. Finkelstein⁴, N. A. Grogin³, N. P. Hathi⁵, D. Kocevski⁶, A. M. Koekemoer³, P. Kurczynski⁷

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

²*University of California Observatories/Lick Observatory, University of California, Santa Cruz, CA 95064, USA*

³*Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA*

⁴*The University of Texas at Austin, Austin, TX 78759, USA*

⁵*Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France*

⁶*Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506, USA*

⁷*Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA*

Received 2013 December 17

ABSTRACT

We present the results of a study investigating the rest-frame ultra-violet (UV) spectral slopes of redshift $z \approx 5$ Lyman-break galaxies (LBGs). By combining deep *Hubble Space Telescope* imaging of the CANDELS and HUDF fields with ground-based imaging from the UKIDSS Ultra Deep Survey (UDS), we have produced a large sample of $z \approx 5$ LBGs spanning an unprecedented factor of > 100 in UV luminosity. Based on this sample we find a clear colour-magnitude relation (CMR) at $z \approx 5$, such that the rest-frame UV slopes (β) of brighter galaxies are notably redder than their fainter counterparts. Our determination of the $z \approx 5$ CMR is well described by a linear relationship of the form: $d\beta = (-0.12 \pm 0.02)dM_{UV}$, with no clear evidence for a change in CMR slope at faint magnitudes (i.e. $M_{UV} \geq -18.9$). Using the results of detailed simulations we are able, for the first time, to recover the intrinsic (i.e. free from noise) variation of galaxy colours around the CMR at $z \approx 5$. We find significant (12σ) evidence for intrinsic colour variation in the sample as a whole. Our results also demonstrate that the width of the intrinsic UV slope distribution of $z \approx 5$ galaxies increases from $\Delta\beta \simeq 0.1$ at $M_{UV} = -18$ to $\Delta\beta \simeq 0.4$ at $M_{UV} = -21$. We suggest that the increasing width of the intrinsic galaxy colour distribution and the CMR itself are both plausibly explained by a luminosity independent lower limit of $\beta \approx -2.1$, combined with an increase in the fraction of red galaxies with increasing UV luminosity.

Key words: galaxies: high-redshift – galaxies: evolution – galaxies: formation – galaxies: starburst

1 INTRODUCTION

The rest-frame ultra-violet (UV) properties of star-forming galaxies at $z \gtrsim 3$ can potentially provide a powerful insight into the formation and evolution of galaxies at high redshift. In common with local star-burst galaxies (Steidel et al. 1999; Meurer et al. 1999), the UV light of high-redshift galaxies is dominated by short-lived massive stars which should provide a good probe of the current metallicity and dust conditions within rapidly evolving stellar populations.

These galaxies are identified by spectral discontinuities at rest-frame 912 Å and 1216 Å, so are typically termed

Lyman-break galaxies (LBGs, see Giavalisco 2002) regardless of how they are selected. Redward of the Lyman break, the rest-frame UV continuum of star-forming galaxies is traditionally parametrized by the spectral index β , under the assumption (see Leitherer et al. 1999) that the UV continuum can be approximated by a power-law: $f_\lambda \propto \lambda^\beta$. Studies of star-forming galaxies at low redshift have revealed a strong relationship between UV slope and reddening (Meurer et al. 1999), such that the spectral slope β and differential extinction $E(B - V)$ are often quoted interchangeably. More recent studies exploiting *Herschel* data have shown that the relationship between β and dust attenuation appears to be already in place by $z \approx 2$ (Reddy et al. 2012), even though the youngest galaxies (< 100 Myr) may

* E-mail: abr@roe.ac.uk

† Scottish Universities Physics Alliance

follow a different extinction curve to their older counterparts (Reddy et al. 2010).

At higher redshifts ($z \geq 3$) the relationship between UV slope, dust attenuation, stellar population age and metallicity is still unclear. However, many previous studies have shown that β reddens with decreasing redshift and increasing UV luminosity (e.g. Meurer et al. 1999; Shapley et al. 2003; Labbé et al. 2007; Overzier et al. 2008; Wilkins et al. 2011; Bouwens et al. 2013; Dayal & Ferrara 2012). This trend is often assumed to reflect increasing dust attenuation at a fixed redshift, and increasing stellar population age at a fixed luminosity (e.g. Labbé et al. 2007).

The obvious prediction from the apparent trend between UV slope, redshift and luminosity is that the bluest galaxies will naturally be found amongst the faintest detectable galaxies at $z \geq 7$. Indeed, initial analysis of ultra-faint LBGs detected in near-IR *HST* imaging of the Hubble Ultra-deep field (HUDF) suggested a population of galaxies with very blue UV slopes (i.e. $\beta \leq -3$), which would require dust free, ultra-low metallicity stellar populations (Bouwens et al. 2010, but see Finkelstein et al. 2010). However, careful consideration of the observational biases (e.g. Dunlop et al. 2012; Finkelstein et al. 2012; Rogers et al. 2013) and the availability of yet deeper *HST* imaging (Ellis et al. 2013; Koekemoer et al. 2013) has led to a consensus that the UV slopes of currently observable $z \approx 7$ galaxies are not significantly bluer than those of moderately young, but otherwise unremarkable, stellar populations (Dunlop et al. 2013; Bouwens et al. 2013).

As a consequence, attention has now turned to understanding the details of how the relationship between UV slope and luminosity, hereafter the colour-magnitude relation (CMR), evolves at high redshift ($z \geq 4$). At present, there is no consensus on the strength or functional form of the CMR at high redshift, with the two largest studies at $z \geq 4$, those of Finkelstein et al. (2012, hereafter F12) and Bouwens et al. (2013, hereafter B13), producing seemingly discrepant results. Although both studies are based on samples with a reasonable dynamic range in UV luminosity, selected from high-quality *HST* imaging, B13 find a significant CMR in redshift bins at $z = 4, 5, 6$, & 7 , whereas F12 see no significant correlation between β and M_{UV} in the same redshift range. In contrast, F12 find that β is more strongly coupled to stellar mass M_* , with more massive galaxies displaying redder UV slopes.

In addition to fundamental questions related to the existence and form of the CMR at high redshift, constraints on the *intrinsic*¹ colour distribution of $z \geq 4$ galaxies are clearly of interest.

For instance, at $z < 3$, Labbé et al. (2007) found evidence for moderate intrinsic colour-variation in the blue sequence, which they attributed to stochastic star-formation histories. However, previous studies addressing this issue at $z \approx 4 - 5$ have been hampered by a lack of dynamic range (Castellano et al. 2012; Bouwens et al. 2012; Wilkins et al. 2011) and no luminosity-dependent trend is currently clear.

The initial aim of this paper is to combine the strength of deep, small area, *HST* imaging with shallower, but wide

area, ground-based imaging from the UKIDSS Ultra Deep Survey (UDS) to provide a large sample of $z \approx 5$ LBGs spanning an unprecedented dynamic range in UV luminosity (≈ 5 magnitudes). We focus entirely on $z \approx 5$ galaxies because this is the highest redshift for which it is possible to consistently select large samples of galaxies free from Ly α contamination using the deep z -band imaging available across our *HST* and ground-based datasets. Based on the techniques developed in Rogers et al. (2013), we use bias-free measurements of β to provide the best available constraints on the form of the $z \approx 5$ CMR, before proceeding to exploit the results of detailed simulations to investigate the *intrinsic* galaxy colour distribution as a function of luminosity.

The remainder of this paper is organised as follows. In Section 2 we describe the selection of our $z \approx 5$ galaxy sample from our chosen *HST* and ground-based datasets. In Section 3 we briefly recount our method for measuring rest-frame UV colours and present our constraints on the $z \approx 5$ CMR. In Section 4 we describe the detailed simulations which were necessary to accurately quantify the contribution of photometric uncertainties to the observed galaxy colour distribution. Based on the results of these simulations, in Section 5 we present our measurement of the intrinsic scatter in the galaxy colour distribution as a function of UV luminosity. Our conclusions are summarised in Section 6. Throughout the paper we quote magnitudes in the AB system (Oke & Gunn 1983) and assume a cosmology with $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. We refer to *HST*'s ACS F435W, F606W, F775W, F814W, F850LP and WFC3/IR F098M, F105W, F125W, F140W and F160W filters as B_{435} , V_{606} , i_{775} , I_{814} , z_{850} , Y_{098} , Y_{105} , J_{125} , JH_{140} and H_{160} respectively.

2 DATA AND SAMPLE

In this section we describe the selection of our $z \approx 5$ galaxy sample. The sample is selected from four survey fields: the UKIRT Infrared Deep Sky Survey (UKIDSS) Ultra Deep Survey (UDS), the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) programme's coverage of the Great Observatories Origins Deep Survey (GOODS) North and South fields, and the Hubble Ultra Deep Field (HUDF). Summary details of the four fields and their respective $z \approx 5$ LBG samples are given in Tables 1 and 2.

2.1 Description of imaging

Here, we briefly describe the available imaging in each field. The fields themselves are shown in Fig. 1, while the band-passes of the filters are shown in Fig. 2.

2.1.1 UKIDSS Ultra Deep Survey (UDS)

The UDS is covered by public Subaru B, V, R, i, z' -band (Furusawa et al. 2008) and UKIRT J, H, K -band imaging², with a co-imaged area of 0.6 square degrees. In addition, we have

¹ Throughout this work, we refer to the noise-free distribution of colours as 'intrinsic', i.e. free of observational effects.

² The images we use are from DR10. See <http://www.ukidss.org>

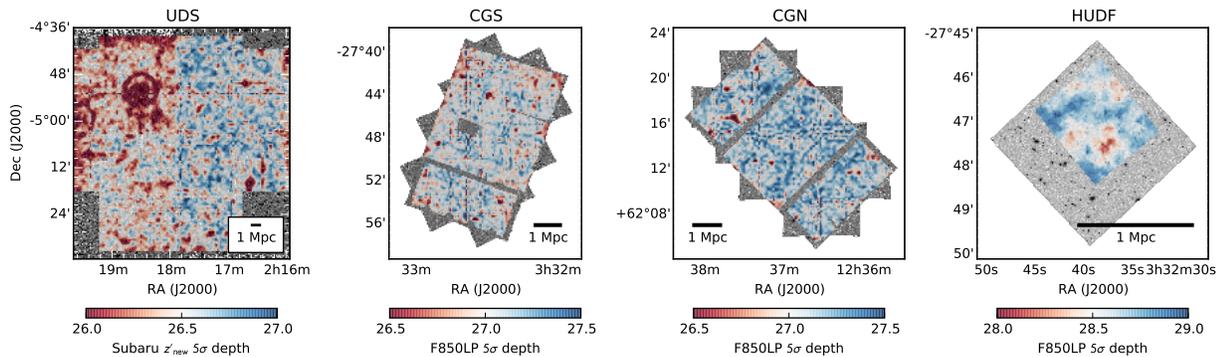


Figure 1. The four fields analysed in this study. In each, the grayscale image shows the z -band imaging in which the galaxies are detected, while the colour-map shows the local 5σ z -band depth from which SNR cuts are applied. Depths are computed at nodes of a 100×100 grid, based on each node’s 200 nearest source-free apertures (see Section 2.3 for a discussion). The footprint of each depth-map defines our search area, i.e. the area in which imaging exists in all wavebands. The scalebar denotes a *physical* size of approximately 1 Mpc at $z = 5$.

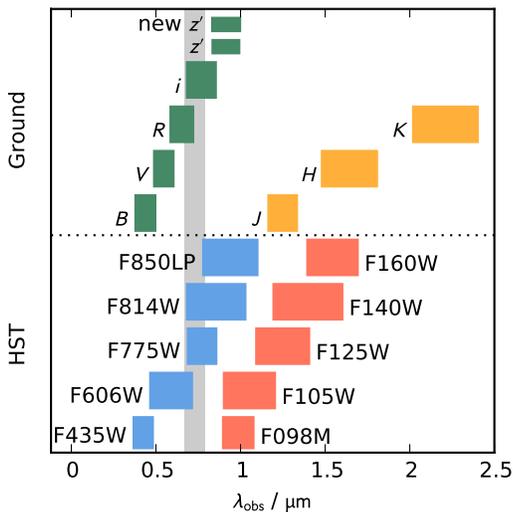


Figure 2. The photometric filter bandpasses for the images used in this work are shown by the coloured regions. Blue and red regions show the *HST*’s ACS and WFC3/IR filters, while green and yellow regions show the Subaru and UKIRT filters used in the UDS. The vertical, grey region denotes the wavelength range in which the 1216 Å Lyman break is observed at $4.5 < z < 5.5$.

made use of additional z' -band imaging taken after Subaru’s SuprimeCam was refurbished with CCDs with improved red sensitivity. For the remainder of this work, the public z' -band is referred to as z_{old} and the new, deeper, proprietary z' -band as z_{new} .

2.1.2 CANDELS GOODS-N (CGN) and CANDELS GOODS-S (CGS)

The two CANDELS GOODS fields together provide 283 arcmin^2 of *HST* ACS and WFC3/IR imaging. The survey and data reduction are described by Grogin et al. (2011) and Koekemoer et al. (2011) respectively. In each field the WFC3/IR imaging consists of a deep central region, flanked by two ‘wide’ strips. In CGS, one of these wide strips is pro-

vided by the Early Release Science field (ERS, Windhorst et al. 2011), which features deep Y -band coverage through the Y_{098} filter rather than the Y_{105} filter which is employed over the rest of the CGN and CGS fields (see Table 1).

2.1.3 Hubble Ultra Deep Field (HUDF)

The HUDF has been imaged by multiple programmes, most recently in the near-IR by the UDF12 campaign (Ellis et al. 2013; Koekemoer et al. 2013). We have analysed the UDF12 near-IR imaging in tandem with the ACS optical B_{435} , V_{606} , i_{775} , z_{850} -band imaging provided by Beckwith et al. (2006) and the more recently assembled I_{814} imaging provided as part of the XDF (Illingworth et al. 2013). For this work, we searched only the area covered by the deepest WFC3/IR imaging ($\approx 4 \text{ arcmin}^2$).

2.2 Photometry

Fixed-diameter circular apertures were used to construct photometric catalogues from each image. In the UDS imaging, 1.8-arcsec diameter apertures were used in all bands, enclosing $\approx 80\%$ of a point source’s flux. While a point source is a reasonable approximation to a $z \approx 5$ galaxy in the ground-based data, *HST* allows many of the brighter galaxies to be resolved. This is advantageous, as it allows stars to be easily distinguished from galaxies using their measured half-light radii ($r_{1/2}$). However in small PSF-matched apertures, broader sources lose a larger fraction of their light in short wavebands, resulting in a red colour bias for extended objects. There are various options to alleviate this: PSF homogenisation, which relies on a well-constrained transfer function to match the PSF of each image to that of the H_{160} data (which has the poorest spatial resolution); a measured-size dependent correction to the aperture photometry, which relies on well-measured half-light radii; or the adoption of sufficiently large apertures that realistic high-redshift galaxy sizes ($r_{1/2} \lesssim 1.5 \text{ kpc}$) have a negligible impact, at the expense of image depth. For this work, we assume the third approach and use apertures of diameter 0.6 arcsec for all *HST* photometry. These apertures enclose a sufficient fraction of the

Field	Area/arcmin ²	SNR cut	5 σ depth/AB mag									
			<i>B</i> ₄₃₅	<i>V</i> ₆₀₆	<i>i</i> ₇₇₅	<i>I</i> ₈₁₄	<i>z</i> ₈₅₀	<i>Y</i> ₀₉₈ ^a	<i>Y</i> ₁₀₅	<i>J</i> ₁₂₅	<i>JH</i> ₁₄₀	<i>H</i> ₁₆₀
HUDF	4	5	29.2	29.6	29.2	28.4	28.6	–	29.5	29.2	29.2	29.3
CGS	143	8	27.6	27.8	27.3	–	27.0	27.1	27.5	27.5	–	27.3
CGN	140	8	27.6	27.8	28.4	–	27.2	–	26.9	27.1	–	27.0
			<i>B</i>	<i>V</i>	<i>R</i>	<i>i</i>	<i>z</i> _{old}	<i>z</i> _{new}	<i>J</i>	<i>H</i>	<i>K</i> _s	
UDS	2701	12	27.8	27.5	27.2	27.1	26.1	26.5	25.6	25.1	25.3	
TOTAL:	2988											

Table 1. Summary of the fields used in this work. Columns 1 and 2 list the field names and the corresponding area of each used to search for $z \approx 5$ galaxies. Column 3 lists the z -band SNR threshold adopted for detecting galaxies in each field. Columns 4–13 list the depth of imaging in each field. Quoted depths are medians across the field, measured in circular apertures of diameter 0.6 arcsec (*HST*) or 1.8 arcsec (UDS); however, for selection, local depth measurements are used (see Section 2.3). The CANDELS CGS and CGN have variable depth NIR imaging, typically ranging from 27 – 28 mag, while the UDS imaging varies in the z_{new} selection band as shown in Fig. 1. ^a – Y_{098} imaging is used in place of Y_{105} for the northern strip (ERS; 44 arcmin²) of the GOODS-S data.

total light in all bands that biases in the UV slope are at the level of $|\delta\beta| \lesssim 0.2$ (see Section 4).

2.3 Image depths

Given the variable exposure-time maps of each survey field, and the importance of consistent signal-to-noise cuts across our sample, the SNR of the photometry for each candidate, in each filter, has been computed relative to the local image depths. To do so we created an object (segmentation) mask for each mosaic using *SEXTRACTOR* v2.8.6 (Bertin & Arnouts 1996), set to mask out any area where two or more pixels rise above 1.4σ , and placed non-overlapping apertures across the remaining source-free sky regions. The local depth at a given point on the mosaic was then measured by analysing the nearest 200 source-free apertures and computing the 5σ depth using the Median Absolute Deviation (MAD) statistic, which yields the standard deviation of the distribution of fluxes by $\sigma \approx 1.4826 \times \text{MAD}$. The MAD is ideal for these measurements as it is robust against very non-Gaussian distributions. The depth maps shown in Fig. 1 were created by computing local depths at nodes of a grid across the images, while the values quoted in Table 1 are medians over each field.

2.4 Selection of $z \approx 5$ galaxies

Within each field, candidates were initially detected using *SEXTRACTOR* in dual-image mode, detecting in the z -band image and measuring from all others in fixed, circular apertures as discussed above. SNR thresholds were then applied in the z -band, to remove sources detected at low significance, and in the B -band, since both the Lyman break ($\lambda_{\text{rest}} = 1216 \text{ \AA}$) and limit (912 \AA) of a genuine $z \approx 5$ galaxy would lie redward of the B -band. To do so, the catalogues were first reduced in size by keeping only candidates with z -band detections brighter than the $3\sigma_{\text{global}}$ depth (in the HUDF) or the $5\sigma_{\text{global}}$ depth in CGN, CGS and the UDS, where σ_{global} is the median of local depths across a given image. At the same time, the catalogues were pruned of any object with a B -band detection at the $2\sigma_{\text{global}}$ level. For the remaining candidates, local depth estimates were computed in each band, using the procedure outlined in Section 2.3. A second cull of $B \geq 2\sigma_{\text{local}}$ detected sources removed the

few contaminants lying in deeper parts of the image. Finally, refined z -band signal-to-noise cuts were imposed at local $5\sigma, 8\sigma, 8\sigma, 12\sigma$ thresholds for candidates in the HUDF, CGN, CGS and UDS (as per Table 1). These thresholds were chosen as compromises between sample size and data quality, which is itself dependent on the homogeneity of depths and the number of bands probing the rest-frame UV. In particular, the data quality of the five bands in the HUDF sampling the rest-frame UV (z_{850} to H_{160}) allow reliable results to be drawn for objects detected at $z_{850} > 5\sigma$, whereas the other fields require $> 8\sigma$ detections for similarly reliable results.

2.4.1 Photometric redshift analysis

Redshift $4.5 \leq z \leq 5.5$ galaxies were selected using the photometric redshift code *LEPHARE* (Arnouts et al. 1999; Ilbert et al. 2006), adopting the ‘COSMOS’ galaxy models of Ilbert et al. (2006) and galactic dwarf star templates from the *SpeX* library³. The ‘COSMOS’ spectral energy distributions (SEDs) include elliptical, spiral and star-burst templates, which were fit over the redshift range (0, 10) allowing dust reddening of $E(B - V) \leq 1.5$ with a Calzetti et al. (2000) reddening law. While our local depth estimates are robust, we ensure the entire observed SED contributes to the fit by imposing *minimum* flux errors of 10%⁴.

Candidate $z \approx 5$ galaxies were thus selected, regardless of how marginally the $z \approx 5$ photometric redshift solution was preferred to any secondary solution. As discussed at length by Rogers et al. (2013) and Dunlop et al. (2013), introducing any form of $\chi^2(\text{primary}) - \chi^2(\text{secondary})$ threshold inevitably biases the selection against faint, intrinsically (or photometrically scattered) red galaxies, although this effect is mostly alleviated for galaxies detected at $\text{SNR} \gtrsim 8\sigma$ (Dunlop et al. 2012). In reality, given our chosen redshift and signal-to-noise windows and large apertures, lower redshift models typically struggle to fit the shape of the observed purported Lyman breaks. Fig. 3 shows an example of the

³ <http://pono.ucsd.edu/~adam/browndwarfs/spexprism>. For further details see Acknowledgments.

⁴ In practice, these minimum errors were adopted in the NIR bands for around 10% of candidates brighter than $z > 5\sigma$ in the UDS, to 50% of candidates in the HUDF.

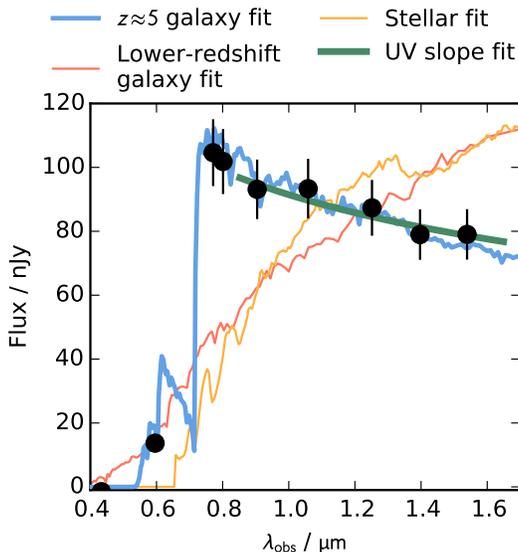


Figure 3. The spectral energy distribution of an example $z \approx 5$ galaxy in the HUDF is shown by black points, in the observed frame. The lines show the various fits used to analyse the galaxy. The error bars are inflated to at least 10% of the flux for the fitting (as shown here; and see Section 2.4.1). In blue is the best-fitting primary redshift solution. The possibility of it being a contaminant is ruled out by the poor lower-redshift galaxy and stellar fits (red and yellow). The rest-frame UV spectral index β is deduced by fitting with a power-law model (green) to the observed near-IR data.

model fitting procedure. Genuine high-redshift galaxies were considered to be those for which the primary galaxy model SED (with an assumed four degrees-of-freedom) was acceptable at the 2σ level, i.e. $\chi^2 \leq 11.3, 9.7, 7.9$ in the HUDF, UDS and CGS/N fields respectively. Potential stellar contaminants were rejected if the best stellar fit was acceptable at 2σ and if the source’s z -band half-light radius $r_{1/2}$ was consistent with being stellar at that luminosity.

Finally, the imaging and SED fits for each candidate were visually inspected. Having survived the selection procedure thus far, objects were only deleted if they were deemed to lie too close to the image edge, or if their SED was acceptably stellar but their proximity to another source led them to be measured with an excessive $r_{1/2}$.

Having applied the above selection procedure, our sample consists of 584 LBG candidates at $z \approx 5$, selected over 2988 arcmin². The sample covers a long luminosity baseline of $-22.5 < M_{1500} < -17.5$, large enough to provide strong leverage on the CMR. A summary of the sample’s properties, broken down by field, is given in Table 2.

2.5 Selection method validation

To check our sample selection method, we have compared our CANDELS GOODS-S (CGS) sample to other catalogues. Comparing to the ESO GOODS/CDF-S Spectroscopy master catalogue⁵, we find 23 sources with spec-

⁵ <http://www.eso.org/sci/activities/garching/projects/goods/MasterSpectroscopy.html>

Field	N	$\langle z \rangle$	$\langle M_{UV} \rangle$	$\langle \beta \rangle$	Mean($\delta\beta$)	
					(data)	(sim)
HUDF	33	5.1	-18.5	-2.04 ± 0.05	0.26	0.26
CGS	112	4.9	-20.2	-1.82 ± 0.04	0.27	0.30
CGN	163	5.0	-20.1	-1.90 ± 0.04	0.29	0.29
UDS	276	4.9	-21.2	-1.90 ± 0.03	0.39	0.36

Table 2. Sample properties of galaxies from the four field analysed in this work (column 1). Column 2 lists the number of galaxies selected from each field. Columns 3–5 list the mean redshift, mean absolute magnitude, and the mean UV slope and its standard error. The final two columns list the mean error on the β measurement for an individual galaxy from the data and, for comparison, the simulations described in Section 4.

troscopic redshifts from Vanzella et al. (2008), and three with spectroscopic redshifts from Popesso et al. (2009) and Balestra et al. (2010). All 26 spectroscopic redshifts lie in the range $4.4 < z_{\text{spec}} < 5.6$, in excellent agreement with our $4.5 < z_{\text{phot}} < 5.5$ defined selection window. The photometric redshift accuracy is $\sigma[(z_{\text{spec}} - z_{\text{phot}})/(1 + z_{\text{spec}})] = 0.032$. Of our 112 CGS galaxies, 110 have also been studied by Dahlen et al. (2014, in prep.), who compared photometric redshift estimates from 11 different codes (Dahlen et al. 2013). Comparing to their results, we estimate that our lower-redshift interloper contamination fraction is around 5%. This is due to our inclusion of candidates with good high-redshift solutions, but which are only marginally distinguished from lower-redshift solutions. Fortunately the exclusion of these potential contaminants, which span a wide range of luminosities, does not affect the results that follow.

2.6 Measuring M_{1500}

Absolute UV magnitudes ($M_{UV} = M_{1500}$) were determined using a top-hat filter centred on 1500 Å in the rest-frame of the best-fitting SED template. To make our results comparable to other studies, the absolute magnitudes were boosted by around 0.2 mag, to account for the still incomplete enclosure of a PSF’s flux within our already large apertures. The luminosity distribution of our complete sample is shown in the top panel of Fig. 4. By design, this definition of M_{1500} does not include any correction for dust attenuation; it is the ‘observed’ absolute magnitude. This choice serves to avoid imparting an artificial colour–magnitude relation onto the data. For instance, the simplest correction would be to follow the Meurer et al. (1999) relation, which yields a dust attenuation correction A_{1600} based on the observed colour β : a low β suggests a small A_{1600} , so only a minimal correction is required on M_{1500} . While less direct, the same must be true if the dust attenuation is estimated by SED fitting, since the SED fit has essentially no other information from which to constrain the dust attenuation. Thus, had we adopted such a correction, red galaxies would be luminosity-corrected by a larger degree than blue galaxies at the same *observed* luminosity. Under the null hypothesis that M_{1500} and β are uncorrelated, or are only weakly correlated, the result is to skew the observed β vs. $M_{1500, \text{obs}}$ space, generating a fake dependency of β on $M_{1500, \text{corrected}}$.

3 MEASUREMENT OF UV SLOPES AND THE COLOUR–MAGNITUDE RELATION

In this section, we use our sample to constrain the colour–magnitude relation at $z \approx 5$.

3.1 Measuring the UV slope

Following Rogers et al. (2013) and Dunlop et al. (2013), we fitted for β at fixed redshift using a variety of power-law SED models. Adopting the photometric redshift z_{phot} derived at the selection stage, we re-ran LEPHARE treating z_{phot} as a spectroscopic redshift. The models were pure power-law models ($f_{\lambda} \propto \lambda^{\beta}$), truncated shortward of 912 Å, with no dust attenuation allowed. Attenuation due to the IGM is provided by LEPHARE. An example of these power-law fits is shown in Fig. 3. Minimum photometric errors of 10% are still used; removing these has little effect, with β changing by < 0.1 . The fitting procedure yields, for each object, a distribution $\chi^2(\beta_i)$ for a finely gridded set of templates $-8 < \beta_i < +5$. The best-fitting UV slope was taken to be that which minimized χ^2 , and the error was obtained by finding the values of β for which $\Delta\chi^2 = 1$ from the χ^2 minimum (where the χ^2 distribution is minimized over the SED normalisation). This procedure excludes the colour uncertainty induced by the photometric redshift uncertainty, but in the vast majority of cases this is minimal since the Lyman break falls a good way blueward of the β -measuring wavebands (see Fig. 2). In the UDS, we used the two z -bands, J and H to fit β . In the CANDELS fields, we used z_{850} , $Y_{098|105}$, J_{125} , and H_{160} , and in the HUDF, we used z_{850} , Y_{105} , J_{125} , JH_{140} , and H_{160} .

3.2 Linear fits to the colour–magnitude relation

A simple linear fit to the colour–magnitude distribution of the combined galaxy sample, with each galaxy weighted by its colour error $1/\delta\beta$, yields a slope

$$\frac{d\beta}{dM} = -0.12 \pm 0.02,$$

and a zeropoint of

$$\beta(M_{1500} = -19.5) = -1.93 \pm 0.03.$$

Fig. 4 shows this relation fitted to the sample. These values are in excellent agreement with the results of B13, who used space-based data alone. This parametrization does differ from the results of F12, who found a much weaker correlation, but our relation is still statistically consistent with their binned data points. As F12 illustrated, restricting their faintest bin to galaxies from the HUDF alone (which moves the bin fainter) also yields a stronger relation, bringing it further in line with this work.

Following the suggested existence of a piecewise-linear relation by B13, whereby $d\beta/dM$ is steeper at the bright end (e.g. $M_{1500} < -18.9$) than the faint end, Fig. 4 also shows their two-component broken linear fit. Our data show no clear evidence to support the broken power-law CMR (see the binned mean results in Fig. 4). Fixing the break at B13’s suggestion of $M_{1500} = -18.9$ and fitting the bright- and faint-end slopes of our data does yield values similar to those reported by B13; of -0.16 ± 0.02 and -0.04 ± 0.12 respectively. However, given the uncertainties, there is clearly

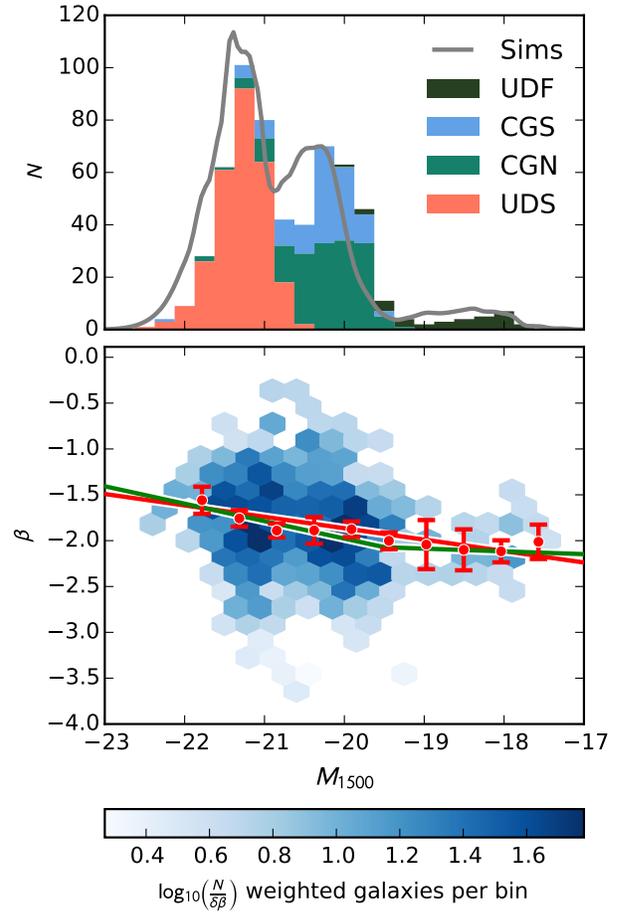


Figure 4. *Top:* the luminosity distribution of the sample used to constrain the colour–magnitude relation, shown as a stacked histogram, split by field. The grey line shows the expected distribution based on the $z = 5$ luminosity function and the size and selection function of each field (see later simulations). *Bottom:* the density map shows the combined sample in colour–magnitude space. The red circles are binned means, and their error bars are $2\times$ the standard error on the mean in each bin (inflated for visibility). These error bars are driven by photometric scatter, so do not directly constrain any intrinsic variation that may exist in the colours. The red line is a fit to all the individual galaxies, with slope -0.12 . Both the density map and fit are weighted by the uncertainty, $\delta\beta$, on each galaxy’s β colour. The green line is the two-component piecewise-linear fit reported by Bouwens et al. (2013). Their single linear relation is in near-perfect agreement with ours, so is not shown.

no significant evidence for the CMR being non-linear. Moreover, even the qualitative nature of this result is heavily dependent on the exact choice of the break luminosity. For example, floating the break luminosity fails to yield any meaningful constraint, and fixing the break instead at $M_{1500} = -19.5$ results in a steeper faint- than bright-end CMR gradient. In summary, we confirm the existence of a significant CMR with our sample. However at least in this redshift window (B13 studied a range of redshifts) we lack strong evidence to either confirm or refute the existence of a characteristic luminosity at which the colour–magnitude relation changes gradient.

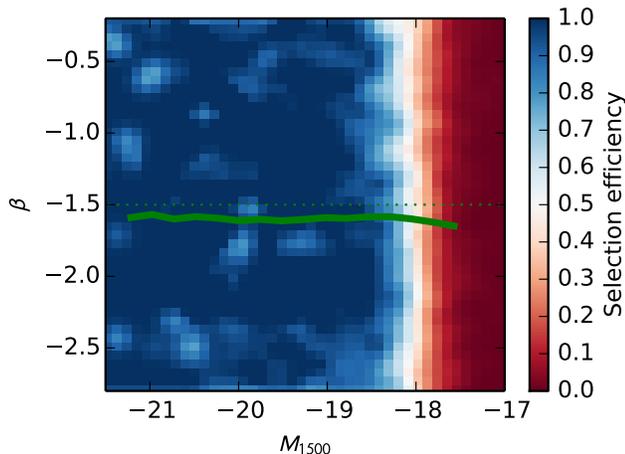


Figure 5. The selection efficiency (recovered fraction of total input galaxies) is shown in bins of colour–magnitude space for the HUDF simulation. The completeness limit lies at a fairly uniform magnitude across the range of β s, meaning no colour bias exists in any luminosity bin. As an example we show the worst-case scenario, picking a $\beta = -1.5$ sample from the simulations (dotted green line) which, after observation, is measured as the biased solid green line. This line is typical for colours $\beta > -2$; the bias line is even flatter for $\beta < -2$. This small offset is (the worst) example of the systematic offsets from the simulations, $\Delta\beta \lesssim 0.1$, due to the way galaxy templates deviate from a perfect power law. The bias line is truncated at the faint limit of our sample.

4 IMAGE SIMULATIONS

In this section we describe the creation and injection of simulated galaxies into the images, and explain how their detection and selection efficiencies compare to the real galaxy sample.

4.1 Model galaxies

Simulated galaxies were defined on a grid spanning β , M_{UV} and z , in order that the selection probability could be computed for any galaxy with a given intrinsic redshift, luminosity and colour, and so that mappings could be made between intrinsic and observed values of β . The luminosity distribution from which they were drawn conforms to the simple redshift-evolving luminosity function of McLure et al. (2013, section 2.7). This was needed since, to simulate some scattering between redshift bins, we input galaxies in the redshift range $4 < z < 6$.

Each model’s SED and M_{UV} were derived from a BC03 (Bruzual & Charlot 2003) stellar population synthesis model with metallicity $0.2 Z_{\odot}$, a declining star formation history with $\tau = 0.03$ Gyr and a Chabrier (2003) IMF. The galaxy population was defined with a uniform distribution of galaxies in the β dimension, achieved by mapping β to pairs of stellar population age t and Calzetti et al. (2000) dust reddening $E(B-V)$ (see fig. 3 of Rogers et al. 2013). While the entire range of β could have been reproduced by modifying $E(B-V)$ at fixed t , the age evolution was imposed to provide a more physically motivated model at each β . Galaxies were allocated half-light radii $0.2 \leq r_{1/2} \leq 1.6$ kpc, according to the $z = 5$ size–luminosity relationship of Bouwens et al.

(2004) with a small scatter ($\sigma = 0.2$ kpc) around the relation. This was implemented in the simulations by broadening the PSF with an appropriate smoothing kernel. As discussed earlier, unknown sizes in this range, when convolved with the PSF and measured in our adopted apertures, imply errors on β of < 0.2 . We assume a wavelength-independent morphology over the fairly narrow rest-frame wavelength range of interest ($\lambda \sim 1300\text{--}3000$ Å).

4.2 Simulation pipeline

The model galaxies were inserted into the images, avoiding existing sources by use of the segmentation map. In the HUDF, 50 copies of the field were used to avoid excess crowding. The simulated galaxies were observed in the same way as in the real data. While no ‘stars’ were injected into the simulations, we performed the same star-rejection routine as for the data such that its effect on the selection efficiency could be determined.

4.3 Selection efficiency

Fig. 5 shows an example of the selection efficiency of galaxies in colour–magnitude space from our HUDF simulation. Reassuringly, considering our conservative approach, there is little preferential selection of galaxies at any colour: the completeness limits are β independent. Our simulations do however show evidence of some systematic offsets, although even in the worst-case scenario these are at the $\delta\beta \lesssim 0.1$ level (see Fig. 5). The remaining bias is due to a combination of aperture/size effects and variations in β which depend on the filter-set or measurement method used (see fig. 2 of Rogers et al. 2013).

4.4 Comparison to data

By multiplying the selection efficiency of the appropriate simulation by the area of each field and integrating down the adopted luminosity function model, a predicted luminosity distribution for our sample can be computed. For this, we adopt the $z = 5$ luminosity function determination of McLure et al. (2009). The prediction is shown as the grey line in Fig. 4, and agrees well with the combined luminosity distribution of the actual sample. The simulations also predict uncertainties on individual galaxy β measurements which are very similar to those found for the sample, as shown in Table 2.

5 INTRINSIC VARIATION

In this section we derive the *intrinsic* variation in colour across the sample, using the image simulations to decouple the observed β distributions into intrinsic variation and photometric scatter. We have approached this problem from three angles. First and most simply, we assumed that the observed β distribution is a convolution of two Gaussians: one representing the intrinsic colour distribution of the galaxy population, and one due to photometric scatter in some fiducial scenario. Second, by drawing realisations from the simulations according to various intrinsic distributions, we compared the data and simulations using an equal-variance test

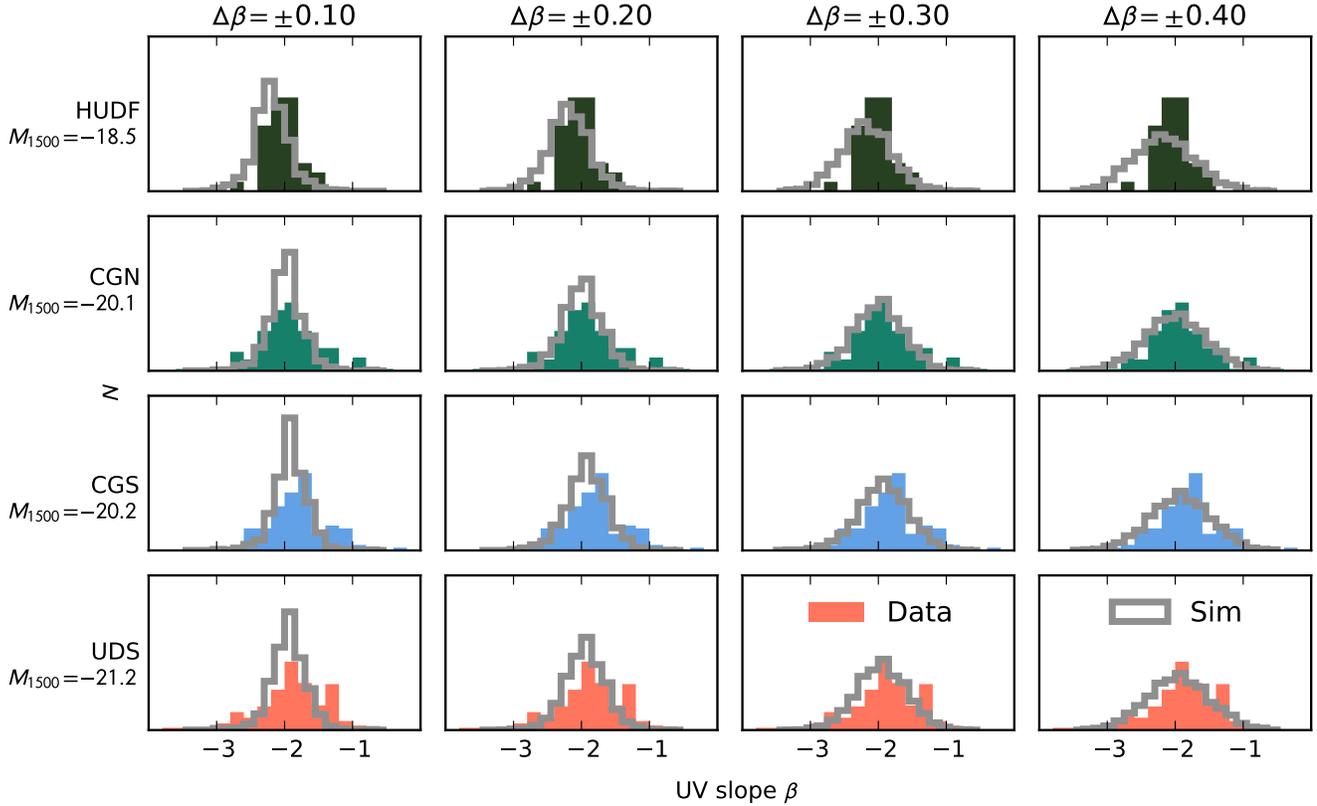


Figure 6. A comparison of the intrinsic scatter of faint galaxies (top panels, from the HUDF sample) through to brighter galaxies (lower panels from CGN, CGS, and the UDS). By comparing the distribution of colours in the data (coloured histograms) to those of the simulations (grey lines), which are designed to replicate the photometric scatter of the data, we can disentangle the intrinsic distribution of colours. Left to right, simulations of increasing intrinsic colour variation are shown superimposed onto the fixed data. For simplicity, we show only four possibilities here: $\Delta\beta \in \{0.1, 0.2, 0.3, 0.4\}$. The fields containing brighter galaxies are better represented by simulations with larger intrinsic colour scatter.

and maximized the probability that the data and realisation are from the same population. Third, the comparisons of the β distributions of the data and simulations were made by a full maximum-likelihood test. Each method relies on a comparison of the observed distribution to some subset of our simulations. A visual comparison of this is given in Fig. 6, where the observed β distribution in each field is shown alongside simulated distributions based on various intrinsic colour scatters. The intrinsic distribution is in all cases assumed to be Gaussian, and is parametrized by $\Delta\beta$, its standard deviation. We have tested a log-normal distribution and, like Castellano et al. (2012), find no convincing preference for it.

5.1 Measuring $\Delta\beta$: Gaussian assumption

The simplest estimate of the intrinsic distribution of colours is to assume that the observed distribution is a combination of two Gaussian distributions: one reflecting intrinsic variation and one due to photometric scatter. Under this assumption, the width of the intrinsic colour distribution is

$$\Delta\beta \approx \sqrt{\sigma_{\text{obs}}(\beta)^2 - \sigma_{\text{photo}}(\beta)^2}, \quad (1)$$

where $\sigma_{\text{photo}}(\beta)$ can be measured by looking at the simulated β distribution of a set of galaxies that were input with $\beta_{\text{input}} \approx \text{median}(\beta_{\text{data}})$. By relying on the varying average luminosity of galaxies in each field, we can make these comparisons along the luminosity baseline without the difficulties of combining the data and simulations of the different survey fields. To better populate the luminosity space, each field and simulation were split into three luminosity bins of equal occupancy. The results of this measurement are shown as \times marks in the first panel of Fig. 7. The colours of the brighter galaxies found in the UDS and CANDELS fields populate a much broader distribution than photometric scatter alone would predict: hence, they stem from a more intrinsically varied population.

5.2 Measuring $\Delta\beta$: equal-variance test

By testing against only a single simulation, the first test could not measure the uncertainty on $\Delta\beta$. So for the equal-variance test a grid of simulated β distributions, each according to a different $0 < \Delta\beta < 1$, was created for each field. In each case, the simulation was centred on the median β of the data. Centring the distribution in this way, rather than around the linear CMR, avoids making any *a priori* assumption about the shape of the CMR. Following Dun-

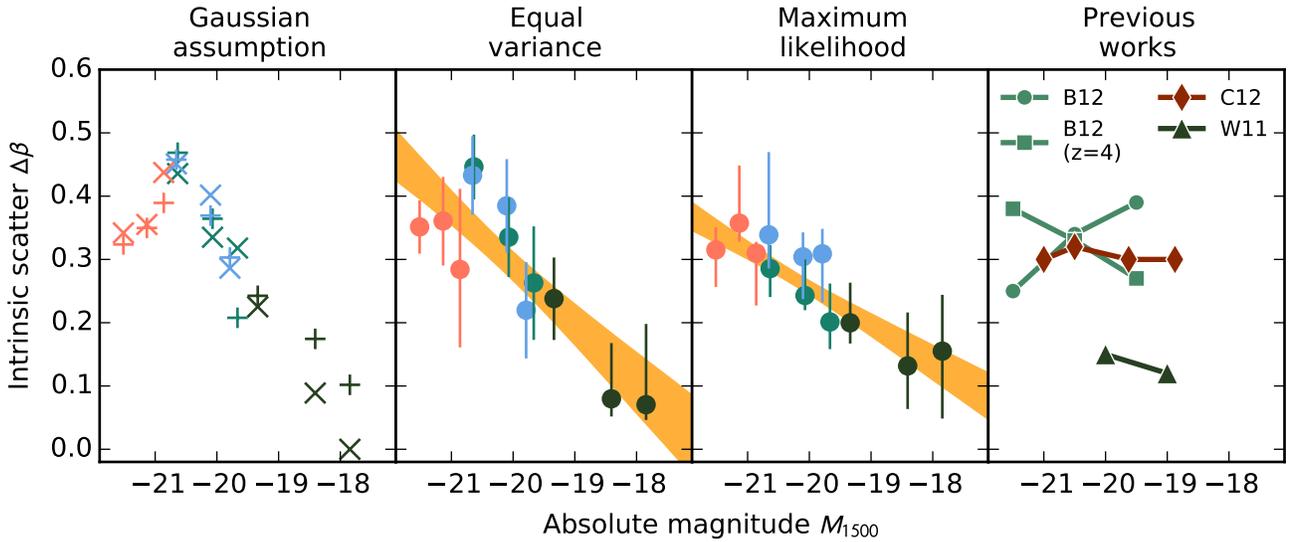


Figure 7. The width of the intrinsic colour distribution of galaxies at various luminosities. The first three panels relate to each of our three test methods, as denoted by the panel title and described in Section 5. In the first panel, \times marks show the results of our Gaussian assumption test, while $+$ marks denote our simulation-free check results. In all cases, each field contributes three bins of equal occupancy. The points are coloured by field, as in Fig. 6: salmon=UDS, blue=CGS, green=CGN, dark=HUDF. The error bars each enclose 68% of the total $p(\Delta\beta)$ where applicable. In each case, the yellow regions show the 1σ error limits of a linear fit to the measurements, weighted by the errors on $\Delta\beta$. In all three tests, brighter galaxies are drawn from a more varied intrinsic population than fainter galaxies. The final panel shows measurements drawn from the literature over narrower dynamic ranges, from Bouwens et al. (2012, B12), Wilkins et al. (2011, W11), and Castellano et al. (2012, C12).

lop et al. (2013), we used a non-parametric test to assess the probability that each bin’s simulated distribution and data arose from the same population. Since for this measurement the mean intrinsic value of β is not of interest, the Brown & Forsythe (1974) test for equal variances, rather than a K-S test, was adopted. In this manner, a probability density function $p(\Delta\beta)$ was created for each of the luminosity bins. By finding the maxima of $p(\Delta\beta)$, and the regions of $\Delta\beta$ enclosing 68% of p , robust measurements for $\langle\Delta\beta\rangle$ and its uncertainty were found. Since the actual subset of simulated galaxies returned is random (in order to approximately populate a Gaussian in intrinsic β), we averaged the best value and uncertainties over multiple realisations at each $\Delta\beta$. The variation in $\langle\Delta\beta\rangle$ between realisations was always much smaller than the error derived from $p(\Delta\beta)$. The results are shown in the second panel of Fig. 7, and are in excellent agreement with the first test.

5.3 Measuring $\Delta\beta$: maximum-likelihood test

For this final test, simulations were created for a grid of $\Delta\beta$ as before. However in this case, each luminosity bin’s data and simulation were binned in β to form histograms with N_B bins spanning $-4 < \beta < 0$. We compared the simulated and actual histograms of β and maximized

$$\mathcal{L} = \prod_i \frac{s_i(\Delta\beta, \mu_\beta)^{d_i} \exp[-s_i(\Delta\beta, \mu_\beta)]}{d_i!}, \quad (2)$$

where s_i and d_i denote the simulated and actual (data) number of galaxies in the i th bin, and s_i depends on $\Delta\beta$ and the centre of the Gaussian distribution from which the simulation is drawn, μ_β . In each case the normalisation of the

histograms was fixed to ensure $\sum_i d_i / \sum_i s_i = 1$, and the central colour μ_β of the simulation was allowed to float and was marginalised over. The simulation realisation was also marginalised out. The maximum-likelihood (ML) results are shown in the third panel of Fig. 7, and are in good agreement with both of the previous tests.

5.4 Measuring $\Delta\beta$: simulation-free check

The three methods above all rely on image simulations. To avoid complete reliance on these simulations, we repeated the Gaussian assumption test in a simulation-free way. For each luminosity bin, we created z - to H -band observer-frame photometry for test galaxies, all with $\beta = -2$ but using the M_{UV} distribution of the data. Using the median image depths given in Table 1, we perturbed the photometry with appropriately scaled Gaussian random numbers. The UV slope was then measured using a simple power-law SED fit to the generated photometry, using the same set of filters as for the data. This process yields a measurement of the expected photometric scatter on β which depends only on the image depths (and an assumed fiducial value of $\beta = -2$). Using equation 1 again to compare to the observed data gives excellent agreement with our earlier Gaussian assumption test. These simulation-free results are shown as the $+$ marks in the first panel of Fig. 7.

5.5 Discovery of significant colour scatter

In all but the faintest bins considered, each test shows significant evidence of intrinsic colour scatter, $\Delta\beta > 0$. To

quantify the significance of this, we compared a null hypothesis, where all luminosity bins contain intrinsic colour variation in line with the faintest bin ($\Delta\beta = 0.1$), to an alternative hypothesis, where $\Delta\beta$ grows with luminosity ($\Delta\beta = s \times M_{1500} + c$), motivated by the trend of the ML results. Using a maximum-likelihood estimated linear relation, the likelihood ratio test statistic

$$D = -2 \ln \left(\frac{\mathcal{L}(\Delta\beta = 0.1)}{\mathcal{L}_{\text{linear}}} \right) \approx 150. \quad (3)$$

Since the linear fit has two extra degrees of freedom (s, c), this equates to a significance of $\approx 12\sigma$. We have therefore discovered very significant intrinsic colour variation in our sample of $z \approx 5$ galaxies.

5.6 A colour-scatter-magnitude relation?

In all of the tests considered, there is clear evidence for the intrinsic variation in colour being not only non-zero, but increasing at brighter luminosities (just as the average $\langle\beta\rangle$ is redder for brighter galaxy populations). To measure the significance of the trend quantitatively, linear fits, weighted by the uncertainties, were made for both the equal-variance test and the ML test results. These resultant ‘colour-scatter-magnitude relations’ are shown as the yellow regions in Fig. 7. Formally, the slopes of the linear fits differ from 0 by $\approx 4.8\sigma$ and 5.1σ for the equal-variance and ML tests respectively. As a further test of the significance, we computed a likelihood ratio of two hypotheses: that the colour variation is a constant at all luminosities $\Delta\beta(M) = 0 \times M_{1500} + c = c$, or that it grows with increasing luminosity $\Delta\beta(M) = s \times M_{1500} + c$. Taking the maximum likelihood under each hypothesis, the likelihood ratio test statistic

$$D = -2 \ln \left(\frac{\mathcal{L}_{\text{flat}}}{\mathcal{L}_{\text{relation}}} \right) = 6.7, \quad (4)$$

which, with the slope s being one degree of freedom, makes the growing relation more likely than the flat relation at the 2.6σ level. This is a more conservative approach and result than the linear-fit derived significance, but the two approaches are in reasonable agreement. In all cases, the existence of a scatter-magnitude relation is confirmed at better than 2.5σ significance.

5.7 Effect of varying signal-to-noise thresholds

The four fields from which our sample has been assembled were treated with different SNR thresholds when selecting galaxies. An obvious concern is that this choice may affect our results. However, we can be confident this is not the case for a number of reasons. Firstly the photometric uncertainty on measuring β does not scale directly with the z -band detection threshold, since $\delta\beta$ depends on the noise in all the bands from which β is measured. Second, the simulated galaxies inhabit the same noise as the data, since we injected sources into the real images rather than merely perturbing a photometry catalogue according to some noise parametrization. As such they are subject to the same photometric scattering and SNR cuts as the data, so the comparison of data to simulations is ‘fair’. Third, the trend for brighter bins to show more intrinsic scatter is visible within each of the

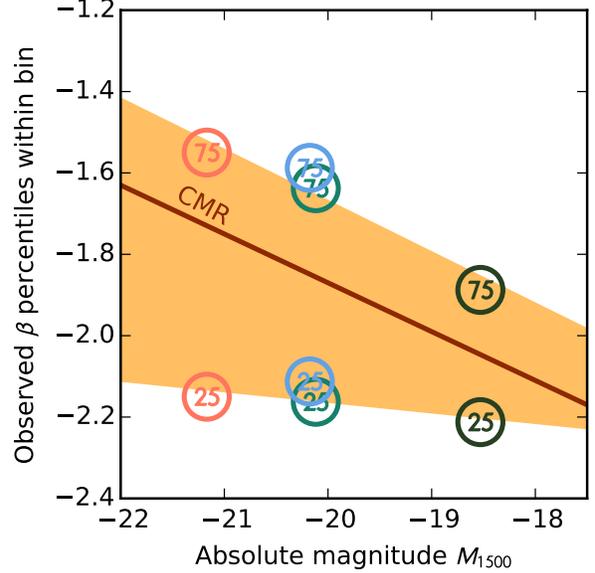


Figure 8. The 25th and 75th percentiles of β in each field are shown by coloured circles (marked 25 and 75, respectively). The edges of the shaded region are linear fits to the points, weighted by errors determined via bootstrapping. The region thus denotes a simple measure of how the scatter in β increases to bright luminosities as in Fig. 7. Here however, it is clear that while the bluest galaxies are similarly blue at all luminosities, the redder average colours and greater colour scatter at bright M_{1500} are driven by establishing a redder sub-population in the brighter bins. The red line is the average colour-magnitude relation from Fig. 4. Otherwise, the field colouring matches that of Fig. 6: salmon=UDS, blue=CGS, green=CGN, dark=HUDF.

fields, as well as between fields. Finally, if the simulations were systematically underestimating $\delta\beta$ then lower-SNR selected fields (i.e. the HUDF) would require more, not less, intrinsic scatter to match the observed distribution than the higher-SNR selected fields.

5.8 Comparison to previous works

Estimates of $\Delta\beta$ at $z = 4$ and $z = 5$ have been published by Castellano et al. (2012), Bouwens et al. (2012) and Wilkins et al. (2011), but each covers only a narrow range of luminosity. Individually the results show little evidence of luminosity trends; in particular the trend reverses between the $z = 4$ and $z = 5$ samples of Bouwens et al. (2012), suggesting that the luminosity dependence of $\Delta\beta$ is poorly constrained in both cases. The fourth panel of Fig. 7 includes these previous estimates alongside our results.

5.9 Asymmetric colour scatter

Is the colour-magnitude relation itself merely a product of the scatter-magnitude relation? For this to be true, the scatter must expand in such a way that brighter bins include more red galaxies than fainter bins without many more blue galaxies. Fig. 8 offers some evidence that this may be the case. Between the four fields, the 25th percentiles of the observed β distribution grow no bluer ($\beta \approx -2.1$) from

the faintest HUDF galaxies to the brightest UDS galaxies. Meanwhile the 75th percentiles redden from $\beta = -1.9 \rightarrow -1.5$. The bluest galaxies are not any bluer in the brightest bins, but the reddest galaxies are much redder. A distribution which grows to the red would be in line with the ‘blue-ridge–red-tail’ distribution of blue-sequence field galaxies in the $z \lesssim 3$ study of Labbé et al. (2007).

6 CONCLUSIONS

The rest-frame UV colours of high-redshift galaxies provide a probe of the metallicity and dust conditions within which their stellar populations are growing. While recent attention has been focused at $z \geq 7$, moving just 400 Myr later to $z = 5$ has allowed us take a more conservative approach to the detection, selection, and colour analysis of a sample of $z \approx 5$ Lyman-break galaxies. Detected mostly at $\text{SNR} > 8\sigma$, and with four or five imaging bands spanning the rest-frame UV, we have been able to robustly determine the UV continuum slope β of each of 584 $z \approx 5$ LBGs (typically to better than $\delta\beta \lesssim 0.4$). Crucially these galaxies span a factor of 100 in luminosity, allowing us to constrain the colour–magnitude relation. Comparing our samples to closely representative image simulations of mock galaxies, we have also disentangled the intrinsic variation in colour at each magnitude from the photometric scatter. Our findings can be summarised as follows.

(i) A linear colour–magnitude relation, whereby brighter galaxies are redder than fainter galaxies by $d\beta/dM_{\text{UV}} = -0.12 \pm 0.02$ provides a good fit to our data.

(ii) The data show no convincing evidence either for or against a piecewise-linear relation, whereby galaxies cease to get bluer with decreasing luminosity below some point, as had been suggested elsewhere.

(iii) For the first time over a wide range of luminosities at high redshift, we have discovered significant (12σ) evidence for *intrinsic colour variation* within the LBG population.

(iv) This intrinsic colour variation is significantly larger (2.5σ) in high-luminosity bins than low-luminosity bins, after accounting for photometric scatter using our detailed simulations. This result was confirmed by multiple statistical tests, as well as by a final check which was not reliant on our simulations in any way.

(v) The luminosity-dependent colour scatter and average colour–magnitude relation appear to be due to the evolution of bright, red galaxies. This appearance of bright, red galaxies coincides with a seemingly luminosity-independent blue floor: in each luminosity bin, the 25th percentiles of colour are always $\beta \approx -2.1$, while the 75th percentiles grow redder from $\beta = -1.9 \rightarrow -1.5$ over $M_{1500} = -18.5 \rightarrow -21.2$.

The rest-frame UV colour β is dependent on all parameters of the stellar population, but is particularly sensitive to light-weighted age and dust attenuation. Our measurements of lower-luminosity galaxies, $\langle\beta(M_{1500} = -18)\rangle \approx -2$, are not so blue as to require dust-free stellar populations. However the lack of intrinsic scatter there ($\Delta\beta < 0.2$) shows that, if the galaxies are dust reddened, it is by similar amounts for all galaxies in that bin. Similarly the light-weighted ages of galaxies in the faint population must be fairly similar to one another, and < 100 Myr. This can be interpreted as all

of those galaxies undergoing intense present or recent star-formation. Comparatively, the average brighter galaxy, with colour $\langle\beta(M_{1500} = -21)\rangle \approx -1.7$, must have built an older stellar population, or have higher metallicity or dust reddening. However at all luminosities a quarter of galaxies have colours bluer than $\beta \lesssim -2.1$, so even at $M_{1500} = -21$ the low reddening, young galaxies remain common.

These observations are consistent with at least two simple scenarios, between which our current observations cannot differentiate: the build up of dust as the galaxy grows and brightens, with some galaxies oriented such that the UV light escapes with less than average reddening; or stochastic star-formation histories, where faint galaxies are always currently star-forming but brighter galaxies are observed during various phases of their star-formation duty cycle.

ACKNOWLEDGMENTS

This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013), NumPy and SciPy (Oliphant 2007), Matplotlib (Hunter 2007), IPython (Pérez & Granger 2007), NASA’s Astrophysics Data System Bibliographic Services and has benefitted from the SpeX Prism Spectral Libraries, maintained by Adam Burgasser at <http://pono.ucsd.edu/~adam/browndwarfs/spexprism>.

ABR and EFCL acknowledge the support of the UK Science & Technology Facilities Council. RJM acknowledges the support of the European Research Council via the award of a Consolidator Grant (PI McLure). JSD, RAAB, and PD acknowledge the support of the European Research Council via the award of an Advanced Grant to JSD, who also acknowledges the support of the Royal Society via a Wolfson Research Merit award and the contribution of the EC FP7 SPACE project ASTRODEEP (Ref. No: 312725). This work is based in part on observations made with the NASA/ESA *Hubble Space Telescope*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

References

- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
- Astropy Collaboration et al., 2013, A&A, 558, A33
- Balestra I. et al., 2010, A&A, 512, A12
- Beckwith S. V. W. et al., 2006, AJ, 132, 1729
- Bertin E., Arnouts S., 1996, A&AS, 117, 393
- Bouwens R. J., Illingworth G. D., Blakeslee J. P., Broadhurst T. J., Franx M., 2004, ApJ, 611, L1
- Bouwens R. J. et al., 2012, ApJ, 754, 83
- Bouwens R. J. et al., 2013, preprint (arXiv: 1306.2950)
- Bouwens R. J. et al., 2010, ApJ, 708, L69
- Brown M. B., Forsythe A. B., 1974, Journal of the American Statistical Association, 69, pp. 364
- Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
- Castellano M. et al., 2012, A&A, 540, A39

- Chabrier G., 2003, *PASP*, 115, 763
Dahlen T. et al., 2013, *ApJ*, 775, 93
Dayal P., Ferrara A., 2012, *MNRAS*, 421, 2568
Dunlop J. S., McLure R. J., Robertson B. E., Ellis R. S., Stark D. P., Cirasuolo M., de Ravel L., 2012, *MNRAS*, 420, 901
Dunlop J. S. et al., 2013, *MNRAS*, 432, 3520
Ellis R. S. et al., 2013, *ApJ*, 763, L7
Finkelstein S. L., Papovich C., Giavalisco M., Reddy N. A., Ferguson H. C., Koekemoer A. M., Dickinson M., 2010, *ApJ*, 719, 1250
Finkelstein S. L. et al., 2012, *ApJ*, 756, 164
Furusawa H. et al., 2008, *ApJS*, 176, 1
Giavalisco M., 2002, *ARA&A*, 40, 579
Grogin N. A. et al., 2011, *ApJS*, 197, 35
Hunter J. D., 2007, *Computing In Science & Engineering*, 9, 90
Ilbert O. et al., 2006, *A&A*, 457, 841
Illingworth G. D. et al., 2013, *ApJS*, 209, 6
Koekemoer A. M. et al., 2013, *ApJS*, 209, 3
Koekemoer A. M. et al., 2011, *ApJS*, 197, 36
Labbé I. et al., 2007, *ApJ*, 665, 944
Leitherer C. et al., 1999, *ApJS*, 123, 3
McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, *MNRAS*, 395, 2196
McLure R. J. et al., 2013, *MNRAS*, 432, 2696
Meurer G. R., Heckman T. M., Calzetti D., 1999, *ApJ*, 521, 64
Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713
Oliphant T. E., 2007, *Computing in Science & Engineering*, 9, 10
Overzier R. A. et al., 2008, *ApJ*, 673, 143
Pérez F., Granger B. E., 2007, *Computing In Science & Engineering*, 9, 21
Popesso P. et al., 2009, *A&A*, 494, 443
Reddy N. et al., 2012, *ApJ*, 744, 154
Reddy N. A., Erb D. K., Pettini M., Steidel C. C., Shapley A. E., 2010, *ApJ*, 712, 1070
Rogers A. B., McLure R. J., Dunlop J. S., 2013, *MNRAS*, 429, 2456
Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, *ApJ*, 588, 65
Steidel C. C., Adelberger K. L., Giavalisco M., Dickinson M., Pettini M., 1999, *ApJ*, 519, 1
Vanzella E. et al., 2008, *A&A*, 478, 83
Wilkins S. M., Bunker A. J., Stanway E., Lorenzoni S., Caruana J., 2011, *MNRAS*, 417, 717
Windhorst R. A. et al., 2011, *ApJS*, 193, 27

Constraints on the star-formation rate of $z\sim 3$ LBGs with measured metallicity in the CANDELS GOODS-South field

M. Castellano¹, V. Sommariva¹, A. Fontana¹, L. Pentericci¹, P. Santini¹, R. Amorin¹, J. L. Donley², J. S. Dunlop³, H. C. Ferguson⁴, F. Fiore¹, A. Galametz¹, E. Giallongo¹, A. Grazian¹, Yicheng Guo⁵, Kuang-Han Huang^{6,7}, R. Maiolino^{8,9}, R. J. McLure³, D. Paris¹,
D. Schaerer^{10,11}, P. Troncoso¹², and E. Vanzella¹³

¹ INAF - Osservatorio Astronomico di Roma, Via Frascati 33, I-00040, Monteporzio, Italy.

² Los Alamos National Laboratory, Los Alamos, NM, USA

³ Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK.

⁴ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA.

⁵ UCO/Lick Observatory, Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA, USA.

⁶ Department of Physics, University of California, Davis, CA 95616, USA.

⁷ Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA.

⁸ Cavendish Laboratory, University of Cambridge, 19 J.J. Thomson Avenue, Cambridge CB3 0HE, UK.

⁹ Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

¹⁰ Observatoire de Genève, 51 chemin des Maillettes, CH 1290 Versoix, Switzerland.

¹¹ CNRS, IRAP, 14 Avenue E. Belin, 31400 Toulouse, France.

¹² Astronomisches Institut, Ruhr-Universität Bochum, Universitätsstraße 150, D-44780, Bochum, Germany.

¹³ INAF-Osservatorio Astronomico di Bologna, Via Ranzani 1, I-40127 Bologna, Italy.

Received... accepted ...

ABSTRACT

Aims. We aim at constraining the assembly history of high-redshift galaxies and the reliability of UV-based estimates of their physical parameters from an accurate analysis of a unique sample of $z \sim 3$ Lyman break galaxies (LBGs).

Methods. We analyse 14 LBGs at $z \sim 2.8-3.8$ constituting the only sample where both a spectroscopic measurement of their metallicity and deep IR observations (CANDELS+HUGS survey) are available. Fixing the metallicity of population synthesis models to the observed values, we determine best-fit physical parameters under different assumptions on the star-formation history and considering also the effect of nebular emission. For comparison we determine the UV slope of the objects, and use it to estimate their SFR_{UV} by correcting the UV luminosity under standard assumptions.

Results. A comparison between SFR obtained through SED-fitting (SFR_{fit}) and the SFR_{UV} shows that the latter are underestimated by a factor 2-10, regardless of the assumed SFH. Other SFR indicators (radio, far-IR, X-ray, recombination lines) coherently indicate SFRs a factor 2-4 larger than SFR_{UV} and in closer agreement with SFR_{fit} . Such discrepancy is due to the assumption of solar metallicity in the usual $\beta - A_{1600}$ conversion factor. We propose a refined relation, appropriate for sub-solar metallicity LBGs: $A_{1600} = 5.32 + 1.99 * \beta$. This relation reconciles the dust-corrected UV with the SED-fitting and the other SFR indicators. It also implies a revision by a factor ~ 2 of the global SFRD: $\approx 0.37 M_{\odot}/yr/Mpc^3$. We find very young best-fit ages (10-500 Myrs) for all our objects. From a careful examination of the uncertainties in the fit and the amplitude of the Balmer break we conclude that there is little evidence for the presence of old stellar population in at least half of the LBGs in our sample, suggesting that these objects are probably caught during huge star-formation burst, rather than being the result of a smooth evolution.

Key words. Galaxies: distances and redshift - Galaxies: evolution - Galaxies: high redshift

1. Introduction

Lyman Break Galaxies (LBG) represent by far the most numerous population of galaxies that we are able to observe in the early Universe. Their statistical distributions are becoming progressively better and better constrained, most notably their UV luminosity function which has been determined from $z \approx 3$ up to $z \approx 8-9$ (e.g. Bouwens et al. 2007; Reddy & Steidel 2009; Castellano et al. 2010; Grazian et al. 2011; McLure et al. 2013). In turn, the debate about their physical properties and about how these properties change with redshift is more active. In particular, estimates of their dust extinction are needed to convert the UV luminosity density into a star formation rate density (SFRD), and to constrain the amount of obscured star formation occurring in different systems at high redshift. An accurate determination of dust extinction in high-redshift galaxies is also fundamental to enable a proper comparison between predictions from galaxy evolution models and observations, so as to improve our understanding of the earliest stages of galaxy formation (e.g. Lacey et al. 2011; Somerville et al. 2012; Kimm & Cen 2013).

To investigate the dust content of LBGs, great attention has been devoted to the study of the slope of the UV continuum (e.g. Castellano et al. 2012; Bouwens et al. 2012; Finkelstein et al.

2012; Dunlop et al. 2013) which is mainly determined by dust absorption, but also affected by other physical parameters, above all metallicity and age. Adopting the (reasonably accurate) assumption that the spectrum of LBGs between $\lambda \simeq 1500\text{\AA}$ and $\lambda \simeq 2800\text{\AA}$ can be represented by a simple power-law $f_\lambda = \lambda^\beta$, these works have shown that high-redshift LBGs are generally found to have “blue” ($\beta \sim -2$) slopes. Despite the remaining discrepancies among different works on the dependency of UV slopes on redshift and UV luminosity, these results have been univocally interpreted as an indication of relatively low dust obscuration by converting the observed β into extinction assuming standard stellar populations.

A thorough physical interpretation of these results remains however an open problem because of the intrinsic observational degeneracies. In fact, for a given extinction both lower metallicity stars and young ages are responsible for bluer UV slopes, while the contribution from nebular continuum produces a reddening of β . As recently noted by Wilkins et al. (2013) on the basis of galaxy formation models, any variation with redshift of the above factors can introduce systematic biases in the computation of dust extinction and of the corrected SFRD. Unfortunately, photometric data alone do not allow to determine how different properties shape the observed β . In particular, while deep IR photometry leads to tighter constraints on the age of the stellar populations, stellar metallicity remains very poorly constrained even in the deepest photometric dataset.

Similarly, it is still somewhat unclear how to place LBGs in a broad scenario of galaxy evolution, establishing the typical star-formation history (SFH hereafter) that led to their observed properties. Specifically, it would be important to establish whether LBGs that we observe at different redshifts sample the same population observed at different epochs, while they assemble their stellar mass in a smooth, “secular” history of star-formation, or if their SFH is more episode-driven and how it is related to possible phases of intense, dust-obscured star-formation. In the latter case, individual galaxies might move in and out of the LBG selection criteria, and/or in different positions of the UV luminosity function over cosmic time. These questions can be in principle investigated by studying the Spectral Energy Distribution (SED) of LBGs at various redshifts. However, several papers in the past have analyzed the SEDs of LBGs showing that, while stellar mass is reasonably well established, their age and SFH is more difficult to determine because of many degeneracies (e.g. Reddy et al. 2012; Curtis-Lake et al. 2013). These degeneracies result from a set of uncertainties in the current observations of LBGs: photometry with high S/N is both scanty and difficult, especially in the crucial IR region where the contribution of previous generations of stars is appreciable; metallicity is generally not known, even for galaxies with spectroscopic redshift; and the large fraction of current LBG samples even lack spectroscopic redshifts, leading to additional uncertainties in the k -corrections and distance modulus. All these factors lead to larger uncertainties in the SED fitting of LBG samples, that prevent us from constraining both their dust content and SFH.

In this paper we take a different approach. Rather than selecting a complete sample of LBGs, we identify the small set of LBGs that have extraordinarily well constrained properties, and perform a

stringent, state-of-the-art SED fitting on them. In particular, we analyse a unique sample of galaxies at $z > 2.9$ for which not only redshift but also metallicity (either stellar or gas-phase) has been measured, and exquisite deep photometry is available in all bands, from the optical to the crucial IR. At this purpose we have identified 14 galaxies at $2.9 \lesssim z \lesssim 3.8$ in the GOODS-S field with measured metallicity (from deep spectroscopic surveys like AMAZE and GMASS) and we exploit the unique CANDELS dataset including observations from the U band to IRAC mid-IR, to perform SED fitting while fixing the metallicity of population synthesis models to the measured one. In particular, the availability of CANDELS WFC3 observations, of the deep K-band data of the HUGS-CANDELS survey (Fontana et al. in prep.) and of IRAC/SEDs data allows us to accurately sample the Balmer break at these redshifts to constrain the age of the objects in our sample. The available multi-wavelength data covering the rest-frame UV are also exploited to estimate extinction from the slope β of the continuum under commonly adopted assumptions. Despite our sample is not complete in a statistical sense, our galaxies are representative of relatively luminous LBGs, and we will show that the properties that we derive can provide useful information to settle the questions mentioned above.

The plan of the paper is the following. In Sect. 2 we present our sample and summarise the available spectroscopic and photometric information. In Sect. 3 we discuss the SED-fitting estimates of their physical properties, obtained by varying assumptions on the SFH and on the contribution of nebular emission. In particular we compare the resulting $E(B-V)$ and SFRs to estimates obtained from UV slope and luminosity, and discuss independent constraints on the SFR from X-ray and FIR data. We exploit the results of our analysis to define a more appropriate $\beta - A_{1600}$ conversion equation which is then used to compute the SFRD at $z \sim 3$ (Sect. 4). A detailed discussion of the age of the galaxies in our sample is presented in Sect. 5. Finally, a summary is given in Sect. 6.

Throughout the paper, observed and rest-frame magnitudes are in the AB system, and we adopt the Λ -CDM concordance model ($H_0 = 70 \text{ km/s/Mpc}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$).

2. Objects at $z \sim 3-4$ with metallicity from deep spectroscopy

2.1. The sample

We consider objects at $z \approx 2.8-3.8$ in the GOODS-S field for which a spectroscopic estimate of their stellar and/or gas phase metallicity is available. The sample includes seven objects with measured stellar metallicity from UV absorption features: the subsample of four LBGs at $3.4 \lesssim z \lesssim 3.8$ in the CDFS from the AMAZE survey (Maiolino et al. 2008) presented in Sommariva et al. (2012) (S12 hereafter) and three galaxies at $2.9 \lesssim z \lesssim 3.4$ from the public release of the GMASS survey (Kurk et al. 2013) whose stellar metallicity estimates are presented here for the first time. Stellar metallicities are measured from the equivalent width of photospheric absorption lines sensitive to metallicity and independent to the other stellar parameters (such as age and IMF). In particular, the 1460\AA and 1501\AA features (introduced in S12), and the stellar features at 1370\AA , 1425\AA and at 1978\AA proposed by Rix et al. (2004) and recalibrated by S12 with updated stellar libraries.

Table 1. Objects at $z \sim 3-4$ with metallicity from deep spectroscopy

ID	ID_{CANDELS}	R.A.	Dec.	Redshift	Z/Z_{\odot}^a	References for Z/Z_{\odot}
CDFS-2528	2405	53.1898	-27.8925	3.689	$0.30^{+0.11}_{-0.08}$ (G)	2
CDFS-4417	5001	53.0972	-27.8657	3.470	$0.23^{+0.15}_{-0.09}$ (S)	1
CDFS-5161	5955	53.0941	-27.8549	3.660	$0.24^{+0.17}_{-0.10}$ (G) ^b	2
CDFS-6664	8005	53.1388	-27.8353	3.790	$0.23^{+0.06}_{-0.05}$ (G) ^b	2
CDFS-9313	12329	53.0717	-27.7984	3.647	$0.24^{+0.06}_{-0.05}$ (G) ^b	2
CDFS-9340	12341	53.0718	-27.7981	3.658	$0.10^{+0.09}_{-0.05}$ (G)	2
GMASS-920	15555	53.1999	-27.7776	2.828	$0.023^{+0.06}_{-0.016}$ (S)	3
GMASS-1160	16841	53.1955	-27.7680	2.864	$0.26^{+0.90}_{-0.21}$ (S)	3
CDFS-11991	17345	53.1770	-27.7643	3.601	$0.25^{+0.06}_{-0.05}$ (G)	2
CDFS-12631	18372	53.0752	-27.7551	3.709	$0.27^{+0.07}_{-0.05}$ (G)	2
GMASS-1788	19760	53.1512	-27.7429	3.414	$0.069^{+0.11}_{-0.04}$ (S)	3
CDFS-14411	21187	53.0872	-27.7295	3.599	$0.28^{+0.07}_{-0.06}$ (G)	2
CDFS-16272	22942	53.0713	-27.7049	3.619	$0.24^{+0.06}_{-0.05}$ (G)	2
CDFS-16767	26121	53.1498	-27.6972	3.615	$0.38^{+0.17}_{-0.12}$ (G)	2

Notes. ^(a) G=gas-phase metallicity; S=stellar metallicity; ^(b) The stellar metallicity for these objects has been measured by stacking their spectra (S12). Stellar metallicity of the stacked object: $Z=0.17^{+0.12}_{-0.07}Z_{\odot}$

References. (1) Sommariva et al. (2012); (2) Troncoso et al. A&A submitted.; (3) This paper;

To this sample we add 7 objects from the final release of the AMAZE dataset (Troncoso et al., A&A submitted) whose gas-phase metallicity has been measured from diagnostics based on the [OII]3727, [OIII]5007 and $H\beta$ emission lines: a detailed discussion on the calibration of diagnostic diagrams can be found in Maiolino et al. (2008).

The objects in the sample and relevant metallicities are summarised in Tab. 1: all the 14 galaxies have sub-solar metallicity in the range $Z=0.07-0.39Z_{\odot}$. In this paper we will consider the best available metallicity estimate for each object, with no distinction between gas- and stellar-based measures. As discussed in S12, the gas-phase metallicity is found to be $\sim 30\%$ higher than the stellar one: given the small range of metallicities available in current stellar libraries, this has no effect on the results presented here.

It is worthwhile to remark that our sample, regrettably, is not complete in any statistical sense. As such, extrapolating our conclusion to the whole population of LBG is in principle unfair. However, the objects that we have selected are reasonably representative of the general population of bright LBGs. They have typical UV luminosity in the range $L \simeq 0.4 - 3 \cdot L^*(z = 3)$ and their selection was done (at least as far as the AMAZE sample is considered) in order to be statistically representative of average LBGs. In particular, the distributions of the AMAZE targets in two color planes (R vs. $(G - R)$, and I vs. $(I - z)$) are consistent with those of the parent spectroscopic samples of Steidel et al. (2003) and Vanzella et al. (2006), respectively. We will also show in the following that, a posteriori, these galaxies are reasonably representative of standard LBGs also in terms of their fitted physical properties.

2.2. Photometric data

The GOODS-S field covers a $\sim 10' \times 16'$ region of the Chandra Deep Field South (Giacconi et al. 2002) centered at (J2000) = 03h32m30s and (J2000) = 27d 48'20", provided with publicly available observations ranging from X-ray to the radio.

CANDELS WFC3/IR observations of the GOODS-S field include a ‘deep’ region of $\sim 6.8 \times 10$ square arcmin and a ‘wide’ field of $\sim 4 \times 10$ square arcmin both observed with the $F105W$, $F125W$ and $F160W$ filters ($Y105$, $J125$ and $H160$ hereafter). Limiting magnitudes for point sources (5σ) in the deep region are 28.2, 27.9 and 27.6 in $Y105$, $J125$ and $H160$ respectively. The corresponding limiting magnitudes in the wide region are 27.2, 27.2 and 26.7 (see Grogin et al. 2011; Koekemoer et al. 2011, for details).

CANDELS WFC3 images were combined with the available observations in the ~ 40 square arcmin of the Early Release Science area (ERS, Windhorst et al. 2011), and in the 4.7 square arcmin of the *Hubble* Ultra-Deep Field (HUDF, Oesch et al. 2010; Bouwens et al. 2010). The final WFC3 mosaics cover all the area that was observed by *HST*/ACS in $F435W$ (B), $F606W$ (V), $F775W$ (I), and $F850W$ (Z) bands as part of the GOODS (Giavalisco and the GOODS Team, in prep.) and Hubble Ultra Deep Field surveys (Beckwith et al. 2006), as well as by ACS $F814W$ ($I814$) CANDELS parallel observations (Koekemoer et al. 2011), and by VLT/VIMOS U -band (Nonino et al. 2009). The ACS mosaics used here are the version v3.0 which includes all the observations of the field carried out up to Cycle 13.

The K_s band data were acquired as part of the HAWK-I UDS and GOODS-S survey (HUGS; VLT Large Programme 186.A-0898; Fontana et al. in prep.). We exploit the full HAWK-I GOODS-S coverage comprising the central region (92.24 square arcmin, total integration time 62hs, included in the official CANDELS GOODS-S catalogue by Guo et al. 2013) and the northern and southern regions of the field (total integration time 31hs). The 5σ depth (in 1 FWHM ≈ 0.4 arcsec) of the HAWKI K_s band data ranges from 26.5 AB in the central area to 25.8 in the northern and southern pointings. The Spitzer/IRAC 3.6 and 4.5 μm observations are part of the Spitzer Extended Deep Survey (SEDS; PI G. Fazio, Ashby et al. 2013). SEDS data include observations performed both during the cryogenic and the warm Spitzer missions, reaching 5σ depths of 26.25 and 26.52 AB magnitudes (1 FWHM aperture) in channel 1 and 2 respectively. IRAC channel 3 and 4 (5.8 and 8.0 μm) observations are part of the GOODS Spitzer Legacy project (PI: M. Dickinson), and reach a 5σ limiting magnitude 23.7 AB.

A description of the different steps of the catalog conception can be found in Guo et al. (2013) (see also Galametz et al. 2013). In brief, the source extraction was done on the CANDELS $H160$ image with SExtractor (Bertin & Arnouts 1996) using a two-step detection process. SExtractor was run twice in a ‘cold’ mode (that correctly deblends extended sources) and in a ‘hot’ mode (that pushes the detection to fainter sources). The cold+hot catalog of the GOODS-S field contains 34930 sources. Total $H160$ magnitudes have been computed using the MAG_AUTO of SExtractor. Colors in all other ACS and WFC3 bands have been measured running SExtractor in dual image

mode, using isophotal magnitudes (MAG_ISO) for all the galaxies, after smoothing each image with an appropriate kernel to reproduce the resolution of the $H160$ WFC3 image. IRAC and K-band magnitudes were obtained through the Template-FITting photometry software TFIT (Laidler et al. 2007) which uses information (position, profile) of sources measured on a high resolution image ($H160$ in our case) as priors to determine photometry in the lower resolution images.

3. Physical properties from SED-fitting

We estimate physical parameters by fitting the observed photometry with the Bruzual & Charlot (2003) (BC03 hereafter) synthetic models through a χ^2 minimization as described in e.g. Fontana et al. (2003) and Santini et al. (2009). We exclude from the fit the filters sampling wavelengths below the Lyman break (on an object by object basis), in order to avoid systematic effects due to the treatment of the IGM absorption. In the fitting procedure for each object we fixed the redshift to the spectroscopic one and the stellar metallicity to the value nearest to the measured one among the ones available in the BC03 library ($Z/Z_{\odot} = 0.02, 0.2, 0.4, 1.0$). We consider the following range of physical parameters: $0.0 \leq E(B - V) \leq 1.1$, Age ≥ 0.01 Gyr (defined as the onset of the star-formation episode) and we assume a Salpeter (1955) IMF. With the goal of performing a comparison (Sect 3.1) between SED-fitting results and previous UV slope - extinction conversion equations we adopt a Calzetti et al. (2000) extinction law.

We introduce two improvements in the SED-fitting procedure:

1) We adopt four different parametrizations for the star-formation history (SFH):

- Constant SFH.
- Exponentially declining laws (SFH $\propto \exp(-t/\tau)$) with timescale $\tau = 0.1, 0.3, 0.6, 1.0, 2.0, 3.0, 5.0, 9.0$ Gyr (“ τ -models”).
- “Inverted- τ ” law (SFH $\propto \exp(+t/\tau)$) with the same range of timescales as above.
- “Rising-declining” star-formation history (SFH $\propto t^2 \cdot \exp(-t/\tau)$) with $\tau = 0.1, 0.3, 0.6, 1.0, 2.0$ Gyr. This SFH law rises up to $t = 2\tau$ and declines thereafter.

2) We include the contribution from nebular emission computed following Schaerer & de Barros (2009). Briefly, nebular emission is univocally linked to the amount of hydrogen-ionizing photons in the stellar SED (Schaerer & Vacca 1998) assuming an escape fraction $f_{esc} = 0.0$. The ionizing radiation is converted in nebular continuum emission considering free-free, free-bound, and H two-photon continuum emission, assuming an electron temperature $T_e = 10000$ K, an electron density $N_e = 100 \text{ cm}^{-3}$, and a 10% helium numerical abundance relative to hydrogen. Hydrogen lines from the Lyman to the Brackett series are included considering case B recombination, while the relative line intensities of He and metals are taken, as a function of metallicity, from Anders & Fritze-v. Alvensleben (2003).

The SED-fitting has been performed separately for each of the analytical SFHs listed above, both including and excluding nebular emission: the best-fit physical parameters and SFH for the

objects in our sample are listed in Table 2, where we also indicate the objects for which the minimum χ^2 is obtained by models with nebular emission. In the appendix (Figure A.1) we show the observed spectral energy distributions together with all the models with $P(\chi^2) > 32\%$.

Both rising and declining SFHs are effective parametrizations for the objects in our sample: 8 out of 14 objects are best-fit by exponentially declining models, the remaining LBGs are either fit with inverted-tau models or with rising-declining models having $age < \tau$, thus being in a “rising” SFH mode. None of the objects has a best-fit model with constant SFH. Regardless of the SFH, the objects are found to be very young, with best-fit ages lower than 500Myrs, and in the range 10-100Myrs in most cases. The best-fit solution is achieved by templates including nebular emission for nine out of 14 objects, a fraction in agreement with the $2/3$ value found by de Barros et al. (2012). However, best-fit models for different SFHs are still acceptable within a 68% probability threshold: for this reason in the following we will consider the fits obtained with any of the above mentioned libraries.

We verified *a posteriori* that a spectroscopic determination of metallicity is essential to avoid strong systematic effects in the SED-fitting. Indeed, when metallicity is left as a free parameter 11 out of 14 objects are found to have a wrong best-fit Z/Z_\odot , with seven of them being fit to $Z = Z_\odot$ templates.

The need for both very deep spectroscopic and IR photometric observations naturally limits the present analysis to a small number of galaxies. It is thus mandatory to assess whether these 14 objects represent a fair sample of the LBG population at $z\sim 3$ with respect to physical properties besides colour and luminosity. To this aim we performed a comparison between our objects and the full GOODS-CANDELS sample at similar redshift in the Mass-SFR plane. We considered the ~ 3000 galaxies in the official CANDELS catalogue with either spectroscopic or photometric redshift in the range $z=2.8-3.8$ and determined their best-fit SEDs through the same procedure described above. Given that we obviously lack of any information on the true metallicity for these galaxies, we separately performed the fit under two different assumptions: 1) fixing the metallicity to the reference value $Z=0.2Z_\odot$, and 2) allowing for any metallicity in the range $0.02 < Z/Z_\odot < 1.0$ but considering only models within 0.3 dex from the fundamental metallicity relation by Mannucci et al. (2010). We verified that our objects occupy the same region of the M-SFR plane as the photo- z selected objects of similar observed magnitude. A detailed discussion of the M-SFR relation at $z\sim 3$ is beyond the scope of the present paper, however we note that 5 of our objects are consistent with being “main sequence” galaxies, i.e. within 0.6dex from the M-SFR relation by Daddi et al. (2009), while the remaining lie in the “starburst” region, as most of the photo- z selected galaxies of similar luminosity.

3.1. Extinction and star formation rate of young, low-metallicity LBGs

The contribution of LBGs to the SFRD is routinely estimated by converting their UV luminosity into SFR, and applying a correction for extinction inferred from their UV slope. In particular, the

Table 2. Objects at $z\sim 3-4$ with metallicity from deep spectroscopy: best-fit physical properties

ID	β	M_{1600}	Age Myr	E(B-V)	SFR M_{\odot}/yr	log(Mass) $10^9 M_{\odot}$	SFH	Nebular
CDFS-2528	-1.94	-21.37	39^{+358}_{-20}	$0.15^{+0.0}_{-0.09}$	73^{+20}_{-51}	$3.3^{+6.1}_{-1.4}$	declining	Y
CDFS-4417	-1.31	-22.44	32^{+8}_{-6}	$0.30^{+0.0}_{-0.0}$	916^{+166}_{-37}	$32.0^{+3.5}_{-4.5}$	declining	N
CDFS-5161	-1.88	-20.90	398^{+493}_{-272}	$0.20^{+0.05}_{-0.05}$	73^{+61}_{-35}	$8.9^{+3.0}_{-2.4}$	rising-declining	N
CDFS-6664	-2.32	-21.16	18^{+82}_{-8}	$0.06^{+0.0}_{-0.06}$	37^{+9}_{-23}	$0.6^{+0.9}_{-0.3}$	declining	Y
CDFS-9313	-2.14	-21.07	200^{+363}_{-189}	$0.03^{+0.07}_{-0.03}$	15^{+46}_{-7}	$3.7^{+2.0}_{-3.1}$	declining	Y
CDFS-9340	-2.64	-20.04	$10^{+1249}_{-0.0}$	$0.06^{+0.0}_{-0.06}$	$37^{+0.0}_{-33}$	$0.1^{+1.6}_{-0.02}$	rising-declining	Y
GMASS-920	-1.88	-21.76	447^{+115}_{-223}	$0.20^{+0.0}_{-0.0}$	143^{+3}_{-12}	$28.0^{+1.8}_{-3.2}$	rising	N
GMASS-1160	-1.39	-20.60	79^{+21}_{-48}	$0.25^{+0.05}_{-0.0}$	83^{+61}_{-7}	$6.8^{+0.7}_{-1.9}$	declining	Y
CDFS-11991	-2.15	-21.85	90^{+23}_{-26}	$0.06^{+0.0}_{-0.0}$	46^{+5}_{-6}	$4.3^{+0.8}_{-1.0}$	declining	Y
CDFS-12631	-1.61	-21.44	63^{+8}_{-13}	$0.25^{+0.0}_{-0.0}$	186^{+9}_{-21}	$11.0^{+0.97}_{-0.8}$	declining	N
GMASS-1788	-1.58	-21.37	282^{+1303}_{-31}	$0.25^{+0.0}_{-0.0}$	158^{+17}_{-3}	$36.0^{+6.1}_{-2.0}$	rising	N
CDFS-14411	-1.99	-21.22	158^{+1254}_{-95}	$0.10^{+0.0}_{-0.04}$	45^{+2}_{-19}	$3.1^{+1.7}_{-0.8}$	rising	Y
CDFS-16272	-2.05	-20.65	90^{+357}_{-54}	$0.15^{+0.0}_{-0.05}$	47^{+19}_{-25}	$1.3^{+1.7}_{-0.5}$	rising-declining	Y
CDFS-16767	-1.87	-21.74	28^{+171}_{-18}	$0.15^{+0.0}_{-0.09}$	112^{+57}_{-78}	$3.3^{+6.3}_{-1.6}$	declining	Y

relation $A_{1600} = 1.99\beta + 4.43$ by Meurer et al. (1999) (M99 hereafter) has been routinely adopted to correct for dust extinction up to the highest redshifts (e.g. Bouwens et al. 2009; Ouchi et al. 2009). However, conversion equations are either calibrated on lower redshift samples or rely on assumptions on the metallicity and age of high- z LBGs. The sample analysed here thus provide a unique opportunity to compare extinction and SFR estimated from an accurate SED fitting procedure, where metallicity is constrained from spectroscopy, to the values obtained with standard fitting formulae.

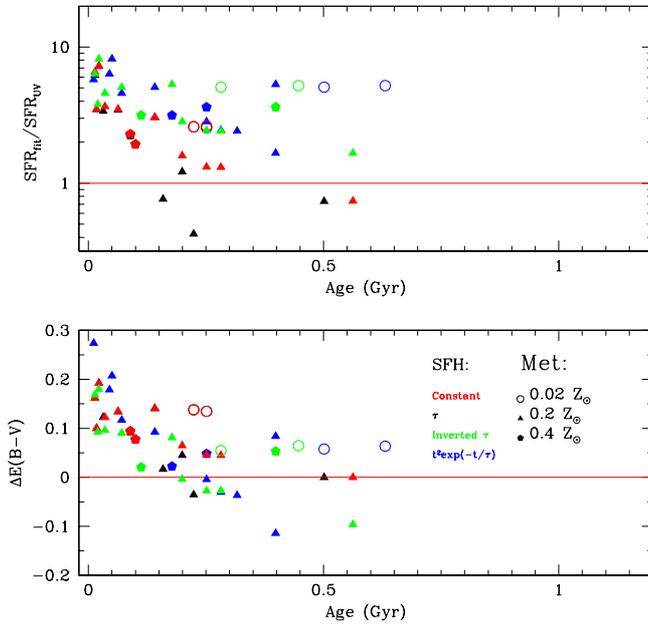


Fig. 1. Comparison between SED-fitting and UV-based extinction and SFR estimates for the 14 objects in the sample: $\Delta E(B-V) = E(B-V)_{fit} - E(B-V)_{M99}$ (top panel) and SFR_{fit}/SFR_{UV} (bottom panel). We show for each of the 14 objects the results obtained for each of the four different SFH (indicated by different colors: see figure insert) as a function of the relevant best-fit Age. SED-fitting has been computed while fixing metallicity of models to the values closer to the measured ones (as indicated by different symbols reported in the insert).

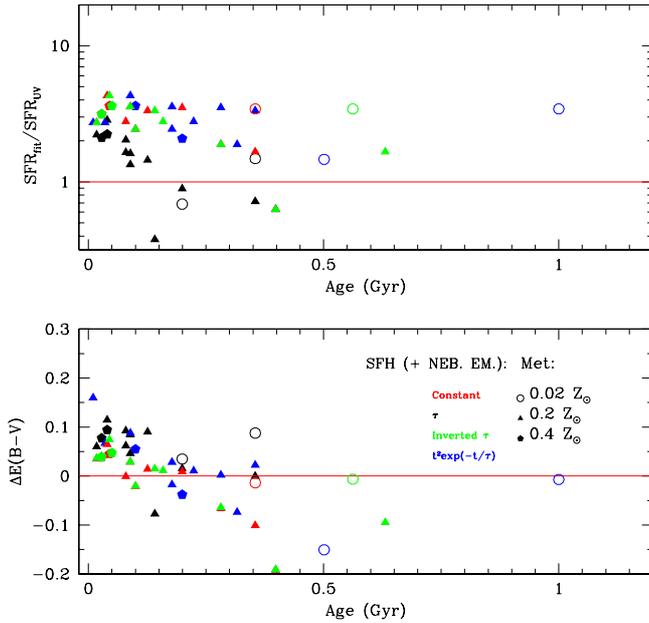


Fig. 2. Same as Fig. 1 but including the effect of nebular emission in the models.

We adopt the common power-law approximation for the UV spectral range $F_{\lambda} \propto \lambda^{\beta}$, and estimate the slope β of our objects by fitting a linear relation through the observed magnitudes spanning the UV rest-frame wavelength range of the objects (Castellano et al. 2012).

We consider the I, Z, Y, J filters for galaxies at $z > 3.4$ and the I, Z, Y only for the two GMASS objects at $z \sim 2.8$, since, at these redshifts the observed J band magnitude samples the rest-frame $\lambda \sim 3500 \text{ \AA}$ which can be affected by the emission of old stellar populations. A linear fit to the UV SED allows us also to estimate the rest-frame M_{1600} at 1600 \AA through a simple interpolation of the slope for each object. Both quantities are reported in Table 2.

The objects in our sample cover a wide range of UV slope values, from $\beta \sim -2.6$ to $\beta \sim -1.3$. We verified that they provide an unbiased sampling of the galaxy population at these redshifts by comparing them and CANDELS-GOODS photo- z selected galaxies in the same redshift range on the usual M_{1600} - β plane (e.g. Castellano et al. 2012).

We exploit the measured values of the UV slope to estimate extinction and color excess following M99 ($E(B-V)_{M99}$), and we estimate the SFR of the objects following Madau et al. (1998) (Ma98 hereafter): $SFR_{UV} = 0.125 \cdot 10^{-27} \cdot L_{UV} M_{\odot}/\text{yr}$, where L_{UV} is the extinction corrected UV emission.

A comparison between the UV based ($E(B-V)_{M99}$, SFR_{UV}) and SED-fitting estimates shows that the latter point to higher extinction and SFRs for all the adopted SFHs: the results are reported in Fig. 1 (stellar templates) and Fig. 2 (stellar+nebular models) and described in detail in the following.

In the bottom panel of Fig. 1 we show, as a function of the best-fit age, the difference $\Delta E(B-V) = E(B-V)_{fit} - E(B-V)_{M99}$ between the extinction estimated from the SED-fitting (considering templates with stellar emission only) and the value obtained through the M99 relation. The resulting

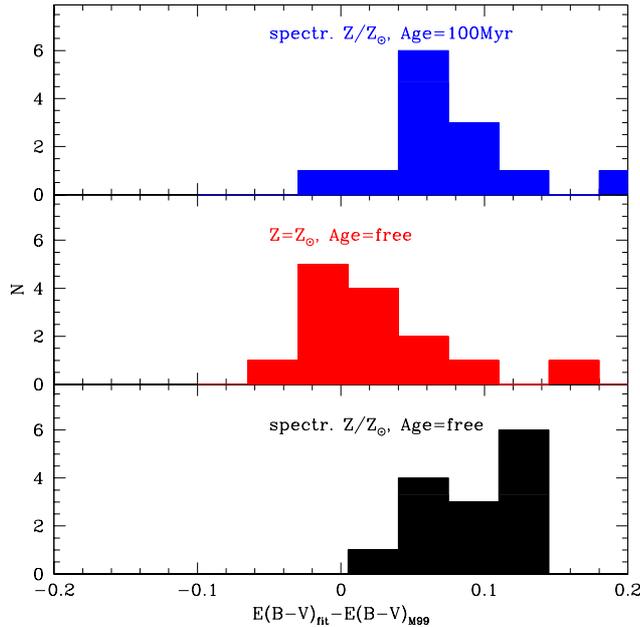


Fig. 3. Comparison between SED-fitting and UV-based extinction considering: best-fit age=100Myr, constant SFH models with metallicity fixed to the measured values (top panel); the best-fit solar metallicity models (central panel); the best-fit models for the 14 objects in the sample (bottom panel).

SFR_{fit}/SFR_{UV} ratio is displayed for all the considered SFHs in the top panel. The same comparison but with the results of stellar+nebular SED-fitting is displayed in Fig. 2. In both figures we plot for each object the best-fit estimates for each of the four SFHs discussed in the previous section. Regardless of the assumed SFH, the $E(B - V)_{M99}$ appear to be *underestimated* by $\Delta E(B - V) \sim 0.05 - 0.2$, with the discrepancy being slightly larger for stellar SED models with the lowest best-fit age (Fig. 1). When nebular emission is also included (Fig. 2), the $\Delta E(B - V)$ turns out to be lower as expected because the nebular continuum 'reddens' the slope at fixed dust extinction. However a systematic difference of $\sim 0.05 - 0.1$ is still present at young ages.

A discrepancy between UV and SED-fitting SFR estimates is in general expected at age < 100 Myr for constant SFH and $t/\tau \ll 1$ for exponentially declining and increasing ones (Reddy et al. 2012; Schaerer et al. 2013). This effect is evident in Fig. 1 and Fig. 2, but, in our case *a discrepancy is found for any SFH and any best-fit age*. Indeed, we find a systematic offset between the extinction corrected SFR_{UV} (from Ma98 equation) and SFR_{fit} with SED fitting indicating SFRs higher by a factor 2-3 (nebular+stellar SEDs) up to $\sim 8-10$ (stellar SEDs at age $\sim 10-50$ Myr). The straightforward explanation for such discrepancies lies in the difference between the sub-solar metallicity of the objects in our sample and the solar metallicity assumed in the M99 equation, which gives rise to the difference between $E(B - V)_{M99}$ and $E(B - V)_{fit}$.

We perform a simple test to constrain this scenario, in particular to assess whether allowing for young formation ages in the fit also play a significant role in determining this result. We determine the best-fit model for each object by 1) fixing age=100 Myr with constant SFH (as assumed in deriving UV-based conversion factors), and 2) fixing metallicity to the solar value while leaving

age as a free parameter. We show in Fig. 3 the resulting $\Delta E(B - V)$ estimated in these two cases, compared to the $\Delta E(B - V)$ of the best-fit models obtained as described in the previous section (metallicity fixed at the observed value, age left as a free parameter). On the one hand, the best-fit age=100Myr, constant SFH models have $E(B-V)_{fit}$ significantly different from the $E(B - V)_{M99}$ obtained through M99 fitting formula (top panel), as it is the case for the best-fit models (bottom panel). On the other hand, the best-fit solar metallicity models yield $E(B-V)_{fit}$ in much better agreement with $E(B - V)_{M99}$ (central panel). On the basis of this test we can conclude that the standard relation between UV slope and extinction (M99, assuming solar metallicity SEDs) yields to significant underestimates of dust corrected star-formation rate, at least for the objects considered here.

3.2. Independent constraints on the star-formation rates

The availability of deep Herschel far-infrared and Chandra X-ray observations of the CDFS allows us to put independent constraints on the SFR.

A 3.1σ detection for CDFS-4417 was found by Fiore et al. (2012) in the 4Ms CDFS X-ray data. On the basis of the X-ray colors and of the low X-ray to z and H band flux ratios, they conclude that the X-ray emission is due to stellar sources rather than to a nuclear source. This conclusion is also supported by the absence of AGN features in the available optical spectrum. The luminosity $\log(L_X)=42.5$ (2-10 keV) of object CDFS-4417 translates into $SFR_X \sim 630 \pm 200 M_\odot/\text{yr}$ using the conversion by Ranalli et al. (2003). Fiore et al. (2012) also report a 2.5σ detection for CDFS-4417 in the VLA-CDFS deep map at 1.4 GHz. The radio flux implies a $SFR_{1.4GHz} \sim 610 M_\odot/\text{yr}$ when applying the conversion factor by Yun et al. (2001) under the assumption of Salpeter IMF and radio spectral index $\alpha = 0.7$.

A check on the deep PACS data obtained by combining GOODS-Herschel and PEP observations (Magnelli et al. 2013) shows a significant detection for CDFS-4417 in both the $100\mu\text{m}$ (3.3σ) and $160\mu\text{m}$ (3.8σ) bands.

We estimate the dust-obscured SFR by converting monochromatic fluxes into total FIR luminosity (8-1000 μm) by means of the main-sequence template built by Elbaz et al. (2011), and applying the SFR- L_{8-1000} conversion by Kennicutt (1998). We obtain $SFR=900 \pm 270 M_\odot/\text{yr}$ from the $100\mu\text{m}$ flux, and $SFR=680 \pm 180 M_\odot/\text{yr}$ from the $160\mu\text{m}$ one. When the starburst template is used instead of the main-sequence one, the estimates are $\sim 10 - 30\%$ lower.

The SED-fitting estimate for CDFS-4417 is $SFR_{fit}=916.0_{-37}^{+166} M_\odot/\text{yr}$ (best-fit value with declining SFH, Tab. 2). We estimate the amount of unobscured SFR in CDFS-4417 from its observed (i.e. not corrected for dust) UV luminosity, obtaining $SFR_{UV,unobs.} = 52.7 M_\odot/\text{yr}$. We can thus predict a dust-obscured star-formation rate $\sim 860 M_\odot/\text{yr}$. This value is in very good agreement with the estimate obtained from the $100\mu\text{m}$ flux, and consistent within the uncertainty with the $160\mu\text{m}$ estimate. The SFR_X and $SFR_{1.4GHz}$ estimated from X-ray and radio emission (which are insensitive to dust) are also consistent with SFR_{fit} within the relevant uncertainties. In turn, the

SFR obtained from dust-corrected UV emission ($SFR_{UV} = 270.5 M_{\odot}/yr$, $SFR = 217.8 M_{\odot}/yr$ when removing $SFR_{UV,unobs.}$) is in clear disagreement with both the FIR and X-ray measures being a factor >2 lower.

The other objects in our sample are not detected in the PACS images. This is not surprising, since the PACS 3σ flux limits (Magnelli et al. 2013) imply star-formation rate upper-limits of the order of $\lesssim 600-800 M_{\odot}/yr$ (in the redshift range considered here), higher than the obscured $SFR \sim 10-160 M_{\odot}/yr$ we estimate for the other objects in the sample. To further investigate this issue, we exploit the public PACS data to build a stacked far-IR image of the 13 objects which are not individually detected. After masking all detected sources, we extract 20×20 arcsec thumbnails centred on the position of the LBGs, which are then combined as a weighted average. We find a $\sim 2\sigma$ detection in both the $100\mu m$ and $160\mu m$ stackings: $F_{100} = 0.120 \pm 0.059$ mJy and $F_{160} = 0.375 \pm 0.202$ mJy. These fluxes imply a $SFR \sim 70-170 M_{\odot}/yr$, considering the stacked object to be at the median redshift of the sample: despite the large uncertainty, this estimate is consistent with the SFR range indicated by the SED-fitting, while being 3-5 times higher than the obscured $SFR \sim 5-30 M_{\odot}/yr$ we infer from the UV, consistently with our finding of SFR_{UV} being a factor $\gtrsim 2$ underestimated.

We finally computed star-formation rates for the 11 objects of the AMAZE sample from the relevant $H\beta$ fluxes. We correct line fluxes through the reddening inferred from the continuum fitting and obtain the $H\alpha$ luminosity by assuming the case B recombination ($H\alpha/H\beta = 2.8$). The $H\alpha$ luminosity is then converted into SFR following Kennicutt (1998). We find good agreement between SFR_{fit} and $SFR_{H\beta}$. The consistency between these two SFR indicators is remarkable since it also confirms the very young ages of the objects in our sample, being recombination lines tracers of the star formation rate on a $t < 20$ Myr timescale (Kennicutt 1998).

4. A revised $A_{1600} - \beta$ relation: consequences on the star-formation rate density at $z \sim 3$

Motivated by the results discussed above we determine a more appropriate conversion between UV slope and $E(B-V)$ on the basis of the average $\langle \Delta E(B-V) \rangle = 0.092$ found from our best-fit results. Such $\langle \Delta E(B-V) \rangle$ translates into an extra $\Delta A_{1600} = 0.89$, which leads to the following modified $A_{1600} - \beta$ relation:

$$A_{1600} = 5.32 + 1.99 * \beta \quad (1)$$

As shown in Fig. 4, systematic discrepancies between SFR estimates are eliminated when extinction is estimated through this new relation. Eq. 1 implies that the UV slope of dust-free objects is $\beta_{dust-free} = -2.67$, significantly bluer than the “zero-point” $\beta_{dust-free} = -2.23$ originally included in the M99 formula. The result in Eq. 1 is in agreement with the analysis of a large $z \sim 4$ LBG

sample by de Barros et al. (2012) (yielding $\beta_{dust-free} = -2.58$), and close to the theoretical value $\beta_{dust-free} \simeq -2.5$ which is found by Dayal & Ferrara (2012) for young, low-metallicity LBGs, albeit at higher redshift ($z > 6$). A check on models from the BC03 library shows that the $\beta_{dust-free}$ we find is consistent with the slope of a dust-free $Z=0.2Z_{\odot}$, age=50Myr, constant SFH galaxy, which, in the light of the previous analysis, can indeed be considered as a good reference model in terms of age and metallicity.

We conclude that Eq. 1 provides on average a $A_{1600} - \beta$ relation which is more appropriate for high-redshift LBGs with young ages and sub-solar metallicity.

We explore the consequences of Eq. 1 on the computation of the corrected SFRD. We consider the relevant $z \sim 3$ UV luminosity function (LF) by Reddy & Steidel (2009) and a Gaussian distribution of UV slope values following the estimate by Bouwens et al. (2009, 2013) for U-drop selected galaxies. The adopted UV slope distribution has an intrinsic scatter $\sigma = 0.35$, and a magnitude dependent average decreasing from $\beta \sim -1.2$ at $M_{UV} = -22$ to $\beta \sim -2.2$ at $M_{UV} = -17$. The dust-corrected SFRD is obtained through the following integral:

$$SFRD = \frac{1.0}{8 \cdot 10^{27}} \int dL \int dA \cdot PDF(A, L) \cdot 10^{0.4A} \cdot L \cdot \Phi(L), \quad (2)$$

where the probability distribution function of extinction correction values ($PDF(A, L)$) is univocally related to the UV slope distribution $PDF(\beta, L)$ through Eq. 1, and $\Phi(L)$ is the UV luminosity function at 1600\AA . The integral is computed up to $M_{1600} = -17.48$ (corresponding to $L = 0.04L^*$) for consistency with previous works. We obtain $SFRD = 0.37 M_{\odot}/yr/Mpc^3$. For comparison, if we instead adopt the M99 $A_{1600} - \beta$ relation, we obtain a factor 2.2 lower $SFRD = 0.16 M_{\odot}/yr/Mpc^3$, consistent with Bouwens et al. (2009). Our estimate is still a factor 1.8 higher than the one by Reddy & Steidel (2009) which adopt a more conservative dust correction factor. In both cases, the total SFRD is higher than so far estimated primarily because of the largest contribution of bright LBGs, that result from our Eq. 1. Fainter galaxies, indeed, remain closer to the previous estimates as they are on average bluer and hence less extinguished.

The SFRD estimate presented above is based on the assumption of Salpeter (1955) IMF and Calzetti et al. (2000) extinction law to allow for a self-consistent comparison between our use of Eq. 1 and the Meurer et al. (1999) one in the computation of dust correction. While these are common assumptions, we must note that the adoption of a different IMF or extinction law can result in changes to the SFRD which can partially compensate systematic errors introduced by the use of the Meurer et al. (1999) formula: a top-heavy IMF yield to a lower conversion factor between UV and SFR, while a steeper extinction curve (e.g. SMC-like) would imply a lower extinction at fixed UV slope, and thus a lower dust-correction factor.

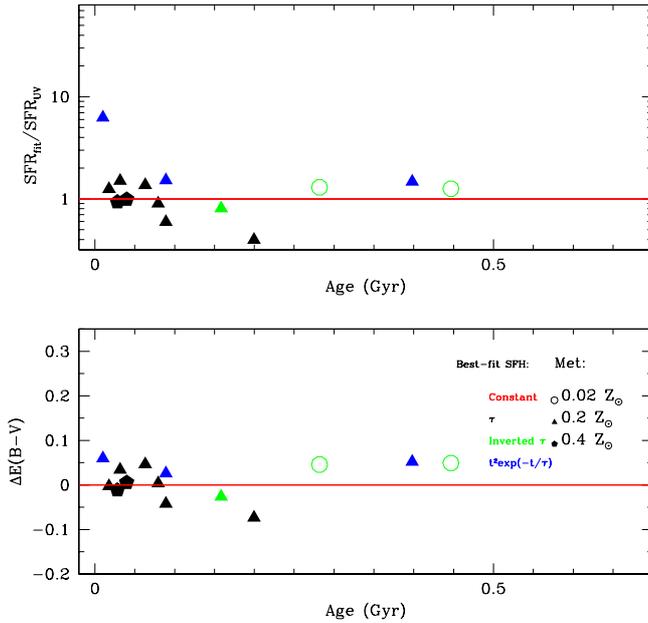


Fig. 4. $\Delta E(B-V)$ vs. Age (upper panel) and SFR_{fit}/SFR_{UV} vs. Age applying the modified $\beta - A_{1600}$ relation in Eq. 1.

5. Constraints on the past evolutionary history

A remarkable feature highlighted by the SED-fitting results is the very young age of the objects, although this is not fully surprising given the very low metallicity found through spectroscopic measurements. The agreement between SFR_{fit} and the star-formation rate inferred from $H\beta$ fluxes (Sect. 3.2) is a further indication that these objects are dominated by stellar population with ages of few 10 Myrs. Indeed, ten of the LBGs in our sample have ages < 100 Myrs, one of them (CDF5-9340) having the best-fit solution at the minimum allowed age in the fit (10 Myrs). Moreover Fig. 1-2 demonstrate that this result does not strongly depend on the choice of the SFH and on the inclusion of nebular emission, since all the fits we performed indicate ages $\lesssim 300$ Myrs with very few exceptions.

At the redshifts we are considering age estimates are mainly driven by the younger, UV-bright population. While this is not a concern when focusing our attention on the star-formation rate, it is nonetheless of great importance to assess the presence of any significant contribution from older stellar populations to the SEDs. To constrain the past evolutionary history of the objects in our sample we will analyse here in more detail their age estimates by assuming both parametric and non-parametric star-formation histories.

A first, straightforward test can be performed by looking at the amplitude of the Balmer break, which is the most evident age-dependent feature sampled by the observations analysed here. To this aim, we define a colour term (“Balmer colour”) bracketing the 4000\AA break: $B\text{Col} = H160 - 0.5 \times (K + 3.6\mu\text{m})$. For the two GMASS objects at $z \sim 2.8$ we adopt the J125 band instead of H160 since the latter is itself affected by the break at those redshifts. To disentangle the competing effects of dust extinction and age we perform a comparison between our objects and population synthesis

models in a ‘‘Balmer colour’’ vs. UV slope plane. We fix the metallicity of the models to the reference value $Z=0.2Z_{\odot}$ and measure colours and UV slopes of the templates in the same way as for the observed sample.

As shown in Fig 5 the Balmer colour of our objects lies in the range $B\text{Col} \sim 0-0.8$, and shows - as expected- a clear dependence on β (see also Oesch et al. 2013): objects with larger Balmer colour have also redder UV slope. The same trend is evident in the models, regardless of the SFH. For six of the objects the combination of Balmer colour ($\sim 0.0-0.5$) and UV slope ($\beta \gtrsim -2.5$), unequivocally indicates ages of 10-50 Myrs for any assumed SFH. All the galaxies in the sample have a Balmer break which is only compatible with $\text{Age} \lesssim 300\text{Myr}$ when declining (bottom panel in Fig 5) and constant (central panel) SFH models are considered. However, when an exponentially increasing SFH is assumed (top panel), the age of the redder galaxies turns out to be poorly constrained, with their $B\text{Col} \sim 0.8$ being compatible with templates of age $\sim 0.3-1.0$ Gyr.

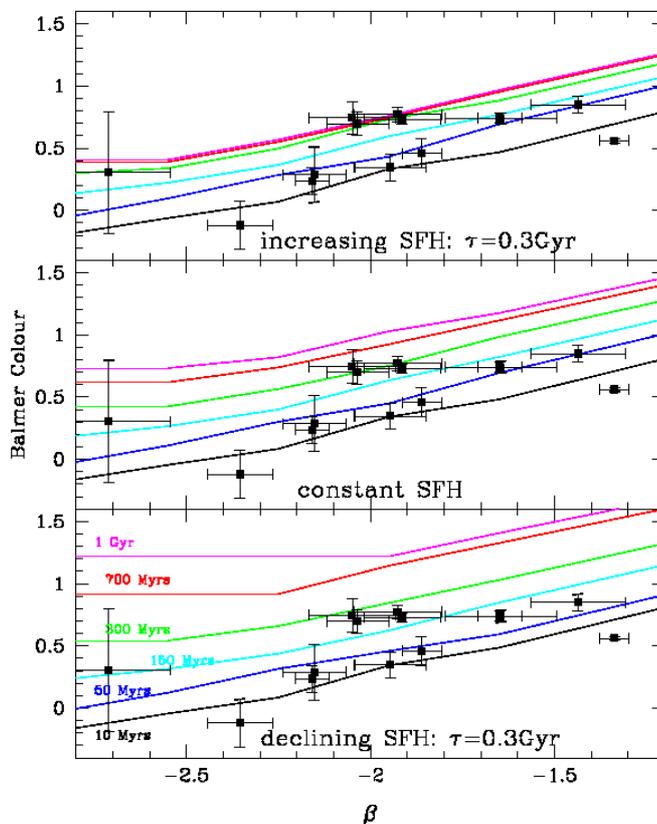


Fig. 5. The amplitude of the Balmer break as a function of the UV slope for the objects in our sample (black squares and error-bars) compared to templates at different ages (continous lines: 0.01 to 1 Gyr from bottom to top) and star formation histories: exponentially declining SFH ($\tau = 0.3$ Gyr, bottom panel), constant SFH ($\tau = 0.3$ Gyr, central panel), exponentially increasing SFH ($\tau = 0.3$ Gyr, top panel). The Balmer colour is defined as $H160-0.5 \times (K+3.6\mu m)$ for objects and templates at $z > 3.4$, and $J125-0.5 \times (K+3.6\mu m)$ for lower redshift ones. The reddening vector in this plot is parallel to the displayed model tracks.

We further compute the maximum allowed age of each object for any of the SFHs adopted in Sect. 3. This is defined as the oldest model with $P(\chi^2) > 32\%$. We note that this is a conservative choice, since the uncertainty is computed from all the bands used for the fit, not only on those 2-3 around the Balmer break where the effect of age is most evident. As shown in Fig. 6, all the objects have a maximum age $\lesssim 500\text{Myr}$ (i.e. a formation redshift $z_{\text{form}} \lesssim 6$) when assuming

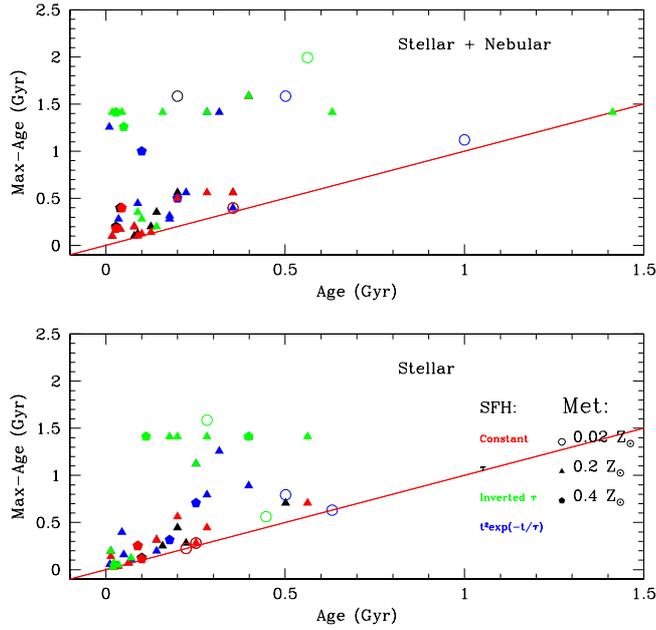


Fig. 6. Maximum allowed age assuming different SFHs, as a function of the relevant best-fit age.

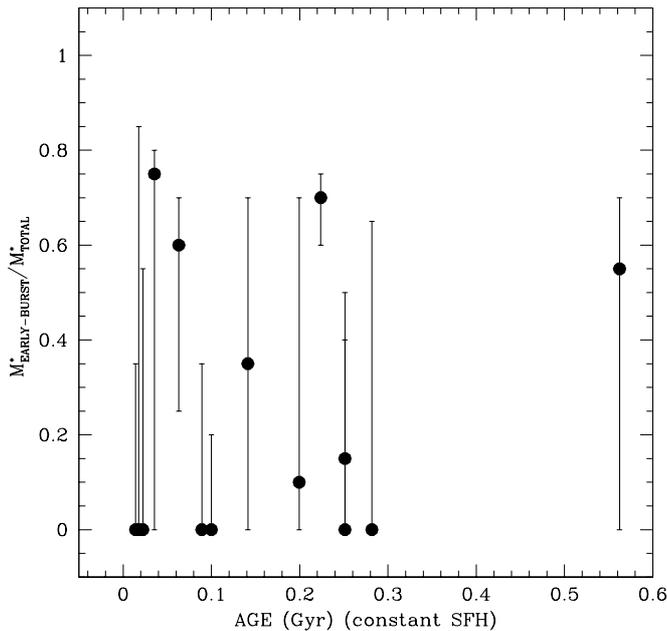


Fig. 7. Best-fit ratio $M_{EARLY-BURST}^*/M_{TOTAL}^*$, where $M_{EARLY-BURST}^*$ is the stellar mass formed during an “early-burst” at age=1.3-1Gyr and the total stellar mass $M_{TOTAL}^* = M_{LATE-BURST}^* + M_{EARLY-BURST}^*$ includes a “late-burst” at age=300Myrs.

declining or constant SFHs. On the other hand, half of the LBGs is compatible with age ≥ 1.0 Gyr ($z_{form} \sim 10$) when assuming rising star-formation histories. We verified that this result does not significantly depend on our definition of age as the onset of the star-formation episode: as an example, the age at which 10% of the stars is already in place is typically 5-20 % lower, and the conclusions of our test do not qualitatively change. Consistently with our previous findings on the amplitude of the Balmer break, this test indicates that the assumed parametric form for the

SFH largely affects constraints on the presence of old stellar populations: while rising and “rising-declining” SFH allow for a high formation redshift within the best-fit uncertainty, both declining and constant SFH univocally indicate ages of few 100Myrs. However, we note that only two of the objects compatible with $\text{age} \gtrsim 1.0$ Gyr actually have models with rising SFHs as preferred solution, so we can conclude that there is little evidence for a significant presence of older stellar populations in the objects of our sample.

We finally build a double-component library of BC03 models, i.e. we assume that the SEDs originate from two different bursts with constant SFH of different intensity: an “early burst” at $\text{age} = 1.3\text{-}1\text{Gyr}$, and a “late-burst” started at $\text{age} = 300$ Myrs. We then fit our objects determining the best-fit ratio $M_{\text{EARLY-BURST}}^*/M_{\text{TOTAL}}^*$ of the stellar mass formed during the early burst and the total stellar mass ($M_{\text{TOTAL}}^* = M_{\text{LATE-BURST}}^* + M_{\text{EARLY-BURST}}^*$). In Fig. 7 we plot the best fit burst ratio and its uncertainty as a function of the best-fit age of single-component constant SFH models. The best-fit ratio turns out to be < 0.2 for nine out of 14 objects in our sample, and equal to zero (no presence of an old burst) for 6 of them, supporting the low formation redshift obtained with parametric SFH models. In turn, four of the objects appear to be dominated by the old burst (ratio > 0.5), with three of them having a low age in the single-component SFH fit. Moreover, the uncertainties on the fit allow for a 20-40 % contribution of the old population even in the SEDs dominated by young stars.

These results coherently point to a minor contribution from older stellar populations to the SEDs of our objects, although significant uncertainties remain, in particular because of the poor constraints on the SFH. Our findings are consistent with previous results indicating ages of $\sim 10^8$ yrs for high-redshift LBGs (Reddy et al. 2012; Curtis-Lake et al. 2013; Oesch et al. 2013) but, at variance with the mentioned works, young ages < 100 Myrs are favoured for a significant part of the objects in our sample. This result, and the robust estimates of high SFRs and moderate extinctions described in the previous sections, agrees with a scenario where luminous LBGs are objects caught in a “bursting”, moderately obscured phase lasting few 100Myrs (e.g. Stark et al. 2009). This scenario is in contrast with the idea of a smooth evolution of the LBG population (e.g. Finlator et al. 2007), since most of these objects at the bright end of the UV LF at $z \sim 3$ were probably populating the faintest end of the UV LF already at $z \sim 5 - 6$.

6. Summary and Conclusions

In this work we have performed accurate SED fitting of a unique sample of 14 galaxies at $2.8 \lesssim z \lesssim 3.8$ in the GOODS-South field. These galaxies constitute the only LBGs with both a spectroscopic measurement of their metallicity (either gas-phase or stellar, as measured from the AMAZE and GMASS surveys) and deep IR observations (obtained combining the CANDELS, HUGS and SEDS surveys). Unfortunately, our galaxies do not make a complete sample in any statistical way; however, a posteriori they appear to have been selected from the general population of massive and luminous LBGs, and as such their analysis can shed some light on the general properties of the overall population of LBGs, or at least their brighter subsample.

We have taken advantage of the 17 bands CANDELS catalogue to perform accurate SED fitting while fixing redshift and metallicity of population synthesis models to the measured values. For the spectral fit we use the BC03 models with Salpeter IMF, and we explore both different SFH (ranging from exponentially declining to rising) as well as models with or without the inclusion of nebular emission. The latter has been computed both in the lines and continuum component following the procedure described in Schaerer & de Barros (2009).

We summarize here our findings, that are connected with two broad areas of investigation about high redshift LBGs: their dust content and their implied contribution to the global SFRD, and their ages and previous SFHs.

UV slope, dust content and star-formation rates. We first compared the SFR obtained through SED-fitting SFR_{fit} with those estimated from the observed UV luminosity after correcting for the observed extinction (SFR_{UV}), in the same manner as typically done on existing large LBG samples. For the latter estimate, we measured UV spectral slopes β through a linear fit of HST magnitudes, and used the relevant extinction estimates (Meurer et al. 1999) to estimate corrected SFR_{UV} according to the standard Madau et al. (1998) L_{UV} -SFR conversion. A comparison between SFR_{fit} and SFR_{UV} shows that the latter are underestimated by a factor 2-8, for all objects. SFR_{fit} exceeds SFR_{UV} regardless of the assumed SFH, and the overestimate is larger for models without nebular emission (where it ranges typically between 3 and 8) rather than in models with nebular emission (ranging between 2 and 5).

This result is supported by independent constraints on the radio (VLA), far-IR (Herschel) and X-ray (Chandra) emission of object CDFS-4417, which give SFR a factor 2-4 larger than SFR_{UV} and in closer agreement with SFR_{fit} . This conclusion is supported by the analysis of the far-IR stacking of the 13 sources which are not individually detected in Herschel data. The stacked object shows a $\sim 2\sigma$ detection at both $100\mu m$ and $160\mu m$, implying an average obscured SFR consistent with the SED-fitting results and 3-5 times higher than UV estimates. The $H\beta$ luminosity measured for 11 of the objects also yield SFRs in agreement with SFR_{fit} , and confirm that the objects are young and intensely star-forming.

We demonstrate that such discrepancies are mostly due to the standard assumption of solar metallicity stellar populations in the widely used Meurer et al. (1999) UV slope-extinction conversion (Fig. 3). On the basis of our results we deduce a new $\beta - A_{1600}$ relation, $A_{1600} = 5.32 + 1.99 * \beta$ (Eq. 1), that we propose to be more appropriate for young sub-solar metallicity LBGs. We note that this formula implies that a “dust-free” UV slope as steep as $\beta = -2.67$, significantly bluer than the current assumption $\beta = -2.23$ based on the (Meurer et al. 1999) formula.

It is interesting to explore the possible implications of these findings, under the assumption that these results can be extended to the overall LBG population. First, the common knowledge of negligible dust extinction in $\beta \lesssim -2.0$ high-redshift galaxies would be seriously challenged. This might bring the current theoretical predictions in better agreement with the observations, given that

such models inevitably predict a rapid formation of substantial amounts of dust in high redshift LBGs.

Another important implication is on the contribution of LBGs to the global SFRD. We exploit our refined $\beta - A_{1600}$ relation to compute the $z\sim 3$ SFRD on the basis of available estimates of the UV luminosity function and UV slope distribution at these redshifts. We find a dust corrected SFRD = $0.37 M_{\odot}/yr/Mpc^3$, more than two times higher than values based on old UV slope-extinction conversions.

Ages and star-formation histories. We finally analysed in detail the constraints that our SED fitting is able to produce on the age and in general on the past evolutionary history of the objects in our sample (Sect. 5).

We note that, on the basis of our fits, we are not able to constrain in a robust way the SFH. In fact, both rising and declining models are found as best-fit solutions: eight out of 14 LBGs are best-fit by exponentially declining models, the remaining are either fit with inverted-tau models or with rising-declining models having $age < \tau$, thus being in a “rising” SFH mode. In addition, the various SFHs are typically able to produce acceptable χ^2 for most objects, such that the preference for a given SFH is not statistically robust even on individual objects.

Our central result here is that we find very young best-fit ages for all our objects, in the range 10-500 Myrs (Tab. 2). This result holds for any assumed SFHs and both including or excluding nebular emission: in all these cases we find best-fit ages $\lesssim 500$ Myrs with very few exceptions.

We have carefully explored whether this result is robust, given the expected prevalence of young stars in the overall SED, that may lead to important underestimates of the true age (the so called “overshining” problem).

We first decided to avoid the possible complications of the SED fitting process and analysed a specific color term designed to be particularly sensitive to age effects. We defined a “Balmer” colour ($B_{Col} = H160 - 0.5 \times (K + 3.6\mu m)$) that brackets the 4000\AA break, and computed B_{Col} as a function of the β value for templates of different SFHs (Fig. 5), at different ages. We show that six of the galaxies in our sample have a combination of low Balmer break ($\sim 0.0-0.5$) and UV slope ($\beta \gtrsim 2.5$), unequivocally indicating very low ages (10-50 Myrs) for any SFH. The remaining galaxies have a position in the $B_{Col}-\beta$ plane indicating Age $\lesssim 300$ Myr when declining or constant SFH are assumed, while only four objects with $B_{Col} \sim 0.8$ are compatible with exponentially increasing SFH templates of age $\sim 0.3-1.0$ Gyr.

Going back to the SED fitting technique, we have also performed an estimate of the “maximal” age compatible for each SFH, defined as the largest age that produce a model with $P(\chi^2) > 32\%$. The “maximal” age remains less 0.5 Gyr for at least half of the sample, regardless of the detailed SFH. The remaining 50% can be reconciled with large ages (equivalent to formation redshifts around 10) only with increasing SFH. Similar results have been found using a specific set of models with double-burst SFH templates, where the relative intensity of the two bursts is left free. Even

in this somewhat extreme case nine out of 14 objects in the sample are found to have a very low fraction (<0.2) of old stellar population to their SED, although significant uncertainties remains.

The results summarised above show that tight constraints on metallicity and on the rest-frame optical regime are fundamental to shed light on two debated issues: 1) the estimate of dust-extinction and dust-corrected SFRs at high-redshift, and 2) the assembly history of Lyman-break galaxies.

On the one hand, we have shown that taking into account the sub-solar metallicity of stellar populations yields to a significant revision of the UV slope-extinction conversion and of the corrected SFRD. On the other hand, the ages of few 10-100Myrs we find for our objects, the low amplitude of their Balmer-break, and the minor impact of older stars to their SEDs, suggest a particular scenario for the assembly of at least a sizeable fraction of the high-redshift LBGs. In fact, these luminous objects are most probably caught during huge star-formation burst moving them on short timescales from the faint to the bright end of the luminosity function, rather than being the result of a constant, smooth evolution.

A final word on these two questions will only come by the analysis of larger, and fainter samples. Extending the present work to a larger number of bright $z \gtrsim 3$ LBGs will be possible thanks to intensive spectroscopic follow-up campaigns such as the ongoing VIMOS Ultra-Deep Survey, which is targeting 10000 sq. deg. including sky regions provided with deep near-IR observations. However, pushing this analysis to fainter galaxies, and to the redshifts approaching the reionization epoch, is beyond the current limits of available instrumentation. Near infrared observations deeper than the ones presented here ($m_{lim} \sim 26.5$ AB), and spectroscopic constraints of absorption features in $M \gg M_*$ LBGs, will be within reach only thanks to JWST and EELT.

Acknowledgements. We acknowledge the contribution of the FP7 SPACE project ‘‘ASTRODEEP’’ (Ref.No: 312725), supported by the European Commission. RJM acknowledges funding via an ERC consolidator grant (P.I. R McLure). PT has been supported by the Marie Curie Initial Training Network ELIXIR under the contract PITN-GA-2008-214227 from the European Commission. MC thanks A. Lamastra for the useful discussions.

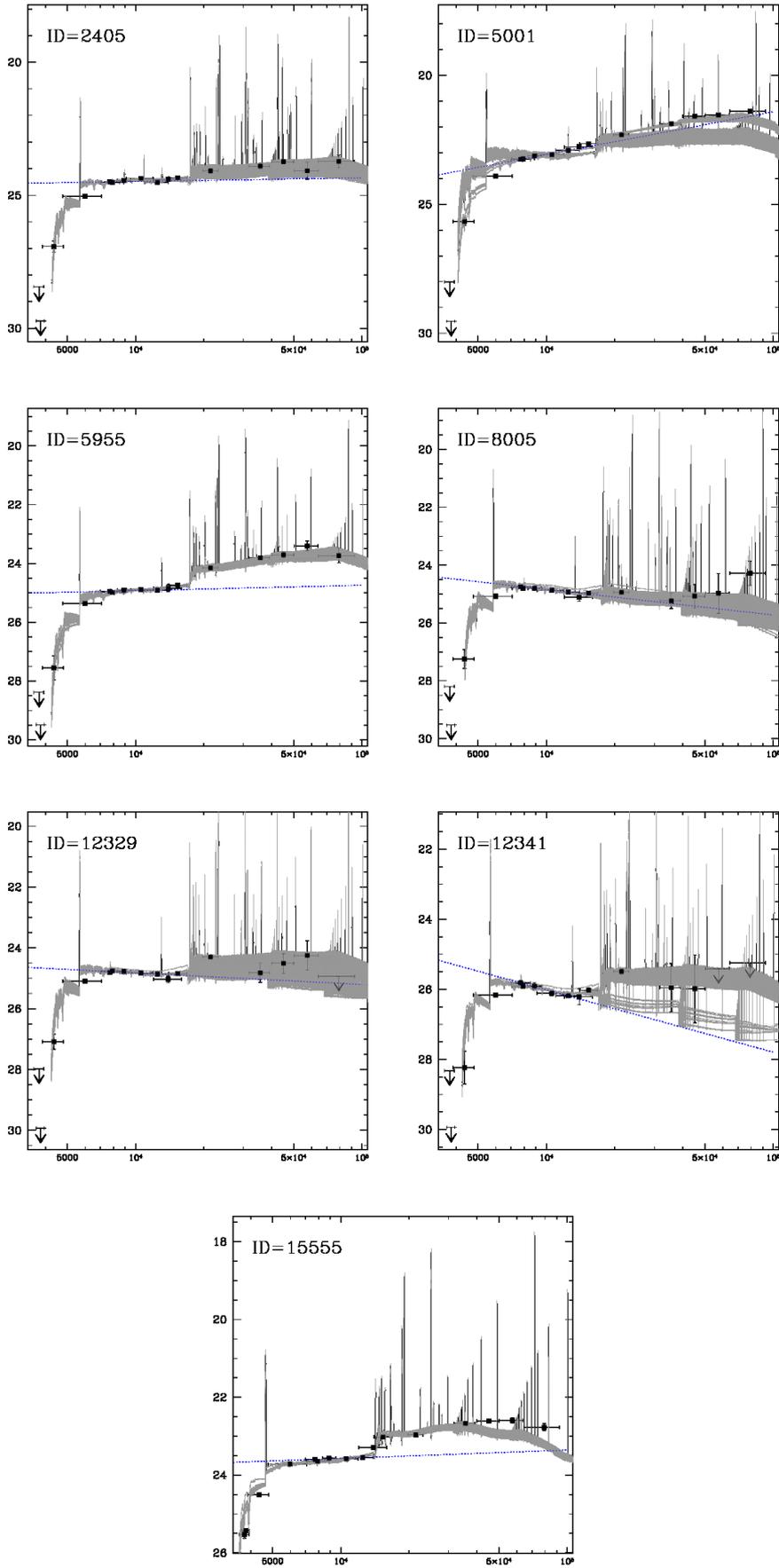
References

- Anders, P. & Fritze-v. Alvensleben, U. 2003, A&A, 401, 1063
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
- Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
- Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, ApJ, 705, 936
- Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928
- Bouwens, R. J., Illingworth, G. D., Gonzalez, V., et al. 2010, ArXiv e-prints
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, ApJ, 754, 83
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2013, ArXiv e-prints
- Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
- Castellano, M., Fontana, A., Grazian, A., et al. 2012, A&A, 540, A39
- Castellano, M., Fontana, A., Paris, D., et al. 2010, A&A, 524, A28
- Curtis-Lake, E., McLure, R. J., Dunlop, J. S., et al. 2013, MNRAS, 429, 302
- Daddi, E., Dannerbauer, H., Stern, D., et al. 2009, ApJ, 694, 1517
- Dayal, P. & Ferrara, A. 2012, MNRAS, 421, 2568

- de Barros, S., Schaerer, D., & Stark, D. P. 2012, ArXiv e-prints
- Dunlop, J. S., Rogers, A. B., McLure, R. J., et al. 2013, MNRAS, 432, 3520
- Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
- Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, ApJ, 756, 164
- Finlator, K., Davé, R., & Oppenheimer, B. D. 2007, MNRAS, 376, 1861
- Fiore, F., Puccetti, S., Grazian, A., et al. 2012, A&A, 537, A16
- Fontana, A., Poli, F., Menci, N., et al. 2003, ApJ, 587, 544
- Galametz, A., Grazian, A., Fontana, A., et al. 2013, ApJS, 206, 10
- Giacconi, R., Zirm, A., Wang, J., et al. 2002, ApJS, 139, 369
- Grazian, A., Castellano, M., Koekemoer, A. M., et al. 2011, A&A, 532, A33+
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
- Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189
- Kimm, T. & Cen, R. 2013, ArXiv e-prints
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
- Kurk, J., Cimatti, A., Daddi, E., et al. 2013, A&A, 549, A63
- Lacey, C. G., Baugh, C. M., Frenk, C. S., & Benson, A. J. 2011, MNRAS, 412, 1828
- Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
- Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106
- Magnelli, B., Popesso, P., Berta, S., et al. 2013, A&A, 553, A132
- Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
- McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, MNRAS, 432, 2696
- Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
- Nonino, M., Dickinson, M., Rosati, P., et al. 2009, ApJS, 183, 244
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2010, ApJ, 709, L16
- Oesch, P. A., Labbé, I., Bouwens, R. J., et al. 2013, ApJ, 772, 136
- Ouchi, M., Mobasher, B., Shimasaku, K., et al. 2009, ApJ, 706, 1136
- Ranalli, P., Comastri, A., & Setti, G. 2003, A&A, 399, 39
- Reddy, N. A., Pettini, M., Steidel, C. C., et al. 2012, ApJ, 754, 25
- Reddy, N. A. & Steidel, C. C. 2009, ApJ, 692, 778
- Rix, S. A., Pettini, M., Leitherer, C., et al. 2004, ApJ, 615, 98
- Salpeter, E. E. 1955, ApJ, 121, 161
- Santini, P., Fontana, A., Grazian, A., et al. 2009, A&A, 504, 751
- Schaerer, D. & de Barros, S. 2009, A&A, 502, 423
- Schaerer, D., de Barros, S., & Sklias, P. 2013, A&A, 549, A4
- Schaerer, D. & Vacca, W. D. 1998, ApJ, 497, 618
- Somerville, R. S., Gilmore, R. C., Primack, J. R., & Domínguez, A. 2012, MNRAS, 423, 1992
- Sommariva, V., Mannucci, F., Cresci, G., et al. 2012, A&A, 539, A136
- Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, ApJ, 592, 728
- Vanzella, E., Cristiani, S., Dickinson, M., et al. 2006, A&A, 454, 423
- Wilkins, S. M., Bunker, A., Coulton, W., et al. 2013, MNRAS, 430, 2885
- Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
- Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, ApJ, 554, 803

List of Objects

Appendix A: Spectral energy distributions



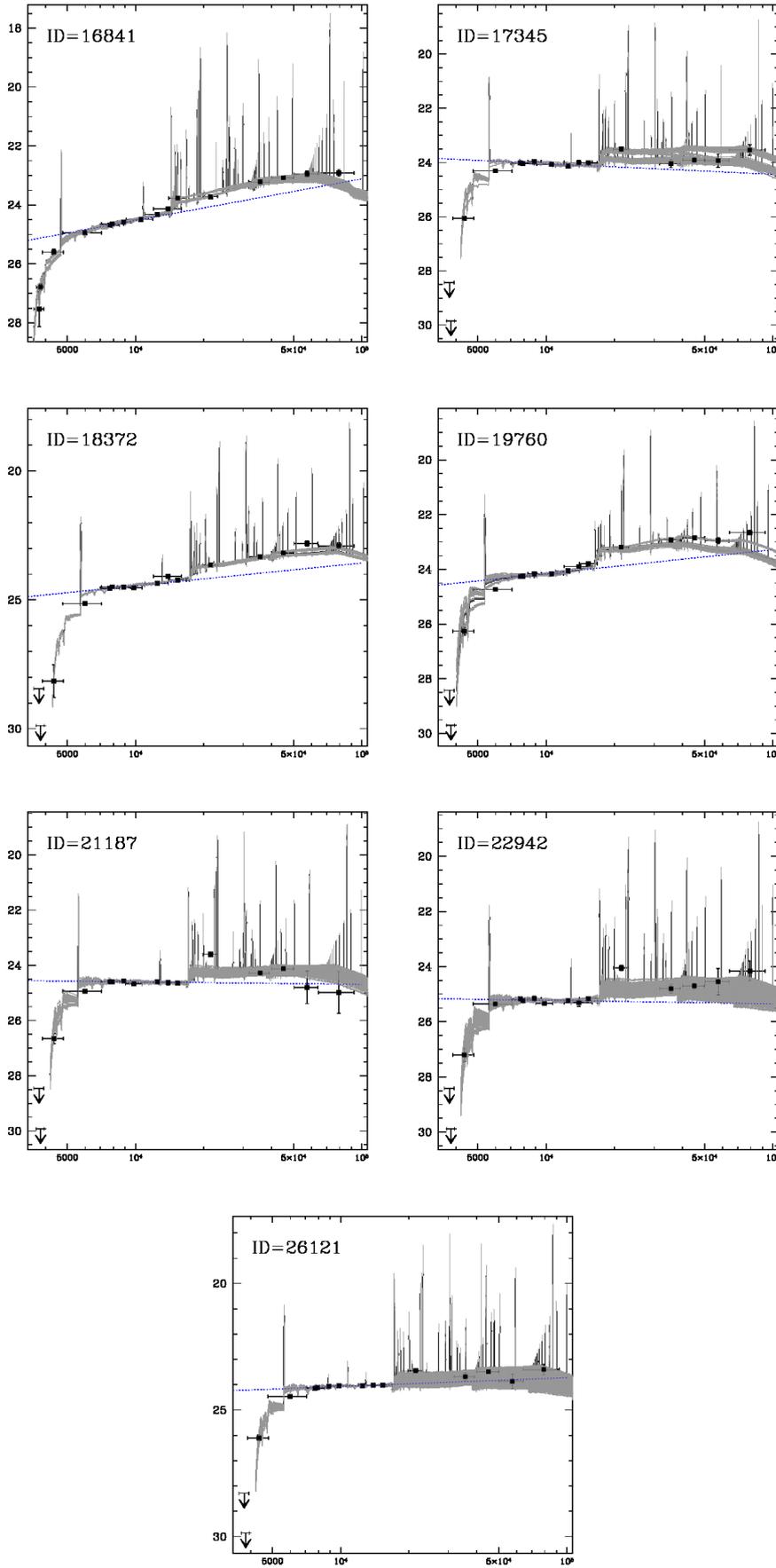


Fig. A.1. Spectral-energy distributions of the 14 objects analysed in this paper. Grey curves in each plot show models with $P(\chi^2) > 32\%$ from the best-fit, considering four different SFH and fits with nebular+stellar and stellar emission only (see Sect. 3). The best-fit UV slope is shown as blue dashed line.

A mass threshold in the number density of passive galaxies at $z \sim 2$

V. Sommariva¹, A. Fontana¹, A. Lamastra¹, M. Nonino², M. Castellano¹, J. S. Dunlop³, H. Ferguson⁴, R. J. McLure³,
A. Galametz¹, M. Giavalisco⁵, A. Grazian¹, Y. Lu⁶, N. Menci¹, D. Paris¹, L. Pentericci¹, P. Santini¹, R. Somerville⁴,
and T. Targett³

¹ INAF - Osservatorio Astronomico di Roma, Via Frascati 33, I00040, Monteporzio, Italy e-mail: veronica.sommariva@oa-roma.inaf.it

² INAF - Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, 34131 Trieste, Italy.

³ SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, UK

⁴ Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

⁵ Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA

⁶ Kavli Institute for Particle Astrophysics and Cosmology, 452 Lomita Mall Stanford, CA 94305-4085

Received ; accepted

ABSTRACT

The process that quenched star formation in galaxies at intermediate and high redshift is still the subject of considerable debate. One way to investigate this puzzling issue is to study the number density of quiescent galaxies at $z \approx 2$, and its dependence on mass. Here we present the results of a new study based on very deep K_s -band imaging (with the HAWK-I instrument on the VLT) of two HST CANDELS fields (the UKIDSS Ultra-deep survey (UDS) field and GOODS-South). The new HAWK-I data (taken as part of the HUGS VLT Large Program) reach detection limits of $K_s > 26$ (AB mag). We have combined this imaging with the other ground-based and HST data in the CANDELS fields to select a sample of passively-evolving galaxies in the redshift range $1.4 < z < 2.5$ (via the $pBzK$ colour-based selection criterion of Daddi et al. 2004). Thanks to the depth and large areal coverage of our imaging, we have been able to extend the selection of quiescent galaxies a magnitude fainter than previous analyses. This has enabled us to establish unambiguously that the number counts of quiescent galaxies at $z \approx 2$ decline at magnitudes fainter than $K_s \sim 22$ (AB mag.), in contrast to the number density of star-forming galaxies which continues to rise to fainter magnitudes. Through extensive simulations we demonstrate, for the first time, that this turn-over in the number of quiescent galaxies is not due to incompleteness, but is real. We show that this trend corresponds to a stellar mass threshold $M_* \approx 10^{10.8} M_\odot$ below which the mechanism that halts the star formation in high-redshift galaxies seems to be inefficient. Finally we compare the observed $pBzK$ number counts with a population of quiescent galaxies extracted from four different semi-analytic models. We find that only two of these models reproduce even qualitatively the observed trend in the number counts, and that none of the models provides a statistically acceptable description of the number density of quiescent galaxies at these redshifts. We conclude that the mass function of quiescent galaxies as a function of redshift continues to present a key and demanding challenge for proposed models of galaxy formation and evolution.

Key words. Galaxies: high redshift - Galaxies: fundamental parameters -

1. Introduction

The cessation, or “quenching” of star formation activity in galaxies at high and intermediate redshifts is one of the key events in the history of galaxy evolution, and one that is not easily explained in the context of hierarchical growth within a Λ CDM cosmology. It is now understood that the (apparently fairly rapid) termination of star formation in some subset of the galaxy population only a few billion years after the Big Bang must be attributed to some aspect of baryonic physics. However, the relevant process remains a matter of debate, with opinion divided over whether the key mechanism is feedback from an active galactic nucleus (AGN), feedback from star formation itself, or termination of the gaseous fuel supply (e.g. due to the difficulty of cool gas penetrating to the centre of a massive halo due to shock heating). Better observational constraints on the prevalence and properties of quiescent galaxies at high redshift are therefore urgently required in order to develop improved models of galaxy formation and evolution over cosmic history.

One key goal is to better establish the mass dependence of star-formation quenching. A number of lines of evidence suggest that this process appears “anti-hierarchical” in the sense that

it is the most massive galaxies that cease star formation at the earliest times, leading ultimately to the present-day Universe in which the most massive elliptical galaxies appear to be the oldest. However, because less-massive passive galaxies are hard to detect and isolate at high redshift (especially at optical wavelengths), it has proved difficult to establish the definitive mass dependence of the number density of passive galaxies at the crucial epoch corresponding to $z \approx 2$. The primary aim of this study is to clarify this situation via new deep multi-frequency imaging data.

Several methods have been developed to efficiently select passively-evolving galaxies at high-redshift based on colour criteria. In recent years, the most extensively used selection technique is the BzK colour-colour selection introduced by Daddi et al (2004). They showed that galaxies in the redshift range $1.4 < z < 2.5$ fall into a specific area of the $(B - z) / (z - K)$ diagram, and demonstrated that galaxies can be easily separated into star-forming ($sBzK$) and passively-evolving ($pBzK$) classes. This technique also has the advantage of not being biased by the presence of dust.

The reddening vector in the BzK plane is parallel to the BzK selection line, making this criterion relatively immune to

dust content. A number of studies have used the BzK -criterion to constrain the properties of star-forming and passive galaxies at $z \approx 2$. For example, Kong et al. (2006) presented K_s -selected samples of BzK galaxies over two fields: a ≈ 920 arcmin² field (with $K_{s,AB} < 20.8$) and a ≈ 320 arcmin² field (to $K_{s,AB} \approx 21.8$). They particularly concentrated their analysis on the clustering properties of BzK galaxies, and concluded that the $pBzK$ galaxies are more clustered than the $sBzK$ galaxies. Meanwhile, Lane et al. (2007) combined the first release of the near-infrared UKIRT Infrared Deep Sky Survey (UKIDSS) Ultra-Deep Survey (UDS; Lawrence et al. 2007) with optical photometry from Subaru imaging. By comparing the commonly-used selection techniques for galaxies at intermediate redshift, they concluded that the brightest Distant Red Galaxies (DRG) have spectral energy distributions (SED) consistent with dusty star-forming galaxies at $z \approx 2$. Moreover, they observed an interesting turn-over in their derived number counts of $pBzK$ galaxies at $K \approx 21$, suggesting an absence of high-redshift passive galaxies at lower luminosities. Grazian et al. (2007), instead, focused on the overlap between DRG and BzK galaxies, discussing their relative contribution to the overall stellar mass density.

The galaxy population in the UKIDSS UDS field was also studied by Hartley et al. (2008), in order to measure the clustering properties, number counts and the luminosity function of a sample of star-forming and quiescent BzK galaxies with a limiting magnitude $K_{s,AB} < 23$. The number counts they derived for the passive objects exhibit a flattening at $K_s \approx 21$ and an apparent turn-over at $K_s \approx 22$: they concluded that the former effect is likely to be real, while the latter is probable but remains uncertain. More recently, McCracken et al. (2010) presented number counts and clustering properties for a sample of $pBzK$ galaxies with $K_s < 23$, selected over a significantly larger area than previous studies (2 deg² in the COSMOS field). They also found some evidence for a turn-over in the number counts of quiescent galaxies, around $K_s \approx 22$.

Is interesting to notice that similar conclusions were already reached by other papers using different techniques. For instance, De Lucia et al. (2007) used a combination of photometric and spectroscopic data to study the cluster of galaxies evolution over a large range of cosmic time, from redshift 0.4 to 1. They found a significant deficit of low mass, faint red galaxies going to high redshift. A similar conclusion, i.e. the truncation of the red sequence at faint magnitude for high redshift galaxies, was confirmed by other authors as Kodama et al. (2004), and Andreon et al. (2011). More recently, Ilbert et al. (2013) presented galaxy stellar mass function and stellar mass density for star-forming and quiescent galaxies in the redshift range $0.2 < z < 4$ using the new UltraVISTA DR1 data release. They study the stellar mass function for both star forming and quiescent galaxies, and found a lack of evolution of the massive quiescent ones. They interpreted this evidence as a direct consequence of star formation being drastically reduced or quenched when a galaxy becomes more massive than $M > 10^{10.7-10.9} M_\odot$. This result is consistent with the one presented in Muzzin et al. (2013), who also computed the stellar mass function for quiescent galaxies using the Ultra- VISTA data. In general, dedicated analysis of the evolution of “Red&Dead” galaxies show a constant decline with redshift (e.g. Fontana et al. 2009, Brammer et al. 2011)

The use of the BzK selection for the search of passively evolving galaxies has the distinct advantages of being relatively easy to perform on data, of being easily and self-consistently reproducible on mock catalogs derived from theoretical simu-

lation, and relatively well tested with spectroscopic follow-up. However, it is crucial to stress that the success rate of the BzK selection is strongly dependent on the availability of deep imaging at both optical and near-infrared wavelengths. All the above-mentioned surveys were limited both by the depth in the near-infrared, and by the relative depth of the bluer bands. The robust detection of passively-evolving BzK galaxies, requires a proper measure of a colour term as large as $B - K \approx 5$. If the B -band imaging is of limited depth, the clean selection of $pBzK$ galaxies cannot reach the faint K -band limits where the bulk of the population is expected to lie, and the observed number counts are severely prone to errors due to incompleteness and noise. Because of these limitations, the existence of a turn-over in the observed number counts of quiescent galaxies at $z \approx 2$ has yet to be firmly established.

In this paper, we use a combination of wide-field and deep optical and infrared images that allow us to robustly select a sample of passively-evolving galaxies at $z \approx 2$ to $K_s \approx 25$, and to explore the dependence of the number density of these quiescent objects as a function of luminosity or, equivalently, stellar mass. The data have been acquired in the context of the CANDELS HST (Grogin et al. 2011; Koekemoer et al. 2011) and HUGS VLT (Fontana et al. in preparation) surveys, that deliver a unique combination of area and depth in the B , z and K_s bands. With this unique sample we study the number counts of the BzK galaxy population, and in particular we focus on the quiescent population to provide improved constraints on the nature of the physical processes involved in the quenching of star formation.

The paper is organized as follows: in Sec. 2 we describe the data and the multiwavelength catalogue for the UDS and GOODS-S fields, in Sec. 3 we select the star-forming and the passively-evolving galaxies with the BzK criterion, and in Sec. 4 we present the simulations performed to quantify the incompleteness of the observed sample. The interested reader may directly go to Sect. 5, where we present the crucial result of the paper, i.e. the luminosity and mass distribution of $pBzK$ galaxies, and to Sect. 6, where we compare our results with the predictions of various semi-analytical models of galaxy formation. Finally, a discussion of our results and a summary of our main conclusions is presented in Sec. 7.

All magnitudes mentioned in the paper are in the AB system.

2. Data

2.1. Imaging data

The present paper is based on data collected over the CANDELS pointings of the UDS and the GOODS-S fields. Both these ≈ 200 arcmin² fields have been observed in many broad bands in recent years, including WFC3/IR and ACS with the *Hubble Space Telescope* (HST), as a part of CANDELS HST Treasury Project (Grogin et al. 2011; Koekemoer et al. 2011). Additional ultra-deep images from the U to the mid-infrared Spitzer bands are also available. A complete description of the imaging data available in these fields along with the procedure adopted for catalog extraction are given in Galametz et al (2013, G13 in the following) for UDS and in Guo et al. (2013, GUO13 in the following) for GOODS-S, respectively.

Since this work makes a crucial use of the B , z and K bands to select passive galaxies, we briefly summarize here the properties of these images, including new imaging data

in the B and K bands that have been recently collected with VLT and are not included in the papers mentioned above.

In the K band, the two fields have been imaged by the High Acuity Wide field K-band Imager (HAWK-I) on the European Southern Observatory’s Very Large Telescope (VLT) as part of the HUGS survey (an acronym for HAWK-I UDS and GOODS-S survey). This program has delivered very deep images in the K_s band over both fields, reaching $5\text{-}\sigma$ detection limits fainter than $K_s \simeq 26$. We refer to Fontana et al. (in preparation) for a complete description of the HUGS survey. The K_s band of HAWK-I will be referred to simply as the K -band in the rest of this paper.

For UDS, the final K band image is already included in the G13 catalog. The whole UDS field is covered with three different pointings. Their seeing is $0.37''\text{--}0.43''$ and the corresponding limiting magnitudes are $m_{lim}(K) \simeq 26$, $m_{lim}(Y) \simeq 26.8$ (5σ in one FWHM) or $m_{lim}(K) \simeq 27.3$, $m_{lim}(Y) \simeq 28.3$ (1σ per arcsec²). In GOODS-S, we use here the final HUGS image, that is considerably deeper than the image used in GUO13. The whole field was covered with 6 (partly overlapping) pointings for total exposure time in the K band (summed over the six pointings) of 107 hours. Because of the complex geometry, this corresponds to an exposure of 60–80 hours in the central area (the one covered by CANDELS “Deep”) and 12–20 hours in the rest (the CANDELS “Wide area”). The final average seeing is remarkably good and constant, with 4 pointings at $0.38''$ (notably including the two deepest) and 2 pointings at $0.42''$. On the finally stacked images, the limiting magnitudes in the deepest area are $m_{lim}(K) \simeq 27$. (5σ in one FWHM) or $m_{lim}(K) \simeq 28.3$, $m_{lim}(Y) \simeq 28.3$ (1σ per arcsec²).

The B and z' images in the UDS were obtained with Subaru/Suprime-Cam (see Furusawa et al. 2008). The Subaru optical data cover the full CANDELS UDS field, and the $5 - \sigma$ limiting magnitudes (computed in 1 FWHM of radius) are 28.38 and 26.67 in B and z' respectively (G13).

Another crucial data set in this context is a recent ultra-deep B-band image obtained with VIMOS on VLT for GOODS-S (Nonino et al in prep.). Their average seeing is $0.89''$. With a VLT/VIMOS total integration time of 24hr, the final mosaics have a median PSF of 0.85 arcsec FWHM, and reach $B = 28.4$ at 5σ , and are thus considerably deeper than the available ACS $F435W$ data (the latest V2.0 version released by the GOODS Team reaches a $5 - \sigma$ limit of $z_{850} \simeq 26.9$). We will use in the following this VIMOS B-Band image.

While B , z and K bands are crucial for the selection of passive galaxies, we have also utilised the full multi-colour data from the U to the mid-infrared Spitzer bands at various stages in our analysis, both to derive photometric redshifts and to determine the shapes of the spectral energy distributions (SEDs) of the selected galaxies. We refer to G13 and GUO13 for a complete description of the imaging data set.

2.2. Catalogs

The catalogs used in this paper are those already published in G13 and GUO13, with the only notable difference that we have included the new B and K images, as described below.

The adoption of the CANDELS catalogs implies that we rely on the H_{160} band for the detection of galaxies. This choice has the practical benefit that we have been able to use the existing CANDELS catalogues, and takes full advantage of the depth and the quality of the HST images. Because

the H_{160} imaging is still significantly deeper than the K -band imaging, this choice does not affect the definition of the K -selected sample; we have performed an independent detection in the K band and verified that all the objects detected in K are also detected in H -band, at least to the magnitude limits we are interested in here.

After removing from our catalogs bad sources (i.e. those not falling on an edge, artifact, star spike or halo) the final UDS sample contains 35933 objects over an area of 201.74 arcmin², while the GOODS sample contains 34932 objects over an area of 137.61 arcmin² (total number are similar since GOODS is about one magnitude deeper than UDS).

Colours in the ACS and WFC3/IR bands have been measured running SExtractor in dual-image mode, using isophotal magnitudes (MAG_ISO) for all the galaxies, after smoothing each image with an appropriate kernel to reproduce the resolution of the H_{160} WFC3/IR image. For all images with poorer resolution, including the B-VIMOS, z' and K used here, the Template-FITting photometry software TFIT (Leider et al. 2007) was used to derive the photometry for all the images with a lower resolution than H_{160} e.g. B , z' and K . This code uses information (position, profile) of sources measured on a high-resolution image (here H_{160}) as priors to establish the photometry in the lower-resolution images. We refer to G13 for a more detailed discussion of the CANDELS UDS multiwavelength catalog. For the new B and K images used here we have applied TFIT exactly as described in G13 and GUO13.

We remark that the TFIT code that we used to measure colors is designed to minimize the effects of blending from nearby sources when extracting the photometry of faint sources in crowded extragalactic fields. The power of TFIT in this context is clearly shown by the simulations presented in Lee et al (2012), where it is shown that the contamination from nearby sources is minimal even in the case of the $\approx 3''$ PSF of the Spitzer images. The ground-based images on which this analysis is based upon (BzK) have PSF much smaller, especially the K band ones that are all below $0.45''$, and we therefore expect possible contaminations to be even smaller. As a further check, we have verified that only 5% of galaxies at $K \simeq 23.5\text{--}24$ (the most typical objects that we target) have a nearby brighter companion within $2''$, that may be affecting the measured colors. We are therefore confident that our results are not significantly affected by photometric contamination.

We also note that the images used in this analysis are remarkably homogeneous in depth over the area that we use. Ground-based B and Z images have been obtained with imagers with FOV larger than the CANDELS fields, and the dithering pattern adopted for Hawk-I has ensured a proper coverage of the full area. We have explicitly verified this by looking at the distributions of the $1\text{-}\sigma$ magnitude limits in the B, z and K images for our BzK galaxies, and we verified that they are reasonably narrow gaussian distributions. The remaining inhomogeneities are accounted for in the simulations that we used to evaluate the systematics in our analysis.

2.3. The SED fitting

While the selection of passive galaxies is performed using only the BzK bands, we exploit the full multiwavelength catalogs delivered by CANDELS to obtain further informations

Table 1. Parameters used for the library of template SEDs.

IMF	Salpeter
SFR τ (Gyr)	0.1, 0.3, 0.6, 1, 2, 3, 4, 9, 15
$\log(\text{age})$ (yr)	7, 7.01, 7.03...10.3
Metallicities	0.02 Z_{\odot} , 0.2 Z_{\odot} , 1 Z_{\odot} , 2.5 Z_{\odot}
E_{B-V}	0, 0.03, 0.06, 0.1, 0.15, 0.2, ..., 1.0
Extinction law	SMC, Calzetti

on the targeted galaxies, like photometric redshifts, stellar masses and other rest-frame quantities.

For all the BzK galaxies without a secure spectroscopic redshift we used the official photometric redshift published by the CANDELS collaboration. The technique adopted is described in Dahlen et al (2013), and photometric redshifts for both fields will be made available in forthcoming papers (Dahlen et al, in prep.). On the sample of BzK galaxies with secure spectroscopic redshift we find that the scatter between photometric and spectroscopic redshift is gaussianly distributed, with a standard deviation $\sigma \sim 0.08$ and a small number of outliers, about 5%.

To compute the physical parameter of the galaxies we used SED fitting technique on the full multiwavelength catalogue available for the two CANDELS fields, that have 18/19 bands on UDS/GOODS, from U band to $8\mu\text{m}$. The SED technique employed for this work has been already intensively tested in previous papers (Fontana et al. 2003, 2004, 2006; Grazian et al. 2006, Santini et al. 2012). It is based on the comparison between the observed multi-colour SED of each galaxy with those obtained from a set of reference synthetic spectra from stellar population model. The redshift of each galaxy is fixed to the spectroscopic or the photometric one during the fitting process.

We decided in this case to use a pretty standard reference spectral library, that can be compared with previous literature. It is based on the Bruzual & Charlot (2003) models, and the code was run using a Salpeter IMF, ranging over a set of metallicities from $Z = 0.02 Z_{\odot}$ to $Z = 2.5 Z_{\odot}$, and dust extinction A_V ranging between 0 and 1.1 assuming a Calzetti extinction curve. We used a smooth exponentially decreasing Star Formation Histories. All the parameters adopted for the SED fitting are listed in Tab. 1. The best-fitting spectrum provides estimates of the star-formation rate (SFR), extinction, stellar population age (the onset of the star formation episode) and galaxy stellar mass, and τ (the star formation e-folding timescale).

Among the outputs of this analysis we will use in particular the stellar masses and the age/τ parameter, that is the inverse of the Scalo parameter. As shown in Grazian et al. (2007) it can be used to discriminate between star forming and passively evolving galaxies. Following their work, we assume $\text{age}/\tau < 4$ for the former and $t/\tau \geq 4$ for the latter.

3. BzK galaxies in CANDELS

3.1. Raw BzK counts

As first proposed by Daddi et al. (2004), the BzK colour-colour criterion is an efficient colour-based method for identifying galaxies in the redshift range $1.4 < z < 2.5$. It also permits segregation of galaxies between actively star-forming galaxies ($sBzK$) and passively-evolving galaxies ($pBzK$). If we define $BzK = (z - K) - (B - z)$, $sBzK$ galaxies have colours consistent

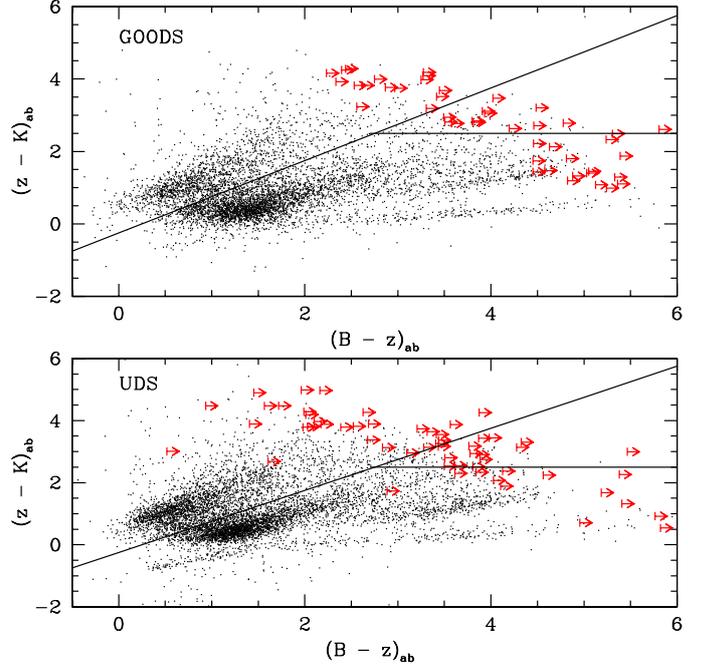


Fig. 1. $B - z$ versus $z - K$ colour-colour diagram for the K -selected galaxies and stars in the CANDELS GOODS-S (upper panel), and UDS (lower panel) fields. Arrows indicate 1σ lower limits on colour for sources which are undetected ($S/N < 1$) in the B -band. The quiescent galaxies are indicated in red, with the star forming galaxies indicated in blue, in the redshift range between $1.4 < z < 2.5$ (as classified using the t/τ parameter as discussed in the text). The green arrows show colour limits for sources that have a photometric redshift outside of the range expected for effective use of the BzK criterion.

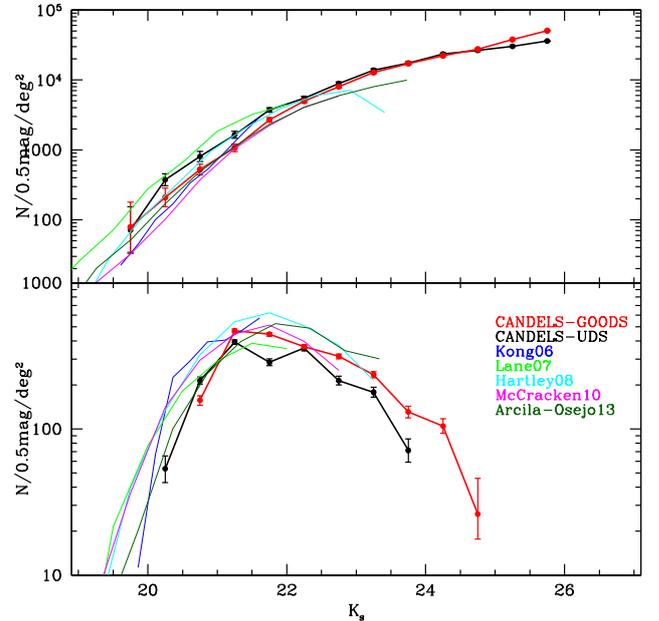


Fig. 2. Number counts for the $sBzK$ (upper panel) and $pBzK$ (bottom panel) galaxies in the two CANDELS fields plotted separately, red GOODS-S and black UDS with $K < 27$. We assume Poissonian error bars. Previous results from the literature are over-plotted: Kong et al. (2006) in blue, Lane et al. (2007) in green, Hartley et al. (2008) in cyan, McCracken et al. (2010) in magenta, and Arcila-Osejo et al. (2013) in dark-green.

with $BzK > -0.2$ while $pBzK$ galaxies have $BzK < -0.2$ and $(z - K) > 2.5$ (corresponding to the upper-left and upper-right regions of the colour-colour diagram respectively; see Figure 1).

To facilitate comparison of our results with previous studies we wanted our photometric selection criterion to match as closely as possible the original BzK selection introduced in Daddi et al. (2004). Since our filters in GOODS-S are not identical to those used in the original BzK paper, we applied a correction term to the colours to account for the different shape of the filters. Correction terms were computed using population synthesis models (Bruzual & Charlot 2003) applied to both our filters and those of Daddi et al. (2004). The result of this analysis was that we applied a small colour correction to $(B - z)$ and $(z - K)$ for the objects in GOODS-S field: $(B - z)_{Daddi} = (B - z)_{GOODS-S} + 0.1$ and $(z - K)_{Daddi} = (z - K)_{GOODS-S} - 0.05$. No correction has been applied to the colours in the UDS field because the filters are nearly identical to those utilized in the standard BzK definition.

In total, we identify 99 $pBzK$ galaxies in the UDS and 86 in GOODS-S. The number counts for star forming and passive BzK galaxies are reported in Table 2 and Table 3, respectively. Figure 1 shows the $B - z$ versus $z - K$ distribution of sources in the CANDELS GOODS-S (upper panel) and UDS (lower panel) fields. Horizontal arrows indicate objects which are non-detected in the B -band ($S/N < 1$). We assume for these sources a lower limit on B -band magnitude equal to their 1σ limiting magnitude. All magnitudes are detected in the z -band images. In principle, BzK galaxies undetected in the B band and located in the $sBzK$ region could be $pBzK$ galaxies scattered left-ward into the $sBzK$ region, as shown also in Grazian et al. (2007), which therefore escape inclusion in the $pBzK$ sample. In sec. 4 we analyze this population using the output of the SED fitting technique to verify the nature of these sources.

Figure 2 shows the raw number counts derived for the both $sBzK$ and $pBzK$ galaxies in the CANDELS fields with $K < 27$ for both fields separately, UDS black and GOODS-S red line, compared to the literature (0.5 magnitude bins). The error bars are computed assuming a Poisson statistics, approximated with the square root of the counts for $N > 10$ and the small-number approximation for the Poisson distribution of Gehrels (1986) for $N \leq 10$. Our number counts are in good agreement with most of the previous studies, especially with the latest results from McCracken et al (2010). However, we find a discrepancy with the number counts presented by Lane et al. (2007). As already proposed by McCracken et al. (2010), this difference could be due to an inappropriate transformation to the Daddi et al. filter set.

It is immediately clear that our observations reach a depth that no previous studies has achieved. It is also very evident the decrease in the number counts of $pBzK$ galaxies at magnitudes fainter than $\approx 21.5 - 22$. To establish whether this turnover is real we have performed dedicated simulations that are described in the following section.

Our data also show a significant different between the two fields, especially for bright $sBzK$ galaxies and for all $pBzK$ galaxies. Since we have already corrected for the slightly different filters, we interpret this difference as an effect of cosmic variance. A simple estimate of the error budget due to cosmic variance, as obtained with the Cosmic Variance Calculator (Trenti & Stiavelli 2008) shows that indeed this is the case. For instance, the total number densities of bright ($K \leq 22$) $sBzK$ galaxies is $\approx 4600 \pm 870$ galaxies arcmin^{-2} in the GOODS/S field, and $\approx 6640 \pm 1020$ galaxies arcmin^{-2} in UDS, respectively. Similarly, the densities of all

Table 2. Observed Number Counts in 0.5mag bins for $sBzK$ and $pBzK$ galaxies in the UDS and GOODS-S fields.

K_{AB}	$N_{sBzK-UDS}$	$N_{sBzK-GOODS}$	$N_{pBzK-UDS}$	$N_{pBzK-GOODS}$
19.75	4	3	-	-
20.25	21	8	3	0
20.75	45	20	12	6
21.25	92	41	22	18
21.75	209	103	16	17
22.25	305	190	20	14
22.75	498	307	12	12
23.25	775	487	10	9
23.75	975	655	4	5
24.25	1318	848	-	4
24.75	1483	1050	-	1
25.25	1697	1446	-	-
25.75	2012	1934	-	-

$pBzK$ are $\approx 2150 \pm 500$ galaxies arcmin^{-2} and $\approx 1770 \pm 380$ galaxies arcmin^{-2} in GOODS-S and UDS, respectively. The different densities are therefore consistent within the errors, when we properly include the cosmic variance effects in their computation. We note instead that the difference at very faint magnitudes ($K > 26$) between UDS and GOODS is likely due to incompleteness in the detection, since UDS approaches the detection limit just at $K \approx 26$.

3.2. The redshift distribution of BzK galaxies

Several of the CANDELS BzK galaxies have published spectroscopic redshifts (35 in the UDS and 87 in GOODS-S). Figure 3 shows the distribution of these sources with a spectroscopic redshift in a BzK diagram; each symbol refers to a different redshift range. The insert box shows the distribution of the spectroscopic redshifts of the BzK -selected galaxies. The total number of $pBzK$ with spectroscopic redshift is 14.

According to the Daddi et al. (2004) definition, BzK galaxies are expected to lie in the redshift range $1.4 < z < 2.5$ (shown by the vertical lines). Figure 3 shows that a large fraction of the BzK -selected galaxy population does indeed fall within this redshift range. Nevertheless, even though the $pBzK$ galaxies with spectroscopic redshifts are few in number, we find that 18% of $pBzK$ galaxies have a spectroscopic redshift that lies outside this redshift window, with two tails extending to $z \approx 1.1$ and $z \approx 2.9$.

In Fig. 4 we present the photometric redshift distribution for both star-forming (upper panel) and quiescent (lower panel) galaxies. From this figure it is easily appreciated that most of the $pBzK$ galaxies have photometric redshifts below $z = 2$, with a $z_{med} \approx 1.8$, while the distribution of $sBzK$ spans a wider range with $z_{med} \approx 1.9$. The redshift distribution of the passively-evolved galaxies is in good agreement with the one presented in McCracken et al (2010). We note that the peak of galaxies around $z \sim 1.6$ could be related to large-scale structures known in both UDS (Papovich et al. 2010; Tanaka et al. 2010) and GOODS-S fields (Castellano et al. 2007; Kurk et al. 2008).

4. Estimating the systematic effects

As evident by looking at Fig 2, the number counts of $pBzK$ galaxies drop below $K \approx 21.5$. Since this is the main focus of the present paper, before analyzing it in more detail it is important to exclude that this behavior is due to systematics in the data. As already discussed in Sect. 2, we can exclude that blending is a major cause of photometric errors, or that image inhomogeneities

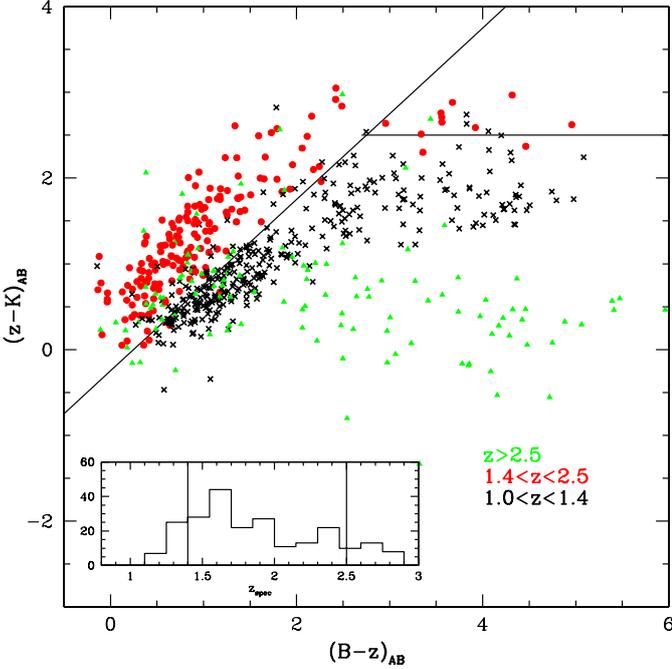


Fig. 3. BzK plot for all CANDELS galaxies with a spectroscopic redshift (35 for UDS and 87 for GOODS-S); each symbol refers to a different redshift range: red points $1.4 < z < 2.5$, black crosses $1.0 < z < 1.4$, and green triangles $z > 2.5$ (different colours in the electronic version). The spectroscopic redshift distribution for the BzK galaxies is shown in the box insert. Vertical lines delineate the standard anticipated redshift range of the BzK selection technique ($1.4 < z < 2.5$).

are a major source of incompleteness. It is well possible, however, that this effect is - at least partially - due to the insufficient depth of the B band images, give the large color terms that must be measured to unambiguously identify the $pBzK$ galaxies. To further verify this, we performed accurate simulations that are described below.

4.1. Incompleteness

The effects of noise and incompleteness can impact in various ways the observed distribution of galaxies in the BzK plot. First, photometric noise may scatter galaxies within the BzK diagram: in particular galaxies that lie close to the lines that define the selection criteria can migrate in or out of the selection windows. For $pBzK$ galaxies this may be particularly important in the case of the horizontal line at $z - K > 2.5$. In our case we can take advantage of the exquisite depth of the CANDELS and HUGS data to keep these effects to a minimum. Even at $K = 24$, the typical error on the K -band magnitude is < 0.05 mag, and the colour error on $z - K = 2.5$ is measured with a total error of typically 0.06 mag. Similarly, the 1σ detection limit in the B -band is $B \approx 30$ in both fields, ensuring detection also in the B -band for most of the $pBzK$ galaxies down to $K = 24$. **On the other hand, we confirm that all the $pBzK$ galaxies have been detected in z -band.** In order to quantify the number of missed objects in our samples, we perform a number of simulations. The ultimate goal is to define a robust limiting magnitude in the K -band, and to correct *a-posteriori* the observed number counts for incompleteness.

We first start from the observed sample of bright $pBzK$ galaxies selected with $K < 20.5$. We use these objects to generate mock catalogues of 5×10^5 objects for each field, normalized at various K -band magnitudes from $K = 21$ to $K = 25$. The

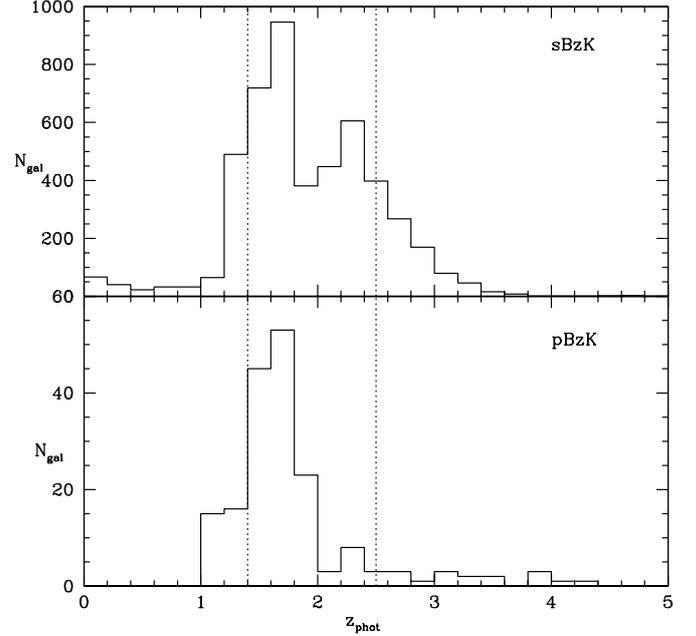


Fig. 4. The photometric redshift distributions of the star-forming (upper panel) and passively-evolving (lower panel) BzK galaxies found in the two CANDELS fields. The dashed lines indicate the anticipated redshift range yielded by the BzK criterion, while the solid line marks the average photometric redshift of each distribution.

computed magnitudes (in all bands) are then perturbed assigning them a noise consistent with the observed error-magnitude relation in each band and field. Our simulations accurately reproduce not only the average S/N as a function of magnitude, but also its scatter, in order to provide the closest match possible between the observed and simulated data (Castellano et al. 2012). The simulated catalogues are then analyzed in a similar way to the ones derived from the observations. Fig. 5 shows the BzK colour-colour diagram for the magnitude bins over which we run the simulations. The blue dots are the simulated samples of $pBzK$ galaxies (as obtained from scaling the observed bright objects to fainter fluxes) and the red horizontal arrows are the objects undetected at 1σ in the B band, similar to those encountered in the actual data. The two numbers in the corner of each window are the limiting magnitude used to run the simulations and the number of simulated $pBzK$ galaxies found respectively.

We have performed a similar test starting from a set of synthetic galaxies computed using the standard spectral synthesis models of Bruzual & Charlot (2003), over the redshift range $1.4 < z < 2.5$. We consider models with exponentially-declining star-formation histories, metallicity $Z/Z_{\odot} = 0.02, 0.2, 1.0$, extinction $0 \leq E(B - V) \leq 0.2$, and ages $1 \leq \text{Age} \leq 13$ Gyr. We assume a Salpeter IMF and a Calzetti et al. (2000) attenuation law. We use this library to generate mock catalogues normalized at various K -band magnitudes from $K = 21$ to $K = 25$.

We find good consistency between the simulated objects obtained in this way and the mock catalogues discussed above. **To correct the observed number counts we used the numbers coming from the mock catalogue. We note that using the simulations from the observed catalogue we obtain the same result.** In fact in both cases, we find that the completeness obviously decreases with increasing magnitude. At $K \sim 23.75$ it is $\sim 73\%$ in the latter case, and $\sim 79\%$ in the former simulations. At $K \sim 25.25$ it is slightly above 40%, in both cases. Based on

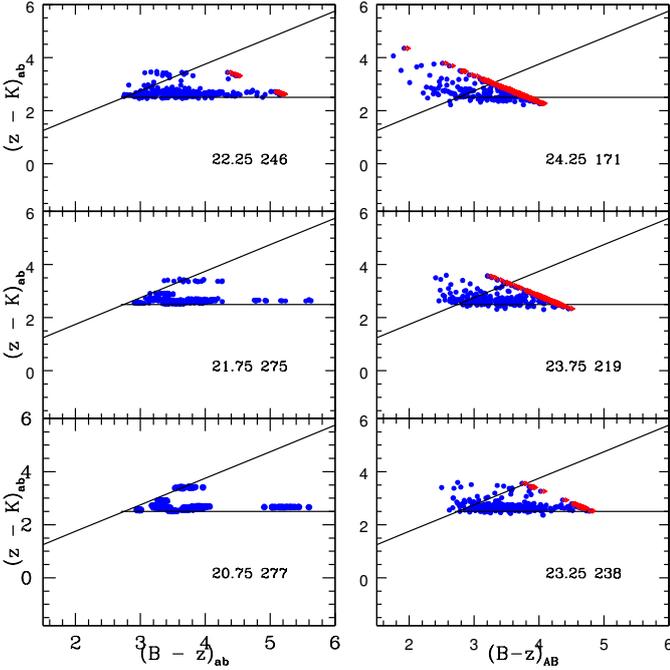


Fig. 5. The $B - z$ versus $z - K$ colour-colour diagram of the simulated sample of passively-evolving galaxies in eight magnitude bins. Each window contains a sample of simulated $pBzK$ galaxies. The number of BzK galaxies in the simulations is ~ 10000 . The labels in each panel precise the limiting magnitudes adopted in the simulations followed by the number of galaxies classified as $pBzK$. **The red horizontal arrows are the $1-\sigma$ upper limits in B -band**

these results, we conclude that incompleteness is safely low up to $K \simeq 24$, and still treatable at $24 \leq K < 25$. In the following we will therefore extend our analysis to $K < 25$. From both these simulations we are able to quantify the number of missed objects in each magnitude bin, and convert it into the incompleteness. We then apply corrections to the distribution of the $pBzK$ number counts that we will present in the following sections.

To confirm the purity of our sample we check if some galaxies may be scattered in the $pBzK$ region due to photometric errors. To investigate this effect we perform some simulations, as described above, but in this case starting from a bright catalogue of non quiescent galaxies place just below the $(K-z)$ threshold, and adding noise to all the photometric bands. We found that the contamination of spurious objects are less than 1% at faint magnitude. Therefore, we confirm that, thanks to our conservative selection, the observed scatter is negligible.

4.2. Spectral classification

We have also compared the galaxies selected with the simple colour selection criteria to the output of a full SED analysis, performed on the multi-wavelength catalogs obtained from CANDELS. To understand the limitations of the BzK selection criterion, derived only from observed colours, we follow a different approach based on the spectral fitting technique. The adopted spectral fitting technique has already been described in Sec. 3.2.

One simple test to verify the validity of the BzK criterion is to analyze the population of BzK galaxies by correlating the age, the τ parameter and the $E(B-V)$. As in Grazian et al. (2007), for each galaxy we compute the t/τ parameter, (where t is the galaxy

age) and we characterize as star-forming the galaxies with $t/\tau < 4$, and as passively-evolving those with $t/\tau \geq 4$.

Therefore, we first verify that the classification based on the t/τ parameter is consistent with the results obtained using the BzK criteria, i.e. we compute this parameter for all the BzK galaxies, and for all of them we examine the correlation between t/τ and the extinction. We verify that both star-forming and quiescent galaxies have the expected value of t/τ , as well as for the extinction. Then, we have looked at the properties of the objects undetected in the B band. As shown in Fig. 1, those located in the $pBzK$ region are consistently fitted with passively evolving models, while those located in the $sBzK$ area are fitted with star-forming models. This confirms our conclusion that a relatively low number of $pBzK$ galaxies are scattered out of the $pBzK$ region because of limited depth of the B -band imaging.

5. Passively evolving BzK galaxies: luminosity and mass distribution

Armed with full characterization of the selection function, we are now in a position to explore the properties of the $pBzK$ galaxies. Figure 6 presents the K -band number counts (in 0.5 magnitude bins) for the $pBzK$ galaxies in the CANDELS fields. We re-emphasize that the galaxies were selected with $K < 24$, and the error bars are computed as described above. The observed number counts show a flattening at $K > 21$, consistent with past studies (e.g. Hartley et al. 2008), and a turn-over at $K \simeq 22$. A decline in the number density of the quiescent population at faint magnitudes was also observed in the H_{160} -band by Stutz et al. (2008).

For comparison, we also overplot the number counts found by Kong et al. (2006), Lane et al. (2007), Hartley et al. (2008), and McCracken et al. (2010). The number counts are in good agreement with previous studies at faint magnitude. **However, we note that at bright magnitude there is a discrepancy with the previous works, especially with McCracken et al. (2010) and Kong et al. (2006). Since the number counts for the SF galaxies are in good agreement with these works, we can rule out that this discrepancy arises from photometric errors. This difference may be the result of cosmic variance, given the strong clustering of $pBzK$ s galaxies respect to the $sBzK$ ones.**

The $pBzK$ galaxy number counts derived from our CANDELS (UDS+GOODS-S) sample are summarized in Table 3.

It is immediately clear that the number counts for the $pBzK$ galaxies shows a flattening at $K \simeq 21$ and a turn-down at $K > 22$. Converting these magnitudes to average rest-frame magnitudes in the I band, (that is sampled by the K band at $z \simeq 1.9$, close to the average redshift of $pBzK$ galaxies with spectroscopic redshift) this corresponds to absolute magnitudes of $M_I \simeq -23$ and $M_I \simeq -22$, respectively, as shown by the upper x-axis in Figure 6.

To further investigate the physical significance of this trend, we translate the galaxy photometry into stellar mass. We compute the stellar masses using our SED fitting to the full multi-wavelength catalogues available for the two fields. The adopted spectral fitting technique is the one already mentioned in Sec 4.2.

We note in passing that the superior quality and spectral extension of the CANDELS imaging data gives us the opportunity of revising the average relation between the K -band magnitude and the stellar mass derived by Daddi et al. (2004) for BzK galaxies with K -magnitude less than 20. This was calibrated on the stellar mass estimates derived from full

SED fitting in the K20 spectroscopic sample (Fontana et al. 2004), where exactly the same numerical code was used to estimate masses, but photometry was inevitably poorer. Daddi et al. (2004) derived a relation between stellar mass and the observed K magnitude of the form:

$$\log(M_*/10^{11} M_\odot) = -0.4(K^{tot} - K^{11}),$$

where K^{11} is the K -band magnitude corresponding on average to a mass of $10^{11} M_\odot$. For the SED fit they found $K^{11} = 21.35$ (AB). We recalibrate the relation based on our sample and find $K^{11} = 21.53$ (AB). We note that this relation works better at fainter magnitude than the one presented in Daddi et al. (2004).

Figure 6 shows the distribution of the stellar masses for the pBzK galaxy sample computed with the SED fitting technique.

For comparison, we overplot the results obtained by Grazian et al. (2006). They computed the stellar masses for the GOODS-MUSIC sample using the SED fitting technique. We compare their distribution with one shown in Figure 6, purple dotted line, to test the goodness of the assumptions for the SED fitting assumed in this work. We confirm the consistency of the two results.

The observed distribution shows that there is a clear decrease in the number density of passively evolving galaxies at stellar masses below $10^{10.8} M_\odot$. This trend departs significantly from the overall form of the galaxy stellar mass function at this redshift, which continues to rise steeply to much lower masses (e.g. Ilbert et al. 2013). We stress that the fraction of pBzK galaxies compared to all BzK galaxies (and hence to most of the galaxies at $z \approx 2$ is large ($\sim 25\%$) at $M_* \approx 10^{11} M_\odot$) and becomes minimal ($\sim 1\%$) at $M_* \approx 10^{10} M_\odot$.

6. Comparison with semi analytical models

In the previous sections we have highlighted the existence of a clear break in the luminosity and mass distribution of passive galaxies at $z \approx 2$, with these objects becoming progressively rarer beyond an observed K -band magnitude of $K \approx 22$. This turnover in the number counts can be translated into a mass threshold roughly placed around $10^{10.8} M_\odot$. This implies that the mechanisms that halted star formation in galaxies at high redshift have been more efficient (or statistically more frequent) in massive galaxies rather than in lower mass ones. It is interesting to investigate whether this basic feature is reproduced by theoretical models of galaxy formation.

We concentrate on four Semi-Analytical Models (SAMs): Menci et al. (2008), Merson et al. (2012) (based on the SAMs of Bower et al. 2006), Somerville et al. (2012), and Lu et al. (2011,2012).

These models vary both in the way they assemble the dark matter halos as well as in the prescriptions adopted for the various physical mechanisms involved in galaxy formation. Here we briefly mention the physical processes relevant in the discussion of our result. We refer the reader to the original papers for full details of the models. In the Menci model the merging histories of dark matter haloes are described through Monte Carlo simulations, the Merson model uses dark matter merger histories extracted from the Millenium simulation (Springel et al. 2005), while the Somerville and Lu models use the Bolshoi N-body simulations (see Klypin et al. 2011 for details). The models based on N-body simulations provide lightcone mock catalogues that mimic the geometry of the UDS and GOODS-S fields. We

Table 3. Differential Number Counts in $\log(N/\text{deg}^2/0.5\text{mag})$ bins for sBzK and pBzK in the (UDS+GOODS-S) fields. The last column shows the corrected pBzK counts.

K_{AB}	N_{sBzK}	N_{pBzK}	$N_{pBzK}^{Corr\ a}$
19.75	1.871	-	-
20.25	2.488	1.503	1.503
20.75	2.839	2.281	2.336
21.25	3.150	2.628	2.701
21.75	3.520	2.544	2.637
22.25	3.720	2.557	2.646
22.75	3.931	2.406	2.531
23.25	4.127	2.304	2.451
23.75	4.238	1.980	2.195
24.25	4.361	1.628	1.892
24.75	4.429	1.026	1.642
25.25	4.523	-	-
25.75	4.623	-	-
26.25	4.663	-	-
26.75	4.590	-	-
27.25	4.324	-	-
27.75	3.849	-	-
28.25	3.175	-	-
28.75	2.473	-	-

Notes. ^(a) The incompleteness in the pBzK counts was computed by performing extensive simulations; see main text for details.

note that the resolution of these N-body simulations in practice sets a lower limit to the mass (hence luminosity) of the smaller galaxies traced in the simulations, and hence to the depth of the luminosity distributions that we are presenting. All four models distinguish between quiescent star formation in galaxy discs, and merger-driven starbursts, although with different prescriptions for the star formation efficiency. The Merson model includes disc instabilities as an extra mechanism to trigger bursts of star formation. Different modes of AGN feedback are implemented in the SAMs: quasar mode (Menci, Somerville), radio mode (Merson, Somerville) and halo quenching model (Lu).

All these models provide simulated galaxy samples, for which magnitudes in any desired filter set are given. Galaxy magnitudes are computed from the predicted star formation and chemical enrichment histories using single stellar population model (SSP). All these models use the SSP model of Bruzual & Charlot (2003) but they assume different initial mass function (IMF): Kennicutt IMF (Merson), Salpeter IMF (Menci), Chabrier IMF (Somerville and Lu). The dust extinction affecting the above magnitudes is computed from the dust optical depth and applying the appropriate attenuation to the luminosity at various wavelengths. All these models are able to compute synthetic galaxy catalogs over given areas of the sky, that have been used for this comparison. In our case, we have extracted the BzK magnitudes and applied exactly the same BzK criterion we applied to real data, as described in the previous sections.

Fig. 7 shows the comparison of the SAMs with the observations: the left-hand panel shows the K number counts of pBzK galaxies, and the right-hand panel compares the mass distributions. We first note that two of these models, Merson et al. (2012) and Somerville et al. (2012), exhibit a global trend that is *qualitatively* consistent with the observed break in the luminosity and mass distributions. Their predicted number density reaches a peak at masses slightly below $10^{11} M_\odot$, similar to what we observe in the data, and decreases at lower masses. However,

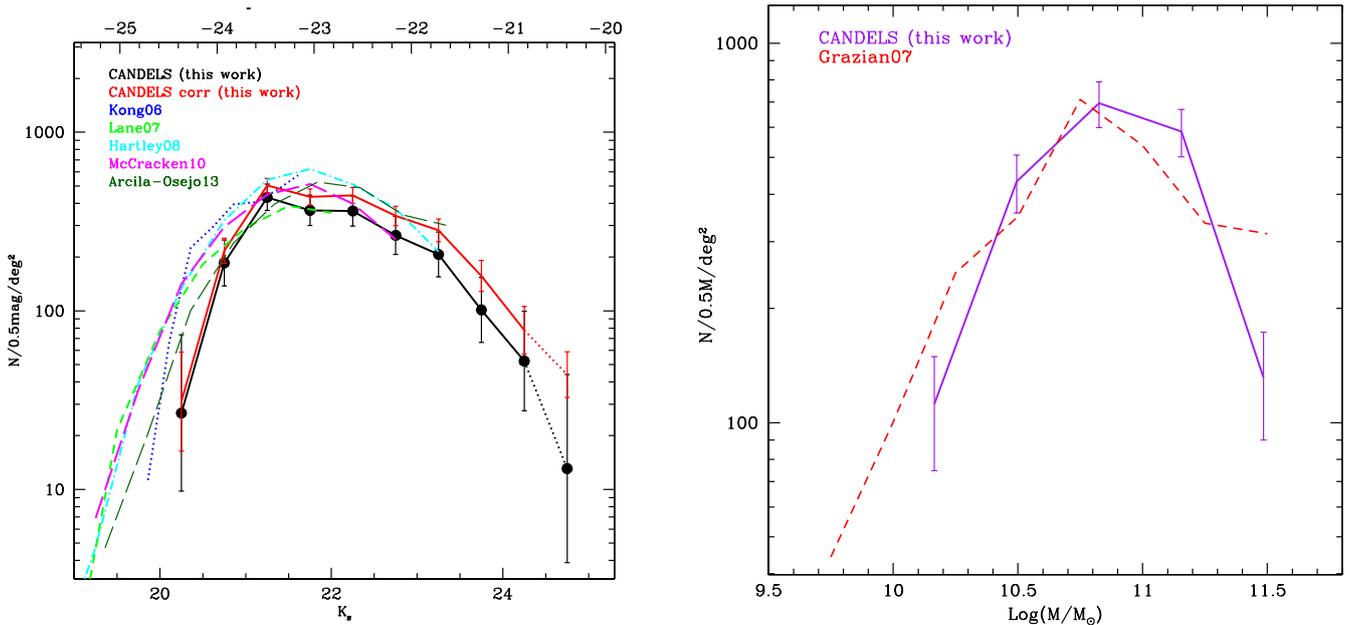


Fig. 6. Left: The number counts of the CANDELS $pBzK$ galaxies (black line with Poissonian errors, see the text.) as a function of the observed K magnitude. The upper x-axis shows the corresponding rest frame absolute magnitude in the I band at $z \approx 2$. **As highlighted in text, for completeness reason, we prefer to limit our analysis at $K < 24$, however we plot the faint end of the distribution until $K = 25$ with a dotted line.** The number counts corrected for incompleteness are shown in red. The blue, green, cyan, magenta and darkgreen lines show the number counts for $pBzK$ galaxies from the literature, from Kong et al. (2006), Lane et al. (2007), Hartley et al. (2008), McCracken et al. (2010), and Arcila-Osejo et al. (2013) respectively. Right: The stellar mass distribution of the CANDELS $pBzK$ galaxies in units of M_{\odot} (logarithmic scale). The black solid line shows the mass distribution computed using only the K -band (Daddi et al. 2004), while the purple dashed line is the result of using the masses obtained with the multi-band SED-fitting technique. For comparison we plot the galaxy stellar mass distribution of the $pBzK$ derived by Grazian et al. (2007) from the GOODS-MUSIC sample (red dotted line).

they do not reproduce the absolute number density of the passive galaxies: the Merson model over-predicts (by a factor of ~ 1.5) the observed counts, while the Somerville model underpredicts them by a similar amount. By contrast, the Menci and Lu models do not even predict a turn-over; in both cases the number counts continue to increase exponentially towards fainter magnitudes (i.e. to lower stellar masses). Both these models deliver an underestimate of the number of massive passive galaxies, and a continuously increasing overestimate at the faint/low-mass end.

Obviously, the counts predicted by the models depend sensitively on the exact choice of the adopted color selection, but the discrepancies found here are clearly larger than can be accounted for by simple noise or systematic effects. For instance, decreasing the threshold on the $(z - K) > 2.5$ color may significantly change the number of quiescent objects found in the models. In particular, to reproduce the density of the observed $pBzK$ galaxies, we would require to lower the $(z - K)$ threshold by 0.2 mag. This is much larger than the typical error on $(z - K)$ color, of about 0.03 mag.

Based on this simple comparison, we derive a first important conclusion: the luminosity/mass distribution of passive galaxies at $z \approx 2$ is clearly a very sensitive and demanding test for hierarchical models.

The importance of this finding - by itself - should not be underestimated. Indeed, current models of galaxy formation are able to provide acceptable fits to the observed properties of the bulk of the galaxy population, such as luminosity functions or color distributions. What we show here is that the physical mechanisms that they implement are not able to ac-

count for the extreme star formation histories characterizing the $pBzK$ galaxy population.

We simply mention that it is not possible to interpret the differences in model predictions in term of the AGN feedback, that is often identified as a major actor in the quenching of star-formation activity, and is suspected to be more effective in massive galaxies. Indeed both the model of Lu et al. (in prep.), in which the AGN feedback is implemented with high efficiency, and the Menci et al (2005) model, where AGN feedback is not so efficient, produce in an over-estimation of the total number of red, quiescent objects.

Locating the origin of the discrepancy between observed and predicted quantities, as well of the differences among the various theoretical predictions requires a detailed comparison among different models, not simply on their output but directly modifying the codes in order to investigate the effects of changing the prescriptions for the feedback and for the star formation. This effort is clearly beyond the scope of the present paper. Here we stress that our findings concerning the abundance of $pBzK$ galaxies push the comparison between models and observations to a higher level of accuracy, required to address the key questions above.

7. Summary and Discussion

In this paper we have exploited new deep, wide-field K -band imaging, performed with the High Acuity Wide field K -band Imager (HAWK-I) on VLT as part of the HAWK-I UDS and GOODS-S survey (HUGS; VLT Large Program), to study the population of passive evolving galaxies at $z \sim 2$. Crucially,

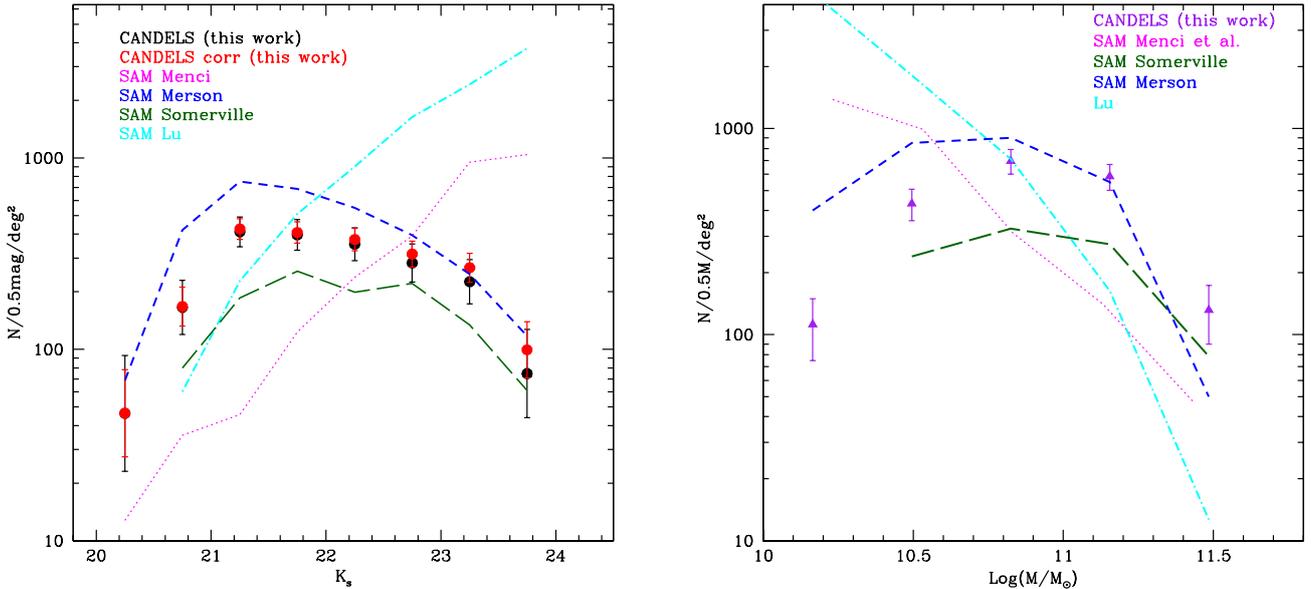


Fig. 7. Comparison of the observations with the Semi Analytical models of Merson et al. (2012) in blue, Menci et al. (2005) in magenta, Lu et al. in prep. in cyan, and Somerville et al. (2012) in green. In the left panel we compare the number counts and in the right panel the mass distribution.

both survey fields possess B and z -band imaging, as well as a wealth of other multi-wavelength data, extending from the U -band to the mid-infrared *Spitzer*, and containing deep HST imaging assembled as part of the CANDELS HST Treasury program.

To define a sample of star forming and quiescent galaxies at redshift between $1.4 < z < 2.5$, we use the BzK criterion proposed by Daddi et al. (2004). Adopting $K_s = 25$ as the appropriate limiting magnitude for our analysis, we find 99 $pBzK$ galaxies and 3071 $sBzK$ in the UDS CANDELS field, and 54 $pBzK$ and 1038 $sBzK$ in the GOODS-S field. At bright magnitudes the observed number counts of the $sBzK$ and $pBzK$ galaxies are in good agreement with previous studies, in particular with McCracken et al. (2010) as shown in Fig. 2. Thanks to the depth of our observations, we have now extended the selection at magnitudes fainter than was possible in previous analysis. We have demonstrated through simulations that the statistics of passive galaxies down to $K \approx 24$ is not significantly affected by incompleteness, and that the latter is still treatable down to $K \approx 25$. Thanks to the depth of our observations we are now able to place on a secure footing the earlier results on the $pBzK$ number counts previously reported Hartley et al. (2008) and McCracken et al. (2010).

Our central result is that the $pBzK$ number counts show a flattening at $K_s \sim 21$, and a turn-over at $K_s \geq 22$, equivalent to rest-frame absolute I -band magnitudes of $M_I = -23, -22$ respectively. Converted into stellar mass, our result corresponds to a decrease in the number density of passive-evolving galaxies at stellar masses below $10^{10.8} M_\odot$ for a Salpeter IMF. As judged against the still steeply-rising number counts of the overall galaxy population at these redshifts, this turnover is fairly abrupt, indicating that at high redshift the mechanism that quenches star-formation activity is much less efficient below this mass limit.

We have compared our observed number counts with the predictions of several semi-analytical models of galaxy formation and evolution, in particular with the models of Menci et al (2006), Somerville et al. (2012), Merson et al. (2013), and

Lu et al. (in prep). Among these SAMs only two, (Somerville et al. (2012) and Merson et al. (2013)) *qualitatively* predict the shape of the number counts, showing a turn-over at a stellar mass close to that observed in the data, but they do not reproduce the absolute observed density of the passive galaxies. In contrast, the other two models fail to show this turn-over, and predict an exponential increase of passive objects to faint magnitudes. This comparison suggests that the distribution of number density (with magnitude or stellar mass) of quiescent galaxies at these redshifts offers a critical test for hierarchical models, and can place strong constraints on the detailed baryonic physical processes involved in galaxy formation and evolution.

The observed discrepancies raise pressing, key questions concerning the comparison between observations and the current models of galaxy formation in a cosmological context.

First, are the color based observational criteria effective to select a the whole population of passive objects (or at least a representative sample of them) ? Indeed, the discrepancy with all present models suggests the possibility that present observations might be missing a whole class of objects in the color-color plane. As shown in Daddi et al. (2004), and in Grazian et al. (2007), the $pBzK$ selection criterion preferentially selects galaxies that have been passively evolving for about 1 Gyr: lowering the threshold in $(z-K)$ corresponds to introducing passive galaxies that are progressively younger, and hence become more numerous. Hence, we should expect that lowering this threshold would also increase the number of observed galaxies, keeping the discrepancy in place. Unfortunately this cannot be easily checked in the data, since lowering the threshold in $(z-K)$ immediately introduces also galaxies at lower redshift that contaminate the selection.

Second, are the current implementations of the star formation quenching process (tuned to match the observed local luminosity functions) effective in providing a fast enough evolution of galaxies in the color-color plane ? For in-

stance, the radio mode feedback (whose effectiveness increases monotonically with cosmic time) could be ineffective in providing the quenching of star formation at $z \gtrsim 2$ required to match the observed abundance of the $pBzK$ population.

Third, is the relative role of the different modes star formation correctly implemented by the models? In particular, it is known that both merging and disk instabilities can provide star bursts which add to the secular conversion of gas into stars (the "quiescent mode" of star formation). Current galaxy formation models are characterized by different implementations of three modes of star formation; while the bulk of the galaxy population could be less dramatically affected by the relative role of such process, the abundance of $pBzK$ galaxies characterized by extremely early star formation histories could be extremely sensitive to the relative importance of the different star formation modes.

Unfortunately, it is not possible to determine the specific features that, within each model, produce the observed trends on the quenching of star formation. This requires a more detailed investigation that we plan to explore in a forthcoming paper.

Acknowledgements. We thank the anonymous referee for a very useful report. The authors are grateful to the ESO staff for their support during the Hawk-I observations. JSD acknowledges the support of the Royal Society via a Wolfson Research Merit award, and also the support of the European Research Council via the award of an Advanced Grant. RJM acknowledges the support of the European Research Council via the award of a Consolidator grant. AF and JSD acknowledge the contribution of the EC FP7 SPACE project ASTRODEEP (Ref.No: 312725). This work is based in part on observations made with the NASA/ESA *Hubble Space Telescope*, which is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS5-26555. This work is also based in part on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under NASA contract 1407. Program based in part on the data collected in the ESO programs 60.A-9284, 181.A0717, LP 186.A-0898 and 085.A-0961.

References

- Andreon, S., Huertas-Company, M., 2011, *A&A*, 526A, 11
 Arcila-Osejo, L., & Sawicki, M., 2013, *MNRAS*, 435, 845
 Blanc, G. A., Lira, P., Barrientos, L. F., et al., 2008, *ApJ*, 681, 1099
 Bower, R., G., Benson, A., J., Malbon, R., et al., 2006, *MNRAS*, 370, 645
 Brammer, G.B., Whitaker, K. E., van Dokkum, et al., 2011, *ApJ*, 739, 24
 Bruzual, G., & Charlot, S., 2003, *MNRAS*, 344, 1000
 Calzetti, D., Armus, L., Bohlin, R., C., et al., 2000, *ApJ*, 533, 682
 Castellano, M., Fontana, A., Grazian, A., et al., 2012, *A&A*, 540, 39
 Daddi, E., Cimatti, A., Renzini, A., et al., 2004, *ApJ*, 617, 746
 Daddi, E., Renzini, A., Pirzkal, N., et al., 2005, *ApJ*, 626, 680
 Dahlen, T., Mobasher, B., and CANDELS Collaboration, 2012, *AAS*, 22013301D
 De Lucia, G., Springel, V., White, S. D. M., et al., 2006, *MNRAS*, 366, 499
 De Lucia, G., & Blaizot, J., 2007, *MNRAS*, 375, 2
 De Lucia, G., Poggianti, B. M., Aragn-Salamanca, A., et al., 2007, *MNRAS*, 374, 809
 Fioc, M. & Rocca-Volmerange, B. 1997, *A&A*, 326, 950
 Fontana, A., Donnarumma, I., Vanzella, E., et al., 2003, *ApJ*, 994, 9
 Fontana, A., Pozzetti, L., Donnarumma, I., et al., 2004, *A&A*, 424, 23
 Fontana, A., Salimbeni, S., Grazian, A., et al., 2006, *A&A*, 459, 745
 Fontana, A., Santini, P.; Grazian, A.; et al., 2009, *A&A*, 501, 15
 Furusawa, H., Kosugi, G., Akiyama, M., et al., 2008, *ApJ*, 176, 1
 Galametz, A., Grazian, A., Fontana, A., et al., 2013, *ApJ*, 206, 10
 Gehrels, N., 1986, *ApJ*, 303, 336
 Giavalisco, M., Ferguson, H. C., Koekemoer, A., M., et al., 2004, *ApJ*, 600, 93
 Grazian, A., Fontana, A., Moscardini, L., et al., 2006, 453, 507
 Grazian, A., Salimbeni, S., Pentericci, L., et al., 2007, *A&A*, 465, 393
 Grogin, N. A., Kocevski, D. D., Faber, S. M., et al., 2011, 2011, *ApJS*, 197, 35
 Guo, Y., Ferguson, H. C., Giavalisco, M., et al., 2013, *ApJS*, 207, 24
 Hartley, W. G., Lane, K. P., Almaini, O., et al., 2008, *MNRAS*, 391, 1301
 Ibert, O., McCracken, H., J., Le Fevre, O., et al., 2013, *A&A*, 556, 551
 Kitzbichler, M. G., & White, S. D. M., 2007, *MNRAS*, 376, 2
 Klypin, A. A., Trujillo-Gomez, S., & Primack, J., 2011, *ApJ*, 740, 102

- Kodama, T., Yamada, T., Akiyama, M., et al., 2004, *MNRAS*, 350, 1005
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al., 2011, *ApJS*, 197, 36
 Kong, X., Daddi, E., Arimoto, N., et al., 2006, *ApJ*, 638, 72
 Lane, K. P., Almaini, O., Foucaud, S., et al., 2007, *MNRAS*, 379, 25
 Lawrence, A., Warren, S., J., Almaini, O., et al., 2007, *MNRAS*, 379, 1599
 Lee, K.-S., Ferguson, H. C., Wiklind, T., Dahlen, T., et al., 2012, *ApJ*, 752, 66
 Lu, Y., Mo, H. J., Weinberg, M. D., & Katz, N., 2011, *MNRAS*, 416, 1949
 Lu, Y., Mo, H. J., Katz, N., & Weinberg, M. D., 2012, *MNRAS*, 421, 1779
 Menci, N., Fontana, A., Giallongo, E., et al., 2005, *ApJ*, 632, 49
 Menci, N., Fiore, F., Puccetti, S., et al., 2008, *ApJ*, 686, 219
 Merson, A. I., Baugh, C. M., Helly, J. C., et al., 2012, arXiv1206.4049
 McCracken, H. J., Capak, P., Salvato, M., et al., 2010, *ApJ*, 708, .202
 Muzzin, A., Marchesini, D., Stefanon, M., et al., 2013, arXiv1303.4409M
 Reddy, N. A., Erb, D. K., Steidel, C. C., et al., 2005, *ApJ*, 633, 748
 Somerville, R. S., Gilmore, R. C., Primack, J. R., et al., 2012, *MNRAS*, 423, 1992
 Stutz, A. M., Papovich, C., & Eisenstein, D. J., 2008, *ApJ*, 677, 828 bibitem[2008]Stu08 Trenti, M., & Stiavelli, M., 2008, *ApJ*, 676, 767

The Hawk-I UDS and GOODS Survey (HUGS): Survey Design and Deep K-band Number Counts

A. Fontana¹, J. Dunlop², D. Paris¹, T. Targett², and al.

¹ INAF - Osservatorio Astronomico di Roma e-mail: adriano.fontana@oa-roma.inaf.it

² University of Edinburgh

Received ; accepted

ABSTRACT

We present the results of an ultra-deep IR imaging survey executed with the Hawk-I imager at the VLT, that observed in the K and Y bands the UDS and GOODS-South fields covered by the CANDELS survey. We dub this public surveys HUGS, an acronym for *Hawk-I UDS and GOODS Survey*. This paper describes the survey strategy, the observational campaign and the data reductions process. We show that, thanks to exquisite image quality and extremely long exposure times, HUGS delivers the deepest K-band images ever collected over areas of cosmological interest, and in general ideally complement the CANDELS dataset in terms of image quality and depth. In the GOODS-S field, the K-band observations cover the whole CANDELS area with a complex geometry made of 6 different, partly overlapping pointings, in order to cover in a proportional fashion the deep and wide areas in CANDELS. In the deepest region, that includes most of the Hubble Ultra Deep Field, exposure times exceed 80 hours of integration, with a $1 - \sigma$ magnitude limit per square arcsec of about 28.0 *mags* (AB). The seeing is remarkably exceptional and constant across the various pointings, ranging between 0.38" and 0.43". In the UDS field the survey is about one magnitude shallower, to match the correspondingly smaller depth of the CANDELS images, but includes also the Y band, where WFC3 is lacking. In the K band, with an average exposure time of 13 hours, and a seeing ranging from 0.37" to 0.43", the $1 - \sigma$ magnitudes limit per square arcsec is about 27.3 *mags*. In the Y-band the average exposure time is of about 8 hours, and an average seeing of 0.45"-0.5", reaching $1 - \sigma$ magnitudes limit per square arcsec of about 28.3 *mags*. We show that the HUGS observations do match the depth of the CANDELS WFC3 data, since most of even the faintest galaxies detected in the H band images CANDELS are detected in HUGS. We finally present the number counts in the K band, as obtained after combining the two fields. We show that the slope of the number counts depends sensitively on the assumed distribution of galaxy sizes, with potential impact on the estimated EBL. All the HUGS images and catalogs are made public at the web site <http://www.oa/roma.inaf.it/HUGS>.

1. Introduction

Ultra-deep imaging surveys are one of the main tools to explore the early phase of galaxy formation and evolution. To some extent, each innovative step in the technology of telescopes and detectors has been immediately applied to obtain the deepest possible observations - the early galaxy number counts obtained with the first CCD (Ellis (1997) and references therein) being an obvious example. In the optical range, the set of the early Hubble Deep Field Campaigns (Williams et al. 1996, 2000) paved the ground to the first exploration of the high redshift Universe. Ground-based telescopes have also been used at similar purpose since the end of the last century, with nearly every new instrumental set-up available (e.g. the NTT Deep Field (Fontana et al. 2000), the Keck Deep Field (Sawicki & Thompson 2005), the VLT Fors Deep Field (Heidt et al. 2003), and others).

In recent years, more emphasis has been progressively devoted to deep imaging surveys in the near-IR. The most recent and spectacular case is the long series of Ultra Deep Field campaigns, obtained with the Wide Field Camera 3, the latest and more efficient instrument on board of HST. The observations secured in the various bands from the Y_{98} to the H_{160} represent our deepest view on the Universe, reaching a final depth that in some case exceeds the 31th magnitude.

The shift to near-IR surveys has not been driven by technological developments only. These surveys are motivated by our desire to compensate the negative effects of k-corrections at

high redshifts, in order to sample the rest-frame optical emission of galaxies up to the earliest phase of galaxy evolution.

In this context, K-band deep surveys have remained a fundamental tool even in the WFC3 era. As an example, the wavelength shift from the H_{160} band (the longest accessible from HST) to the K band enables us to extend the redshift coverage of the rest-frame B band from 2.6 to 4. At $z > 3.5$, imaging longward of H-band is needed to locate and measure the size of the Balmer break, which reaches IRAC 3.6 μ m only at $z > 8$. It is worth remembering that, at $z = 6$, the wavelength gap between H and 3.6 μ m is comparable to the gap between the observed Z and K bands at $z = 3$. Thus, straddling this large spectral range with deep K-band is crucial for an accurate determination of the rest-frame physical quantities (stellar age, stellar mass, dust content) of galaxies in this redshift range. For this reason, most of the newly introduced near-IR imagers have been used to secure progressively deep fields in the K-band, that has been used at the crucial purpose of obtaining mass-selected samples of galaxies at high redshift (Moustakas et al. 1997; Huang et al. 1997; Crist3b3bal-Hornillos et al. 2003; Labb3e et al. 2003; Minowa et al. 2005; Grazian et al. 2006; Cirasuolo et al. 2010).

Needless to say, an ultradeep field in a single bandpass is of relatively little scientific usage. For this reason, most of the surveys mentioned above have been focused on few, selected high latitude fields, in order to accumulate deep multiwavelength exposures across as many bandpasses as possible. Building on the experience of the early HDF, the concepts of color selection cri-

teria and photometric redshifts have become a common tool to explore galaxies at high redshifts.

The CANDELS survey is the latest, and most ambitious enterprise of this kind. It is a 900-orbit HST program (PI S. Faber, Co-PI H. Ferguson), representing the largest HST time award ever made. The new HST WFC3 imaging delivers 0.1-arcsec J (F125W) and H (F160W) imaging reaching 27.2 mag (AB; 5σ) over 0.25 sq. degrees, with even deeper (28 mag., AB; 5σ) 3-band (Y, J, H) imaging over ≈ 120 sq. arcmin (within GOODS-South and GOODS-North). It also delivers the necessary deep optical ACS parallels to complement the deep WFC3 imaging. The major scientific goals of this MCT program are the assembly of statistically useful samples of galaxies at $6 < z < 9$, measurement of the morphology and internal color structure of galaxies at $z = 2 - 3$, the detection and follow-up of SuperNovae at $z > 2$ for validating their use as cosmological distance indicators, and the study of the growth of black-holes in the centers of high-redshift galaxies. With ultra-deep radio imaging available in all 5 fields, deep X-ray imaging either available or planned, and Herschel/Laboca imaging (plus soon SCUBA2) also now provided at sub-mm wavelengths, the legacy value of the Spitzer+HST data is clear and unrivalled.

While optical ground-based or ACS imaging is available over most of the fields, the availability of adequately deep K -band data on the CANDELS fields is much scanty. Because of the small field of view, the depth of the ISAAC mosaic on GOODS-S is also significantly shallower than the new WFC3 observations from CANDELS, and of inhomogeneous quality. To fill this gap, we have designed a survey that makes use of the recently newly available instrument Hawk-I (the High Acuity Wide field K-band Imager, Kissler-Patig et al. (2008)) at VLT. The instrument delivers relatively large field-of-view square images with 7.5 arcmin of size, with exquisite sampling of the PFS (the pixel size is 0.11") and state-of-the-art quantum efficiency and cosmetics. Being optimally matched to the size of the CANDELS fields, it allows to reach an unprecedented combination of depth and area coverage. Our survey targets two of the three CANDELS fields accessible from Paranal, namely GOODS-S and the UDS, since UltraVISTA will already deliver ultra-deep Y, J, H, K imaging within the COSMOS field in the forthcoming years. We dubbed this survey HUGS (Hawk-I UDS and GOODS Survey) to emphasize the unique role of Hawk-I in this task.

Although HUGS is designed to complement the WFC3 CANDELS observations, it will also enable scientific investigations on its own, thanks to the depth of the K band images. Among these, we mention for instance the analysis of the spectral energy distribution of $z \approx 4$ galaxies, where the H band allows us to sample with high accuracy the Balmer break of selected LBGs (see Castellano et al. 2014, *subm*) or the evolution of the galaxy mass function at high redshift (see Grazian et al 2014, *in prep.*), where K -band selected samples are required to be as complete in mass as possible.

This paper describes and the accompanying data release provide a complete compilation of the data obtained within HUGS and also in all the previous observations that we executed with Hawk-I on GOODS. We note indeed that the GOODS-S field was already observed with Hawk-I in the framework of both the Hawk-I Science Verification as well as of a previous ESO Large Program aimed at identifying $z \approx 7$ galaxies. These program has delivered a robust estimate of the Luminosity Function of $z \approx 7$ galaxies (Castellano et al. (2010a,b)), and led to the discovery of the first robust spectroscopic confirmation at $z > 7$ (Vanzella et al. (2011)).

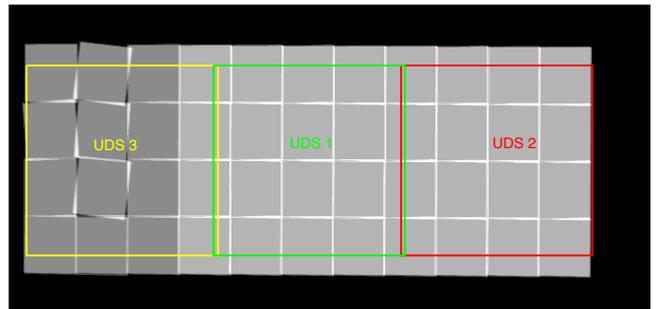


Fig. 1. The location of the three Hawk-I pointings overlaid on the WFC3-CANDELS mosaic of the UDS. Greyscale of the WFC3 images is on a linear stretch from 0 to 4ks.

This paper describes the survey design and the data collected out of the whole survey. We provide accurate estimates of the final quality of the data, in terms of depth and image quality - the latter is particularly impressive, given the long integrations from the ground that have been used. We finally use these data to obtain the deepest galaxy number counts ever secured in the K band over a statistically meaningful area. We have used AB magnitudes throughout the whole paper.

2. Survey strategy

The HUGS survey has been designed to cover the two CANDELS fields that do not have in the present or in the near term future suitably deep K band images: UDS and GOODS-South (hereafter GOODS-S).

The depth of the images in the K band has been tuned in order to match the depth of the WFC3 images produced in the J_{125} and H_{160} filters. In practice, the target depth was 0.5 mag shallower than H_{160} ones, suitable to match the average $H - K$ color of faint galaxies.

In both fields deep Y band images have also been acquired. In the case of UDS, these images complement the CANDELS data set, since neither the Y_{098} nor the Y_{105} has been obtained within CANDELS. In the case of GOODS-S, the images come from an earlier program designed to select $z \approx 7$ galaxies with ground-based images Castellano et al. (2010a,b). These images are slightly less deep than the Y_{105} images that have been later obtained within CANDELS, and cover about 70% of the GOODS-S field, but are nevertheless reduced and made available here. We describe below the details of the two fields, in terms of pointings, exposure time and expected depth.

We note that, since Hawk-I is a mosaic of four square detectors ($2k \times 2k$ each), it delivers images that exhibit a shallower cross at the center of the mosaic. Although our dithering pattern has been chosen to minimize its impact, this feature is inevitable in the output data.

2.1. The UDS field

Thanks to its quite regular shape, the UDS has been straightforward to cover with Hawk-I. Three different Hawk-I pointings are able to cover 85% (**CHECK**) of the UDS field. The layout is shown in Figure 1. We show the position of the three different pointings (named UDS1, UDS2 and UDS3 in the following), assuming a nominal size of $7.5' \times 7.5'$ for the Hawk-I image. It is shown that the three pointings also provide two overlapping regions that have been used to cross-check the photometric and astrometric solutions in the three individual mosaics. The

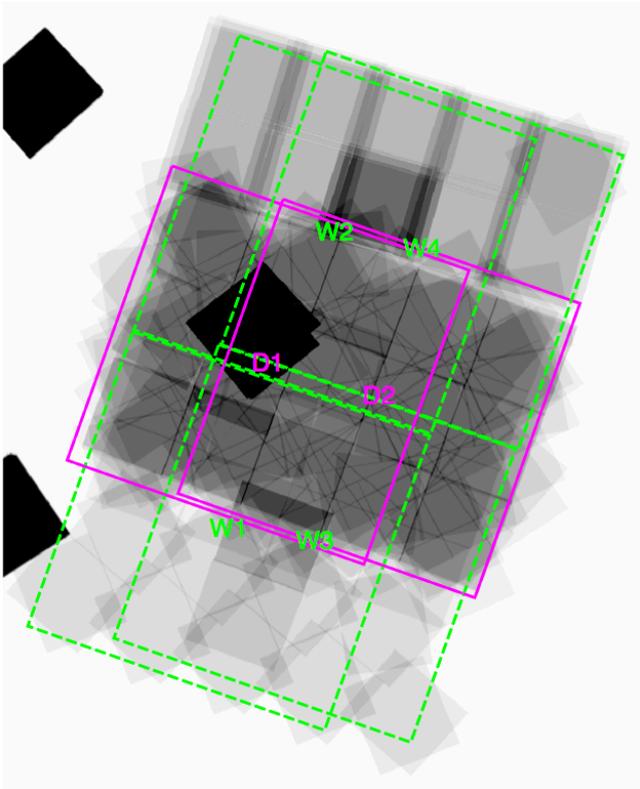


Fig. 2. *Upper:* the location of the Hawk-I pointings overlaid on the WFC3 data available on the GOODS-South field. The black square at the centre is the HUDF12 region. Greyscale of the WFC3 images is on a linear stretch that saturates at the deepest levels of the CANDELS data - the HUDF12 is deeply oversaturated.

three pointings have been exposed with nearly identical exposure times, of 8 hours in the Y band and 13 hours in the Ks band (final exposure times are slightly different since some image has been discarded during the reduction process). Table 1 summarizes the location and exposure time of the various pointings.

2.2. The GOODS-S field

The coverage of the GOODS-S field with CANDELS is more complex, and forced us to a more complicated pattern for the HUGS observations. The WFC3 observations are deeper in a rectangle region (10 by 7 arcmin by side, named GOODS-Deep in CANDELS) that spans the entire width (West-East extension) of the GOODS-S field, and is centered in the vertical (i.e. North-South) direction. Close to the center of this area it is placed the Ultra Deep Field region (UDF in the following). These deep images are complemented by shallower images, partly coming from the ERS survey and partly by CANDELS, that cover the remaining part of the original ACS frames. Our final layout has been designed to deliver a deeper image over GOODS-Deep, while covering nearly the whole CANDELS area. Since the width of the GOODS-S field is 10 arcmin, it cannot be covered efficiently with a single Hawk-I pointing. We therefore decided to cover the whole field with a 2×3 grid of pointings, rotated by -19.5 degrees as the ACS and WFC3 mosaics. The layout is shown in Figure 2: as for the UDS, we show both the position of the 6 individual pointings overlaid upon the WFC3 exposure map (that includes also the position of the UDF and the other parallel deep fields) as well as the final expected exposure map.

The pointings are offset in the W/E direction by 3 arcmin each, and in the NS by 6 arcmin each. This approach has also made optimal use of the Ks-band images obtained in our first program. The two central pointings have been exposed with significantly longer exposure times, for a total of about 31 hours, while the four upper and lower pointings have been exposed for about 11 hours each. We therefore name GD1 and GD2 the two deep exposures, and GW1, GW2, GW3, GW4 the four shallower ones¹. This layout has also the advantage of producing a final mosaic where each region of the GOODS-Deep area is observed with different physical regions of the instrument, further minimizing possible trends due to large-scale residual in the flat-fielding.

Looking at Figure 2 it is immediately appreciated that the coverage of the GOODS-Deep area is not uniform, because of the combined effects of the pointing locations and of the Hawk-I gaps. It reaches nearly 90 hours of exposure time in the very central area, that covers most of the UDF, and in any case more than 40 hours of exposure over the remaining part of the GOODS-Deep area. It currently represents a unique combination of depth and area in the K band.

During the earliest observations, several frames were incidentally acquired in the H and Br- γ filter, over the GD1 pointing. We have reduced also these images, and make them available, although they have not been used in any scientific analysis so far.

Table 2 summarizes the location and exposure time of the various pointings.

3. Data acquisition and reduction

3.1. Observations

All images in the K band has been obtained with individual images of 10 seconds of integration, averaged in sets of 12 images during acquisition (in the ESO slang these two parameters are referred to as DIT and NDIT, respectively). In the case of the Y band images we have adopted DIT=30 and NDIT=4. A random dithering pattern with a typical offset of 12 arcsec has been applied in all cases. Observations were scheduled in Observing Blocks of about 1hr of execution each, corresponding to about 45 and 48 minutes of exposure in K and Y, respectively. The parallactic angle of each OB was rotated by 90 degrees, so that the final mosaics are the results of individual images obtained with different physical regions of the detectors, which allowed us to test the accuracy of the photometric and calibration procedures.

All observations have been executed in Service Mode, over the period comprised between P86 and P90. We have retrieved from the archive all the images obtained during these runs, including those that were not graded within specifications during the observations. All images have been analyzed with an automated pipeline to assess their quality. We have included in the final coadded frames also some of the images graded “out of specs”, except those with wildly discrepant seeing, poor photometric quality or other cosmetic defects. For this reason the actual exposure times listed in Tables 1 and 2 are slightly different from pointing to pointing, although we planned identical exposure times.

¹ For those readers who’d dare downloading and reducing the raw data from scratch, these are named GOODS1 or GOODS-D1, GOODS-D2, GOODS-WIDE1, GOODS-WIDE2, GOODS-WIDE3 and GOODS-WIDE4 in the OB description.

Table 1. Layout and summary of observations for the UDS field.

Pointing	Central RA	Central DEC	Area (arcmin ²)	Exp. time (Sec/Hrs)	Final seeing	maglim ^a	maglim ^b
K band							
UDS1	02:17:37.470	-05:12:03.810	70	48360 / 13.43	0.37	27.4	26.1
UDS2	02:17:07.943	-05:12:03.810	70	46820 / 13.00	0.43	27.3	25.9
UDS3	02:18:06.896	-05:12:03.810	70	45240 / 12.57	0.41	27.4	25.9
Y band							
UDS1	02:17:37.470	-05:12:03.810	70	28800 / 8.00	0.45	28.4	26.9
UDS2	02:17:07.943	-05:12:03.810	70	28800 / 8.00	0.50	28.3	26.7
UDS3	02:18:06.896	-05:12:03.810	70	29400 / 8.17	0.48	28.2	26.6
H band							
UDS1	02:17:37.470	-05:12:03.810	70	13800 / 13.83	0.44	N.A.	N.A.
Br- γ band							
UDS1	02:17:37.470	-05:12:03.810	70	5760 / 1.6	0.41	N.A.	N.A.
UDS2	02:17:07.943	-05:12:03.810	70	5760 / 1.6	0.42	N.A.	N.A.

Notes. ^(a) at 1σ arcsec⁻² ^(b) at 5σ in 1 FWHM

Table 2. Layout and summary of observations for the GOODS-S field. We note that each pointing has been rotated with PA=-19.5degrees

Pointing	Central RA	Central DEC	Area (arcmin ²)	Exposure time (Sec)	Final seeing	maglim ⁽¹⁾	maglim ⁽²⁾
K band							
GOODS-D1	03:32:36.835	-27:47:45.24	70	113520 / 31.53	0.39	27.8	26.5
GOODS-D2	03:32:24.890	-27:48:33.22	70	112800 / 31.33	0.38	27.8	26.5
GOODS-W1	03:32:41.080	-27:51:44.32	70	47220 / 13.12	0.43	27.4	26.0
GOODS-W2	03:32:29.650	-27:44:37.26	70	40800 / 11.33	0.38	27.3	26.0
GOODS-W3	03:32:31.796	-27:51:01.74	70	37320 / 10.37	0.38	27.3	25.9
GOODS-W4	03:32:20.242	-27:44:59.97	70	41880 / 11.63	0.42	27.3	25.8
H band							
GOODS-D1	03:32:36.835	-27:47:45.24	70	21360 / 5.93	0.42	N.A.	N.A.

3.2. Data reduction

We have initially used two pipelines to independently reduce the images acquired in the first year of observations. One pipeline has been developed at the Rome Observatory, and is derived from a pipeline used to reduce LBT imaging data both in the visible and in the IR. In its former version it has been used to reduce the earliest Hawk-I data in GOODS-S (Castellano et al. (2010a,b)). The second pipeline has been developed at the Edinburgh Observatory and used to analyze other Hawk-I data set Targett et al. (2011). We have then compared two pipelines and their produced images have been compared, both in terms of conceptual steps and algorithms adopted, as well as in terms of the final mosaic produced. This comparison has yielded a final version of the Rome pipeline, that has been used to eventually process all the data, including a re-processing of those already reduced. In this sense, the images produced here are slightly different - and better - than those used in Castellano et al. (2010b). This pipeline is described in more details in a separate paper (Paris et al in prep.) but we describe here the basic steps (that follow the usual recipe of IR data reduction) and specific features.

3.2.1. Pre-reduction

The raw images have been retrieved from the ESO archive and each Observing Block has been processed separately at this stage. The procedure of the pre-reduction consists on the removal of the dark current and applying a flat-field in order to normalize the response of each image pixel. At first each scientific frame is subtracted by a median stack dark image obtained by combining a set of dark frames, with the same EXPTIME and NDIT values of the observation set images. Then a median

stack flat image (masterflat) is created, by combining a set of sky flats taken with the same filter of the observation set, each subtracted by its own dark. While combining, each flat is normalized by its own median background level, in order to obtain a final median stack flat normalized to unity. Each scientific image is then divided by the masterflat, so the response of pixels is finally homogenized. During the prereduction also pixel masks are created to flag saturated regions, cosmic rays events, satellite tracks and bad (hot/cold) pixel. We have also developed a specific procedure to take into account the effects of persistency - i.e. the residuals left by bright objects observed within an exposure, that leave residual counts on the subsequent image. After some test, we decided to identify all the pixels that are above 10^4 counts in a given image and mask them out from the following image. Given the large number of subsequent images, this efficiently masks out most of the pixels affected by persistency. We assume that the remaining contribution is efficiently wiped out by the dithering process, such that it does not yield detectable sources.

3.2.2. Background subtraction

After prereduction images are far to be flat. Structures appear both at small and at large scales due to a variety of causes, such as pupil ghosts, dust and sky background variation during the observation, that are particularly remarkable in IR bands, especially in K. We have developed specialised algorithms to carefully remove these structures. Since they are assumed as additive features, the basic operation is to create and then subtract maps of background from each image. The first map is a sigma-clipped median stack image of the observations included in temporal window - typically of about 10 minutes - around the pro-

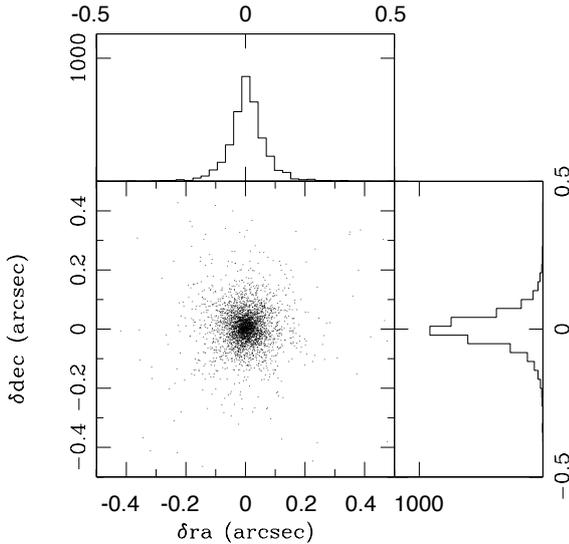


Fig. 3. Differences in RA and DEC for all the objects detected in the overlapping region between the GD1 and GD2 pointings. The two insets on the right and top show the resulting distributions

cessed image. Y-band images show already a good result after the subtraction of this first map, while a two-pass procedure has been developed to subtract such residual features that survive in K-band. At the end of this stage, in which each Observing Block has been processed separately, images are flat and ready to be processed to create the final mosaic.

3.2.3. Astrometric solution

In order to perform the coaddition an accurate astrometric calibration has to be performed. In fact images show geometrical distortions arising from the positional errors of each pixel due to many causes, such as optical distortions, atmospheric refraction, rotation of chips, non-integer dithering pattern, etc. The procedure of astrometric calibration consists on two basic operation: the correction of relative linear offsets between exposures and the refined absolute global calibration. For each exposure a SExtractor catalog is created, and the relative linear calibration between exposures is done by correcting for the offsets between sources coordinates, computed by the cross-matching with a catalog chosen as reference. The absolute calibration is done by giving an absolute reference catalog and correcting for distortions through the sources coordinates cross-matching and by storing the final corrected solution into the header of the images. For both data sets we use as reference the CANDELS WFC3 images. For GOODS we use catalogs from Giavalisco et al. (2004), Koekemoer et al. (2011) and Grogin et al. (2011), while for UDS we refers to Koekemoer et al. (2011) and Grogin et al. (2011).

The ultimate accuracy of the astrometric solution has been tested using the regions with overlapping images. As an example, we show in Figure 3 the distribution of the differences in RA and DEC for the objects that fall in the large overlapping area between the GD1 and GD2 pointings.

3.2.4. Estimate of absolute noise image

Absolute noise map for each exposure are created directly from the raw images. They are based on the assumption that the noise is given by the Poisson statistics of the counts detected in each pixel of the original frames. This contribution is propagated to take into account the scaling applied to each pixel during processing (including flat-fielding, normalization of exposure times rescaling of zero-points etc). The resulting absolute noise map at pixel X, Y , $\sigma(X, Y)_i$ can be obtained by the following formula:

$$\sigma(X, Y)_i = \sqrt{\frac{RAW(X, Y)_i}{gain_i}} \times \frac{1.0}{\sqrt{FLAT(X, Y)_i}} \times \frac{1.0}{\sqrt{ndit}} \quad (1)$$

where $i = 1, 2, 3, 4$ is an index that represents the number of the chip of HAWK-I, $RAW(X, Y)_i$ the raw number counts at pixel X, Y , $gain_i$ is the read-out-gain, $FLAT(X, Y)_i$ is the value of the flat-field image used to calibrate the raw image, $ndit$ the value of $NDIT$

3.2.5. Coaddition

Our pipeline uses SWarp (Bertin et al. (2002)) to resample all the processed images, implementing it into a procedure designed to properly propagate the absolute r.m.s. obtained as above. At first a global header is created from the input images which are resampled according to the geometry described into the resulting global header. In order to obtain a physical exposure and a rms map of the final mosaic, during the resampling the internal WEIGHT_TYPE parameter has to be set to the MAP_RMS modality, so that a noise map has to be given for each exposure. During the resampling stage images with bad astrometry informations are rejected, while for each resampled image SWarp provides a weight map in output. The last step is to perform a weighted summation of all the resampled images, using as weight $w_i(X, Y) = 1/\sigma_i(X, Y)^2$, where σ for each pixel is given by Eq. 1. The final rms map is obtained simply by: $RMS(X, Y) = 1/\sqrt{\sum_{i=1}^n w(X, Y)^i}$. These RMS images are released along with the science data (see below). Since the current version of SWarp is not able to produce these r.m.s. images, these steps have been obtained with a specific pipeline.

3.2.6. Photometric calibration

At the end of each reduction we have adopted a careful procedure to calibrate the photometry and estimate the zero point, independently for each pointing. For each pointing we have chosen at least one OB qualified as photometric during the observations, and we have stacked them in order to obtain a mosaic of the field with about 1 hour of exposure, in good photometric conditions and with consistent airmass. We then retrieved from the ESO archive a set of standard stars observed at the same airmass of the scientific images, and temporarily as close as possible to the observations. We have reduced the standards using the same calibration frames used for the scientific images. At the end of the reduction we have extracted a catalog for each standard calibration image and by comparing the magnitude of the standard star, corrected for extinction, with the magnitude reported in literature, we have got a first estimate of the zero point zp_1 , that we assume can be straightforwardly applied to the stacked OB described above. We have finally extracted and cross-matched two catalogs, the first from the full complete mosaic, and the

second from the 1-hour mosaic, calibrated with zp_1 . By comparing the differences in magnitude between the sources in the two catalogs we have got a final refined estimate of the zero-point. To minimize possible systematic offsets all these operations have been executed using the MAG.BEST magnitude obtained by SExtractor on the brightest objects only. We estimate that the typical uncertainty in the derivation of the zeropoints is of ± 0.02 mags, as obtained from repeated estimate of the ZP on independent sets of OBs and standards of the same pointing.

4. Validation and tests on Photometry

We have performed a number of tests, both on intermediate steps of data reduction as well as on the final images, comparing them with external data sets.

We report here two classes of comparisons, that may be of general interest for the reader. In all these cases we have obtained single-band photometric catalogs using SExtractor and cross-matched the catalogs using the measured RA and DEC. We then use the difference in observed total magnitudes for the objects in common between the various catalogs. We note that the stability of the photometric solution is in general quite good, to the extent that its validation has been ultimately limited by the uncertainties in the photometry. As our fields are relatively devoid of stars, most of the objects that we have used for comparisons are galaxies. It is well known that, when galaxies are observed with different seeing, sampling and depth, the estimate of their total magnitude is affected by systematics that depend on the size, profile and surface brightness of individual objects. To minimize these effects we have performed these tests on bright objects (typically those detected at $S/N > 35$) and used Kron magnitudes, as measured by SExtractor, that are relatively less sensitive to these effects.

First, we compared the catalogs obtained from fully reduced stacks of the different pointings in the overlapping regions. They have revealed small differences (usually within the errors) of the zeropoints that have been averaged out in order to provide internally consistent data set.

All the pointings of our survey have overlapping region with (at least one) nearby pointing. The typical size of these overlapping areas is XX and YY in UDS, and larger in GOODS-S (see Figure). We have used the observed systematic offsets between the magnitudes in the various pointings (as shown in Figure 4) as a measure of the systematic differences in the derived zeropoints. As shown in Figure 4, these offsets are always small (of the order of 0.01 - 0.02 mags), consistent with the uncertainty of the flux calibration. For the UDS, we have tied the photometry to the zero-point of UDS2, i.e. we have renormalized the original zeropoints of UDS1 and UDS3 in order to make the photometry of the overlapping areas fully consistent. In GOODS-S we have similarly tied the photometry to the GD1 one.

We have also compared our final images to the previously available images obtained by wider field imagers. The goal of this exercise is to further check against large-scale trends that may be left in our data, that cannot be identified on the overlapping areas.

For the UDS we have used the DR8 release of the UKIDSS Deep Survey, that has images a 0.77deg^2 field that includes the CANDELS field. For the GOODS-S we use the output of the Extended Chandra Deep Field (ECDS), that was obtained with the SOFI instrument on NTT. Results are shown in Figure 5 and Figure 6. Again, they show no major systematic trend in the data.

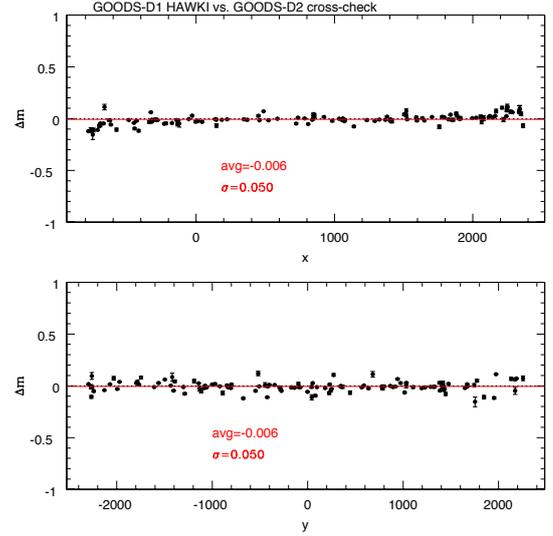


Fig. 4. Difference in magnitude for objects detected in both the GD1 and GD2 fields, as a function of the X, Y position in pixels in the GD2 field. Since the field is rotated by -19.5 degrees, using RA and DEC could hide real trends in the reduced data. This plot is before any renormalization of the zero-point (see text for details).

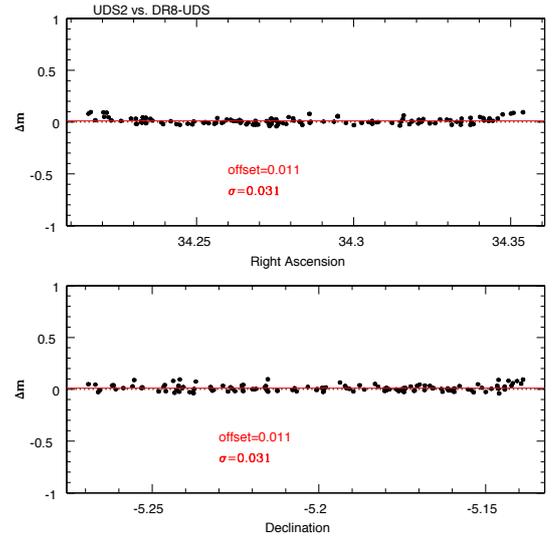


Fig. 5. Difference in magnitude for objects detected in the UDS2 pointing and in the UKIDSS DR8 release, as a function of RA and DEC (see text for details).

5. Data release

5.1. Images

We have finally obtained a set of mosaics of each data set, that we make publicly available. For each data set we release:

- The coadded image for each pointing (UDS1,2,3, GOODS-D1, D2, W1, W2, W3 and W4); These have all calibrated and rescaled to a standard zero-point of 27.5 for the K images, and 27.0 for the Y;
- The relevant absolute r.m.s. images, with the same flux scale;
- A global mosaic of the two fields in each bands, with the relevant absolute r.m.s., after homogenizing all images to the same PSF. This has been done by degrading the highest qual-

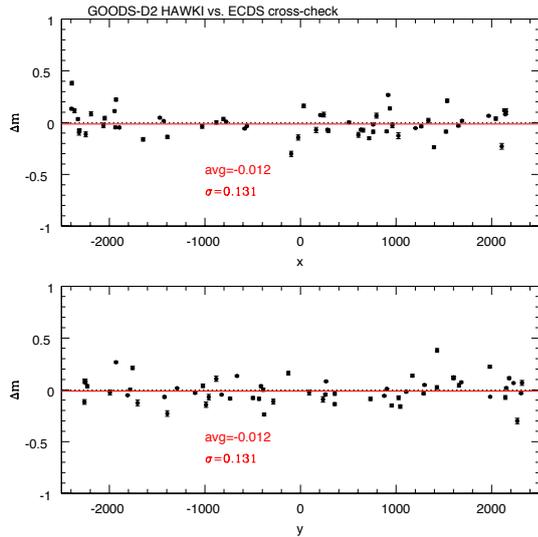


Fig. 6. As in Figure5, for the GD2 and ECDS field (see text for details).

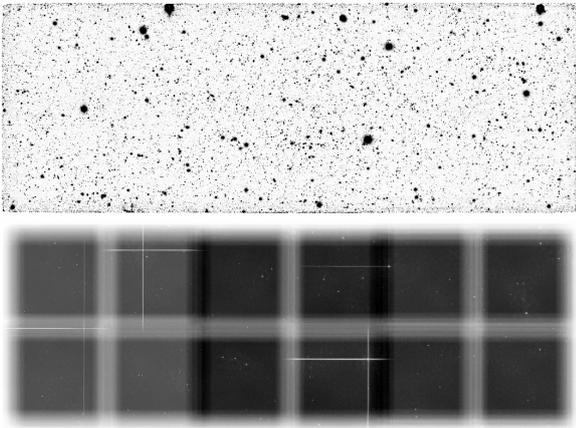


Fig. 7. *Top:* The final image on the UDS field, in the Y band. *Bottom:* The weight image, computed as described in the text. Darker regions have higher weight - and hence correspond to deeper regions of the images. The leftmost pointing (UDS3) is slightly shallower than the other two pointings, despite they have the same exposure time, because the average background observed during the observations turned out to be larger than for the other two pointings. THIS FIGURE MAY GO FULL WIDTH

ity images to the one with higher seeing. Although the seeing are relatively similar, and all excellent, this procedure inevitably degrades part of the information contained in the data. The correlation of the background pixels is also visibly different across each image, because of the different degree of filtering applied to each original pointing.

- A global mosaic of the two fields in each bands, with the relevant absolute r.m.s., without any correction of the different PSF. These mosaic have varying PSF across the fields (in a smooth way across the overlapping region) but do not show a varying degree of correlation in the background. These are probably most useful for illustrative purposes, and have been used in the following illustrations.

We remark that the most accurate scientific analysis is, for most application, obtained by separately using the individual pointings and then combining the output in an appropriate way,

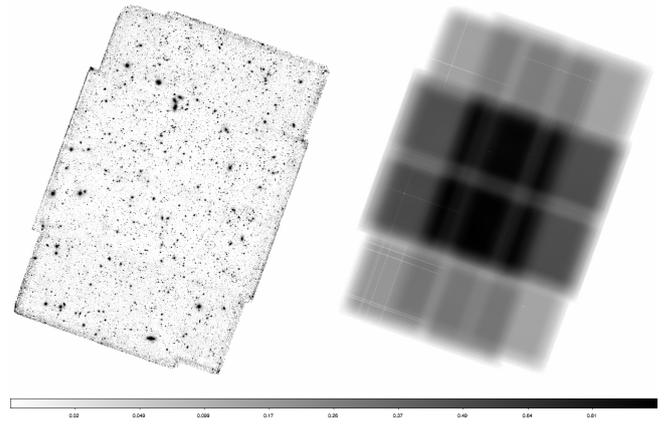


Fig. 8. *Left:* The final image on the GOODS-South field, in the K band. *Right:* The weight image, computed as described in the text. Darker regions have higher weight - and hence correspond to deeper regions of the images. THIS FIGURE MAY GO FULL WIDTH

as we did for the derivation of the multicolor catalogs (Galametz et al 2013, and see below).

The final images for the two fields are shown in Figure 7 and Figure8 The images shown are those obtained by combining the various pointings into a single mosaic, without performing any PSF matching prior to coaddition.

Figure7 and Figure8 show also their r.m.s. images of the final mosaics. The relics of cosmetics defects are clearly visible in the r.m.s. images, that are less exposed in the regions where part of the data have been removed to eliminate defects or trails.

Because of the inhomogeneous depth of the final mosaics, the limiting magnitude is not constant across the fields, in particular for the GOODS-S one. This is shown in Figure10, where we plot the distribution of the magnitude limit in the two fields. This has been computed converting the calibrated r.m.s. contained in each pixel into a 1σ magnitude limit in 1 arcsec^2 . The two peaks in the UDS field come from the slightly shallow exposure obtained in the UDS3 pointing, where the sky background effectively observed in the data was higher than the average in the other two. In GOODS-S, the 4 broad peaks in the distribution of the magnitude limits come from the complex geometry of the exposure, as shown in Figure8.

5.2. Catalogs

Catalogues from the HUGS images can be extracted in two different ways, either as “single band”, i.e. using the HUGS images as detection image, or adding them to the full multiwavelength suite of data in CANDELS.

5.2.1. Single band catalogs

Single-band catalogs are in principle straightforward to obtain. We have used the SExtractor code to obtain single-band detected catalogs that are distributed along with the images. However, a compromise must be achieved in this case between two conflicting requirements. First, to make full use of the largest possible depth, the detection should be made on the global mosaic, especially in GOODS where the overlaps between the various pointings are significant. However, since the detection process must be tuned to the PSF of the images, and given the different seeing of the HUGS images, to obtain a fully consistent catalog we have used as input images the seeing-homogenised images

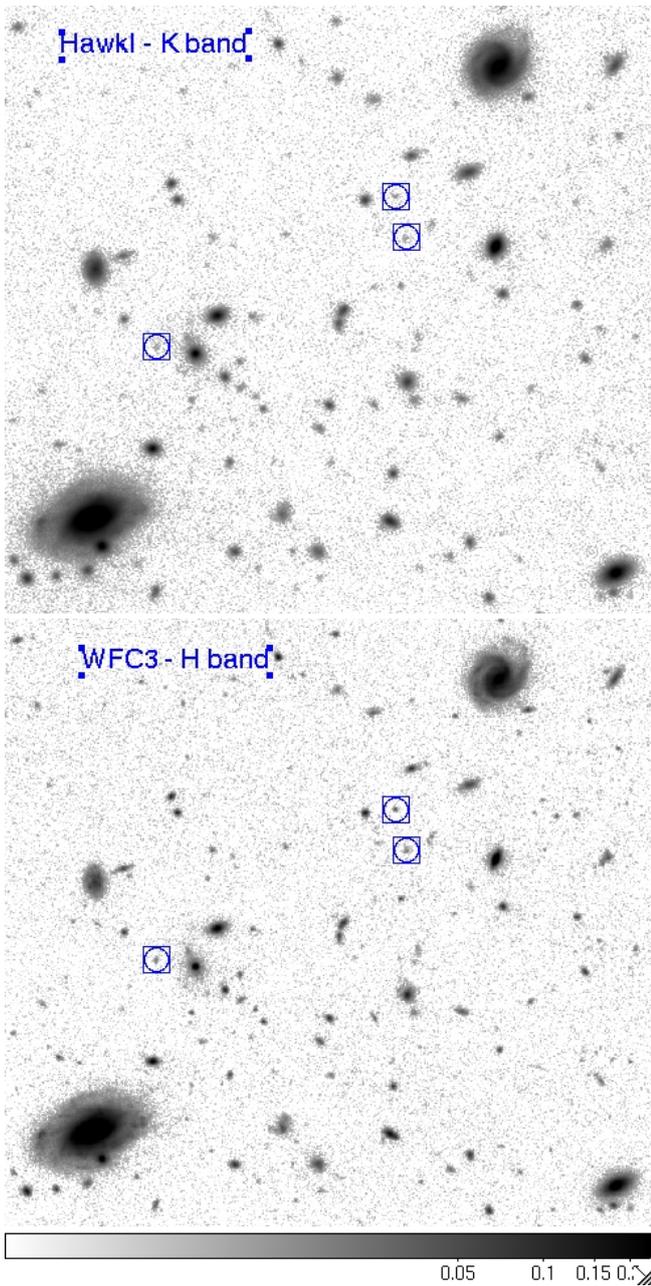


Fig. 9. The center of the GOODS-S field as observed with hawk-I in the K band (upper) and with WFC3 in H band (bottom). The displayed area is 1 arcmin wide, and is extracted from the region where the Hawk-I data have the maximum depth. The H band image is from the CANDELS Deep area. Both images have a dynamical range extending from 0.5σ to 100σ per pixel, on a logarithmic scale. Objects encircled have an H magnitude ≈ 26 and a color $H - K \approx 0.5$, typical of faint galaxies.

presented above. This somewhat limits the possibility of detecting the faintest galaxies in the pointings with the best seeing, although the seeing difference among the various images is not dramatically large (see Tables 1 and 2). We deemed this procedure as the most appropriate to obtain single-band catalogs, that are made publicly available for future uses. For the same reason, the number counts presented in the next sections have been obtained only on a sub-set of the images. Summarizing, the public catalogs that we derive have been obtained using SExtractor on the seeing-averaged mosaics of both UDS and GOODS-S. We

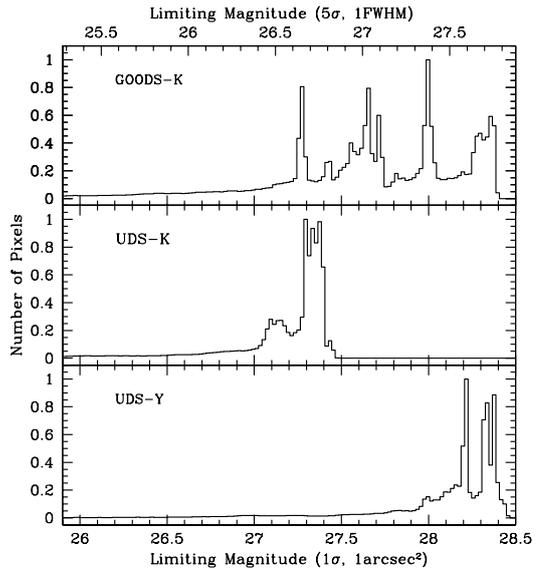


Fig. 10. Distributions of limiting magnitude in the 3 HUGS images, as described in the legend. The lower x-axis shows the limiting magnitude computed at 1σ in an area of 1 arcsec^2 , the upper at 5σ in an area of 0.4 arcsec^2 , comparable to the average FWHM. Note that the latter is not the total magnitude of objects detected at 5σ in an area of 0.4 arcsec^2 , since no aperture correction has been applied.

have applied a smoothing before detection (σ), and used as minimal detection area of (σ), requiring $S/N > 5$ in such area. The background has been estimated .. (PUT HERE DETAILS OF SExtractor runs.)

5.2.2. Multiwavelength catalogs

We have included the HUGS images (both in K and in Y) to the GOODS and UDS multiwavelength catalogs described in Galametz et al. (2013) et al and Guo et al. (2013) respectively. In both cases we have detected the objects in the WFC3 H-band image from CANDELS, and performed a PSF-matched photometry on the HUGS images. This has been accomplished by using the TFIT package to properly take into account the morphology of each object during the deblending process. We defer the reader to those two papers for more details. The Guo et al. catalog was compiled using only the first epoch of Hawk-I images in GOODS-S. We have therefore obtained again the K-band photometry using the final images described here, for all the objects detected in the H-band. This catalog is released here.

We note that, to deal with the different PSF of the various images, we have independently processed each of the final individual pointing described above, and thereafter weight-averaged the photometry of objects detected on multiple images to obtain the final photometry. Clearly, following this procedure the ultimate depth of the catalog is driven by the WFC3 H-band image. It may be of some interest to show how effective are the HUGS images in providing us with useful information on the CANDELS-detected objects, that is one of the main aim of the HUGS survey. This is shown in Figure 11, where we plot the fraction of the objects that are detected at 5σ or 1σ in the K band as a function of the H magnitude. In the case of GOODS, it is shown that as much as 90% of the H-detected galaxies have some flux measured at $S/N > 1$ in the K band, down the faintest

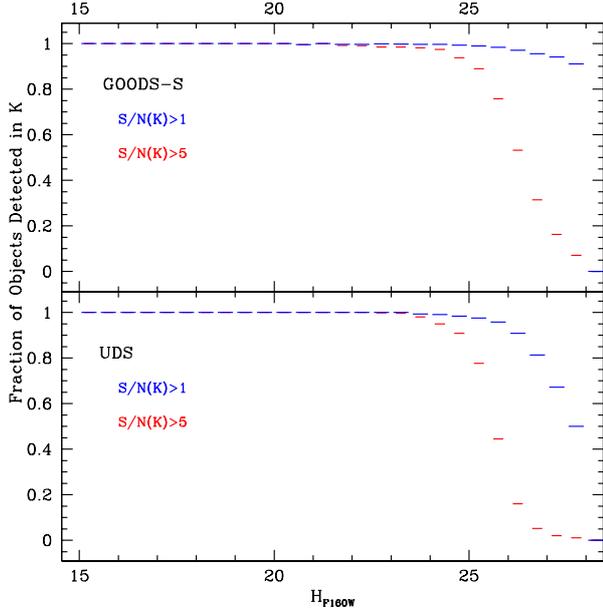


Fig. 11. Fractions of objects in the CANDELS catalogues that have a detected flux in the HUGS data, as a function of the H band. Here the H band is measured on the CANDELS F160W images, and the corresponding K -band flux is measured with the TFIT code on the final HUGS images. Results are shown for two different S/N ratio in the K band, and for the two HUGS fields, as shown in the legends.

limits of the H-band catalog, and that nearly 50% of the $H \approx 26$ galaxies (and 15% of the $H \approx 27$ ones) have a K -band solid detection at $S/N > 5$. We note that these statistics are measured on the full GOODS-S HUGS area, that is highly inhomogeneous in depth both in the H and in the K bands. This result confirms that our original goal has been achieved, and in particular that our pointing strategy has been quite efficient to cover the inhomogeneous GOODS field at the requested depth.

6. Number Counts

We have finally obtained the number counts in the K -band, combining both UDS and GOODS-S images. At variance with the procedure described above, we decided not to use the seeing-homogenized mosaics since their use would limit the final depth of the catalogs, due to the worse seeing. In practice, we have obtained independent catalogs from the three UDS fields (which have notably different seeing), optimizing each catalog to the relevant PSF. In the case of GOODS, we built a specific mosaic using the two deep pointings D1 and D2 as well as W2 and W3. These four pointings have remarkably close seeing, such that we have built a mosaic without any seeing-homogenization, that allows us to exploit the deepest images obtained in HUGS without degrading their quality.

Four catalogs (UDS1, UDS2, UDS3, GOODS-D1+D2+W2+W3) have been obtained by using SExtractor, as described above. Number counts have been derived independently on each image, using the procedure described below, and a weighted average among them has been finally computed to obtain the final number counts. The small overlap between the UDS fields makes the three catalogs not entirely independent, but we have ignored this effect since objects detected in more than one image are about 1% of the total. After trimming the outer regions of the images, the final area over which we compute

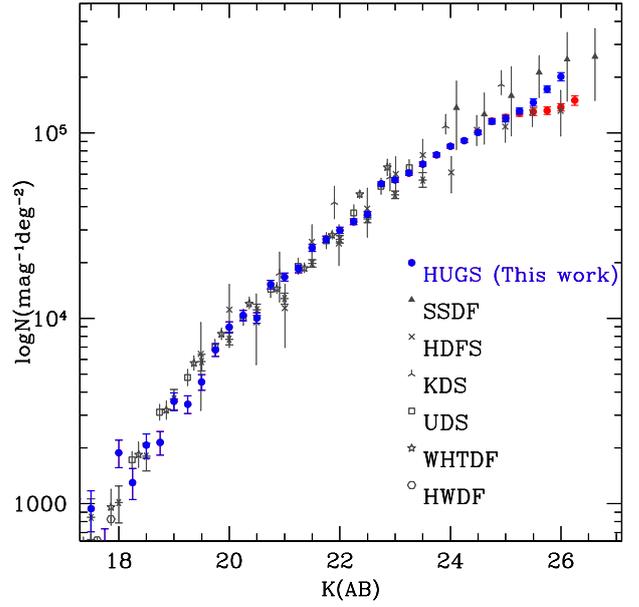


Fig. 12. Number counts in the K band, from HUGS and from recent literature. In blue are HUGS counts corrected for incompleteness assuming a size $r_{hl} = 0.1'' - 0.3''$ for faint galaxies. In red the same counts are corrected assuming point-like sources.

the number counts is 340.58 square arcmin, i.e. about 1/10 of a square degree. We note however that the area over which number counts depend effectively on the magnitude, because of the inhomogeneity of our exposure maps, such that the deepest number counts (those at $K > 25$) are computed over an area of 50.17 square arcmin in GOODS.

The key ingredient to compute number counts is the estimate of the incompleteness and other systematics involved, that is achieved through the use of simulations. As customary in these cases, we have performed these simulations by inserting fake objects in the images with a range of magnitudes (from $K=18$ to $K=27.5$) and sizes (we used exponential profiles with half-light radius randomly chosen in the range $0''-0.4''$). Objects have been convolved with the observed PSF and placed at random positions in the field. Thereafter, detection has been done in the very same way as on the original images and the output magnitudes measured for the fake objects (when detected) have been retained, along with the detected fraction. Only 200 objects have been placed in each run, in order to minimize crowding effects. Simulations have been repeated until 10^6 objects have been simulated in each field.

To use these simulations, we have followed exactly the same approach used in the analysis of the deepest K -band image obtained so far, in the AO-assisted Subaru Super Deep Field (SSDF) Minowa et al. (2005) (see their equations 1) and 2)). This method takes into account three sources of errors and incompleteness. The first is the incompleteness, i.e. the fraction of objects lost as a function of their real input magnitude. It also takes into account the systematic effect that arises in the estimate of their total magnitude: at small S/N, the Kron magnitudes progressively underestimate the real magnitude of the detected galaxies. Our simulations indicate that at faint fluxes this effect can easily be of 0.1-0.15 mags, and neglecting it would bias the estimate of the slope of the number counts. Finally, it includes also the effect of fainter sources that may enter in the number counts when falling on positive noise fluctuation. Simulated ob-

jects are indeed simulated up to $K \approx 27.5$, much deeper than real limits, and extracted following a power-law in the number counts, whose slope is varied until it matches the slope of the recovered number counts. We defer to Minowa et al. (2005) for more details.

The estimate of the correction depends on a critical assumption, namely the distribution of galaxy sizes. At the exquisite resolution of the HUGS images the difference between compact and point-like sources is measurable, as in the Minowa et al. (2005) data. To get a hint of the real size distribution of the galaxies at faint K band magnitudes, we have looked at the half-light radius (r_{hl}) of galaxies as measured by SExtractor in the H band WFC3 images in CANDELS-GOODS. While the r_{hl} of stars is $0.15''$, that is the instrumental limit of WFC3, we find that galaxies at $K \approx 26$ (the typical magnitude where incompleteness is effective in the deepest GOODS data, see below) have r_{hl} typically ranging from $0.15''$ to $0.3''$. There is also a clear trend with magnitudes, with brighter galaxies being even larger than $0.3''$, while the typical r_{hl} for galaxies at $K \approx 27$ appears to be much closer to $0.15''$, the value of unresolved objects. We therefore computed the correction for incompleteness in two cases: assuming point-like sources (as in Minowa et al. (2005)) and assuming a distribution of size between $0.1''$ and $0.3''$. Finally, to minimize the effect of correction, we have used the counts from the various images only when the incompleteness is negligible ($\leq 5\%$), at the only exception of the deepest areas in GOODS, that have been the only used (with the appropriate correction) to reach the faintest limits.

The final results is shown in Figure 12, where we show the HUGS number counts computed with the two assumptions about galaxy size, compared with a number of recent results from the literature (SSDF: Minowa et al. (2005), HDFS: Labbé et al. (2003), KDS: Moustakas et al. (1997), UDS: Cirasuolo et al. (2010), WHTDF: Cristóbal-Hornillos et al. (2003), HWDF: Huang et al. (1997), GOODS: Grazian et al. (2006)). Uncertainties have been estimated assuming simple Poisson error.

As expected, the number counts agree very well with previous results from the literature. It is immediately appreciated that the HUGS number counts exceed in depth and statistical accuracy all previous estimates at the faint side, with the only exception of the faintest bin by the SSDF. The latter used AO-assisted observations to reach a very small PSF, $0.18''$, that makes these observations more sensitive than HUGS to the detection of point-like sources. We note that, instead, their completeness levels assuming extended morphology is comparable to our own.

At the faint limit, it is immediately clear how dramatic the impact of the assumptions on galaxy size is. Assuming point-like sources we confirm the flattening of the number counts at $K=24-26$ detected by Minowa et al. (2005), in our case with a much larger statistical accuracy, given the $50\times$ larger field of view of our images. Assuming instead a typical galaxy size in the range $0.1''-0.3''$, we find that the slope of the number counts remains essentially unchanged up to $K \approx 26$, with a slope of about 0.18 (CHECK THIS NUMBER).

This difference in slope is potentially very important to establish the contribution of detected galaxies to the EBL. A robust determination of this effect, however, requires more refined simulation and analysis of the galaxy size at different magnitudes, possibly to be done with even deeper data as the HUDF12, that is beyond the scope of this paper.

Table 3. K-band galaxy number counts in the HUGS survey, corrected for incompleteness as described in the text. Number densities are given as galaxies per magnitude per square degree. The full table is available on the electronic version of the paper and at <http://www.oa-roma.inaf.it/HUGS>

Magnitude bin	$\log(N)^a$	$\log(\sigma(N))^a$	$\log(N)^b$	$\log(\sigma(N))^b$
24.00	4.9274	3.2965	4.9283	3.2969
24.25	4.9582	3.3285	4.9588	3.3288
24.50	5.0033	3.4016	5.0021	3.4010
24.75	5.0639	3.5820	5.0619	3.5810
25.00	5.0827	3.6783	5.0774	3.6757
25.25	5.1068	3.6904	5.1187	3.6964
25.50	5.1147	3.7863	5.1648	3.8114
25.75	5.1203	3.7891	5.2363	3.8471
26.00	5.1391	3.7985	5.3038	4.0063
26.25	5.1758	3.9423	-	-

Notes. ^(a) assuming unresolved morphologies for galaxies - see text for details ^(b) assuming a distributions of half-light radius from $0.1''$ to $0.3''$ for galaxies - see text for details

7. Summary

We have presented in this paper the ultra-deep images obtained with the Hawk-I imager at the VLT in the K and Y bands on the UDS and GOODS-South fields covered by the CANDELS survey. We dub this public surveys HUGS, an acronym for *Hawk-I UDS and GOODS Survey*. While the bulk of the data presented here comes from a program specifically designed to cover CANDELS, (The Large Program 186...) we have also included in our analysis other data coming from previous observations on GOODS-S, that were acquired either during the Science Verification Phase () or in the framework of another program designed to look for $z \approx 7$ galaxies (Castellano et al 2010, 182...). These data comprise nearly all the available Hawk-I data on GOODS and are discussed and made publicly available here.

This paper describes the survey strategy, the observational campaign and the data reductions process. For the latter, we refer the reader to the previous sections, that give full details. We simply mention here that we have followed standard recipes for these images, adopting a number of validation controls. First, we have used two independent pipelines (one developed in Rome and one in Edinburgh) to reduce the first epoch of data, and cross-compared their results, that turned out to be consistent. In addition, we used the large wealth of independent images acquired over each field to internally validate the data. At the end of these tests, we are confident that the observation uncertainties are under control, typically at the level of few percent.

Similarly, full details of the observing strategy and resulting data are described in the text. We refer the reader in particular to Tables and Figures for further informations. We summarize here the fundamental aspects of our survey.

- HUGS has been designed to complement the CANDELS data in the two fields where crucial IR bands are missing: K band in GOODS-S and Y and K bands in UDS. The depth has been tuned in order to match the depth of the WFC3 images. For instance, the K-band limit is set about $0.5 mag$ shallower than H_{160} ones, to match the average $H - K$ color of faint galaxies.
- Pointings have also been optimized to cover CANDELS. In the UDS, we covered 85% of the CANDELS area with 3 different, marginally overlapping Hawk-I pointings. For the K band in GOODS-S we adopted a more complicated pattern, made of 6 different pointings (two “Deep” and four “Wide”) that is described in Figure2. This has allowed us to vary the exposure

time over the field in order to match the varying depth of the CANDELS images. Y band in GOODS-S covers about 60% of the Eastern CANDELS area.

- In the UDS, the exposure times of each pointing are of about 13 hours in the K band and 8 in the Y band. The seeing is $0.37''$ - $0.43''$ in the K band and $0.45''$ - $0.5''$ in the Y band. The corresponding limiting magnitudes are $m_{lim}(K) \simeq 26$, $m_{lim}(Y) \simeq 26.8$ (5σ in one FWHM) or $m_{lim}(K) \simeq 27.3$, $m_{lim}(Y) \simeq 28.3$ (1σ per arcsec²).

- In GOODS-S, the total exposure time in the K band (summed over six pointings) is of 107 hours. Because of the complex geometry, this corresponds to an exposure of 60-80 hours in the central area (the one covered by CANDELS “Deep”) and 12-20 hours in the rest (the CANDELS “Wide area”). The final average seeing is remarkably good and constant, with 4 pointings at $0.38''$ (notably including the two deepest) and 2 pointings at $0.42''$. On the finally stacked images, the limiting magnitudes in the deepest area are $m_{lim}(K) \simeq 27$. (5σ in one FWHM) or $m_{lim}(K) \simeq 28.3$, $m_{lim}(Y) \simeq 28.3$ (1σ per arcsec²).

- We have derived from the HUGS images two different flavors of catalogues - a K-band selected catalogue, that we use estimate number counts, and PSF-matched catalogues for all the H-selected galaxies detected in the CANDELS images. For the UDS, the corresponding catalog published in Galametz et al 2013 already uses the full HUGS data. On the contrary, the catalog for GOODS-S catalog obtained and made available here supersedes the catalog already published in Guo et al 2013, that used only a preliminary version of the HUGS data, obtained from shallower observations of the central area (pointings D1 and D2) only.

- The crucial goal of matching the CANDELS depth is fully achieved, as shown in Fig. 11, where it is shown that a large fraction of the CANDELS objects is detected in K even at the faintest H band limits of the H-band image.

We finally present the number counts in the K band, as obtained after combining the two fields. We describe the simulations that we have adopted to correct for the incompleteness at the faintest limits, with different assumptions for the size distribution of galaxies. We show that our number counts extend to magnitude limits fainter than any previous survey, with the only exception of the AO-assisted images of the SSDF field. The latter have a FWHM of $0.18''$ and exposure times comparable to our survey, and hence reach fainter limits for unresolved objects, but on such a small area (1 sqarcmin) to be very uncertain in a statistical sense. We show that the slope of the number counts at the faintest sides depends sensitively on the assumed distribution of galaxy sizes. It ranges from 0.18, if we assume that galaxies at $K > 26$ are unresolved, to 0.22, if we assume that such galaxies have a typical distribution of half-light radii from $0.1''$ to $0.3''$, as suggested by a preliminary analysis of the deepest CANDELS images.

We made all the HUGS images and derived catalogs publicly available at the web site <http://www.oa-roma.inaf.it/HUGS>.

Acknowledgements. This work wouldn't have been done without the support and dedication of the whole ESO staff. We thank in particular our support astronomer Monika Petr-Gotzens. We are also grateful to the memory of Alan Moorwood, who was fundamental to motivate the development of Hawk-I, on which our survey is based. AF and JSD acknowledge the contribution of the EC FP7 SPACE project ASTRODEEP (Ref.No: 312725). This work uses data from the following ESO programs: 60.A-9284, 181.A0717, 186.A-0898.

References

- Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 281, *Astronomical Data Analysis Software and Systems XI*, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228
- Castellano, M., Fontana, A., Boutsia, K., et al. 2010a, *A&A*, 511, A20
- Castellano, M., Fontana, A., Paris, D., et al. 2010b, *A&A*, 524, A28
- Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2010, *MNRAS*, 401, 1166
- Cristóbal-Hornillos, D., Balcells, M., Prieto, M., et al. 2003, *ApJ*, 595, 71
- Ellis, R. S. 1997, *ARA&A*, 35, 389
- Fontana, A., D'Odorico, S., Poli, F., et al. 2000, *AJ*, 120, 2206
- Galametz, A., Grazian, A., Fontana, A., et al. 2013, *ApJS*, 206, 10
- Grazian, A., Fontana, A., de Santis, C., et al. 2006, *A&A*, 449, 951
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24
- Heidt, J., Appenzeller, I., Gabasch, A., et al. 2003, *A&A*, 398, 49
- Huang, J.-S., Cowie, L. L., Gardner, J. P., et al. 1997, *ApJ*, 476, 12
- Kissler-Patig, M., Pirard, J.-F., Casali, M., et al. 2008, *A&A*, 491, 941
- Labbé, I., Franx, M., Rudnick, G., et al. 2003, *AJ*, 125, 1107
- Minowa, Y., Kobayashi, N., Yoshii, Y., et al. 2005, *ApJ*, 629, 29
- Moustakas, L. A., Davis, M., Graham, J. R., et al. 1997, *ApJ*, 475, 445
- Sawicki, M. & Thompson, D. 2005, *ApJ*, 635, 100
- Targett, T. A., Dunlop, J. S., McLure, R. J., et al. 2011, *MNRAS*, 412, 295
- Vanzella, E., Pentericci, L., Fontana, A., et al. 2011, *ApJ*, 730, L35+
- Williams, R. E., Baum, S., Bergeron, L. E., et al. 2000, *AJ*, 120, 2735
- Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, *AJ*, 112, 1335

LBT/MODS1 SPECTROSCOPIC CONFIRMATION OF TWO FAINT SOURCES AT $Z = 6.4$ MAGNIFIED BY THE CLASH / FRONTIER FIELDS CLUSTER MACSJ0717.5+3745: TOWARDS THE CHARACTERIZATION OF STAR-FORMING GALAXIES AT THE EPOCH OF REIONIZATION¹

E. VANZELLA², A. FONTANA³, A. ZITRIN^{4,†}, D. COE⁵, L. BRADLEY⁵, M. POSTMAN⁵, A. GRAZIAN³, M. CASTELLANO³, L. PENTERICCI³, M. GIAVALISCO⁶, P. ROSATI⁷, M. NONINO⁸, R. SMIT⁹, I. BALESTRA⁸, R. BOUWENS⁹, S. CRISTIANI⁸, E. GIALONGO³, W. ZHENG¹⁰, L. INFANTE¹¹, F. CUSANO², R. SPEZIALI³

²INAF – Bologna Astronomical Observatory, via Ranzani 1, I-40127 Bologna, Italy

³INAF - Rome Astronomical Observatory, Via Frascati 33, I-00040 Monteporzio Roma, Italy

⁴Cahill Center for Astronomy and Astrophysics, California Institute of Technology, MS 249-17, Pasadena, CA 91125, USA

⁵STScI, 3700 San Martin Dr., Baltimore, MD 21218, USA

⁶Department of Astronomy, University of Massachusetts, Amherst MA 01003, USA

⁷Dipartimento di Fisica e Scienze della Terra, Università di Ferrara, via Saragat 1, 44122 Ferrara, Italy

⁸INFN, National Institute of Nuclear Physics, via Valerio 2, I-34127 Trieste, ITALY

⁹Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

¹⁰Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

¹¹Institute of Astrophysics, Pontificia Universidad Católica de Chile, V. Mackenna 4860, 22 Santiago, Chile and

[†]Hubble Fellow

Draft version December 20, 2013

ABSTRACT

We report the LBT/MODS1 spectroscopic confirmation of two images of faint Lyman alpha emitters at $z = 6.4$ behind the Frontier Fields galaxy cluster MACSJ0717.5+3745. A wide range of lens models suggests that the two images are highly magnified, with a strong lower limit of $\mu > 5$. These are the faintest $z > 6$ candidates spectroscopically confirmed to date. These may be also multiple images of the same $z = 6.4$ source as supported by their similar intrinsic properties, but the lens models are inconclusive regarding this interpretation. To be cautious, we derive the physical properties of each image individually. Thanks to the high magnification, the observed near-infrared (restframe ultraviolet) part of the spectral energy distributions and $Ly\alpha$ lines are well detected with $S/N(m_{1500}) \gtrsim 10$ and $S/N(Ly\alpha) \simeq 10 - 15$. Adopting $\mu > 5$, the absolute magnitudes, M_{1500} , and $Ly\alpha$ fluxes, are fainter than -18.7 and $2.8 \times 10^{-18} \text{ ergs}^{-1} \text{ cm}^{-2}$, respectively. We find a very steep ultraviolet spectral slope $\beta = -3.0 \pm 0.5$ ($F_\lambda = \lambda^\beta$), implying that these are very young, dust-free and low metallicity objects, made of standard stellar populations or even extremely metal poor stars (age $\lesssim 30$ Myr, $E(B-V)=0$ and metallicity $0.0 - 0.2Z/Z_\odot$). The objects are compact ($< 1 \text{ kpc}^2$), and with a stellar mass $M_\star < 10^8 M_\odot$. The presence of the $Ly\alpha$ line and the intrinsic FWHM ($< 300 \text{ kms}^{-1}$) of these newborn objects do not exclude a possible leakage of ionizing radiation. We discuss the possibility that such faint galaxies may resemble those responsible for cosmic reionization.

Subject headings: cosmology: observations — galaxies: formation — galaxies: evolution — galaxies: distances and redshifts

1. INTRODUCTION

The investigation of the distant Universe and the processes that led to the reionization of the intergalactic medium, are amongst the major goals of observational cosmology (Robertson et al. 2010). While there are tens (a few) spectroscopic confirmations of galaxies at redshift 6(7) (e.g., Vanzella et al. 2009, 2011), accessing the faint-luminosity regime down to $\lesssim 0.1L^*$ remains challenging even with 8-10m class telescopes, especially for $z > 6$. Before the advent of next generation observatories like JWST and the extremely large telescopes, the only viable way to pursue extremely faint distant objects, and investigate the nature of their stellar populations (even PopIII), is to exploit strong lensing magnification (e.g.,

Zackrisson et al. 2012, 2013). To this aim, Bradley et al. (2013) (B13, hereafter) selected magnified candidate galaxies at redshift 6 – 8 fully exploiting the 16-bands photometry of the CLASH survey (Postman et al. 2011), and found agreement down to ~ 27 mag with the UV luminosity functions of blank fields. After the completion of the CLASH program, the investigation of the high- z universe is now continuing with the ultradeep HST Frontier Fields campaign (FF hereafter), that includes four CLASH galaxy clusters².

Accessing the faint luminosity regime ($L < 0.1L^*$) at $z > 6$ is crucial in the context of cosmic reionization (e.g., Fontanot et al. 2013): faint galaxies dominate the global ultraviolet luminosity density (Bouwens et al. 2007) and possibly have an escape fraction of ionizing radiation larger than the brighter counterparts (e.g., Ferrara & Loeb 2013; Yajima et al. 2011).

Here we report on the LBT/MODS1 spectroscopic confirmation of two faint $z = 6.4$ sources, significantly magnified by the FF galaxy cluster MACSJ0717.5+3745,

¹ The Large Binocular Telescope (LBT) is an international collaboration among institutions in the United States, Italy and Germany. LBT Corporation partners are: The University of Arizona on behalf of the Arizona university system; Istituto Nazionale di Astrofisica, Italy; LBT Beteiligungsgesellschaft, Germany, representing the Max-Planck Society, the Astrophysical Institute Potsdam, and Heidelberg University; The Ohio State University, and The Research Corporation, on behalf of The University of Notre Dame, University of Minnesota, and University of Virginia.

² <http://www.stsci.edu/hst/campaigns/frontier-fields/>

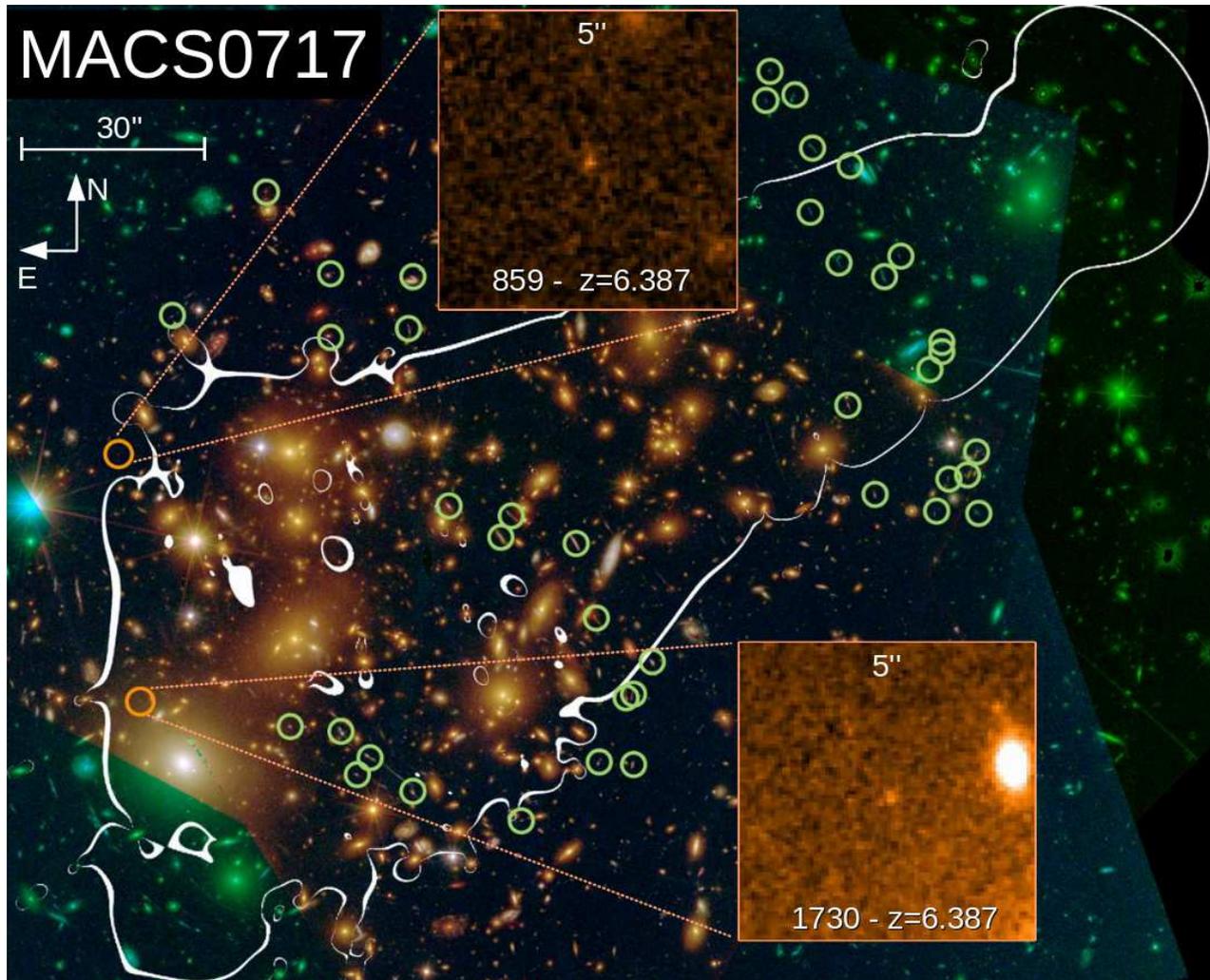


FIG. 1.— Figure shows the 16-band CLASH RGB false-color image of MACSJ0717.5+3745, with the two $z = 6.4$ spectroscopically confirmed images marked with *red circles* (the insets show the J125 zoom). The critical curves ($\mu > 100$ here) for a source at $z_s = 6.4$ from the revised Zitrin et al. model are overlaid in *white*. The *green circles* mark the multiple images used as constraints (see Zitrin et al. 2009, Limousin et al. 2012, Medezinski et al. 2013). As can be seen, the two $z = 6.4$ objects lay (a) close to the critical curves, and (b) in regions in which there are hardly other multiple images known, so that the exact position of the critical curves is not perfectly constrained. The proximity to the critical curves results in very high magnifications, of the order of few to few dozen, and correspondingly, large errors on these estimates. Still, all probed models (see §2.1) yield $\mu > 5$ for both images, which we have adopted throughout this work as our lower limit, for conservative results.

study their physical properties, and discuss the contributions of such objects to the reionization of the IGM.

Throughout this paper magnitudes are reported in the AB scale and the world model, when needed, is a flat universe with density parameters $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

1.1. Target selection and magnification

B13 selected 15 magnified $z \simeq 6$ galaxy candidates behind the FF galaxy cluster MACSJ0717.5+3745, by using their drop-out features and corresponding photometric redshift estimate. In particular, we report here the spectroscopic observations of two candidates from their sample, macs0717_0859 and macs0717_1730 (859 and 1730 for short, hereafter), with photometric redshifts of 6.1 ± 0.2 and $6.0^{+0.2}_{-0.3}$, respectively (and see Table 1). The magnifications reported in B13 were $\mu = 15.6$ (859) and $\mu > 100$ (1730) (i.e. the latter unconstrained since

the object is too close to the critical curves). The magnification estimates were based on the revised lens model by Zitrin et al. (2009; see also Medezinski et al. 2013) who first performed the strong-lensing analysis for this cluster, uncovering that is the largest magnifying lens known to date (see Figure 1). Here we have also estimated the magnifications from several other lens models made for the Frontier Fields program (including a refurbished version of the Zitrin et al. model used in B13), by running the Magnification Calculator³ available online. The estimate from different groups, methods and assumptions span the range between 5 and 70, with some solutions even higher than 100 within the 68% confidence interval. The medians among the different models are: $\mu = 17.4^{+25}_{-13}$ ($^{+50}_{-12}$) for 1730 and $\mu = 6.9^{+1}_{-1}$ ($^{+30}_{-2}$) for 859, where statistical and systematic errors (in parentheses)

³ <http://archive.stsci.edu/prepds/frontier/lensmodels/>

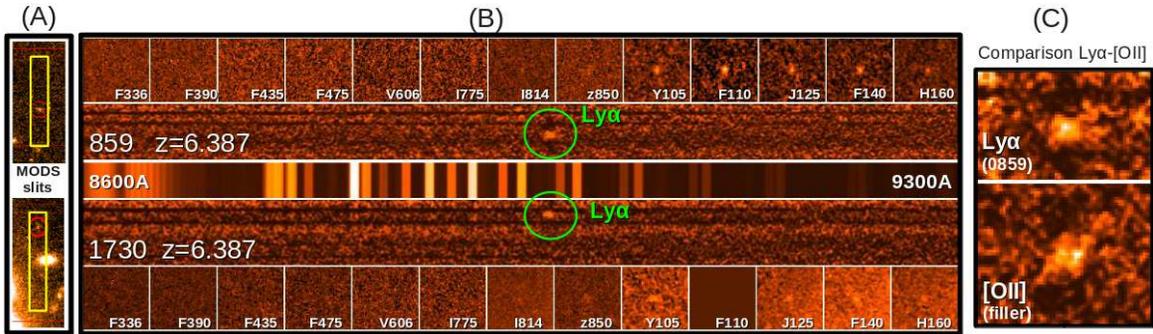


FIG. 2.— **Panel (A)**: the position of targets 859 (top) and 1730 (bottom) in the MODS1 slits over the J125 band are shown. **Panel (B)**: the two dimensional spectra, the $Ly\alpha$ lines (marked with a green circle), the sky spectrum and the CLASH multi-band cutouts ($1.5''$) are shown. **Panel (C)**: The $Ly\alpha$ line of 859 compared with a low- z [OII] doublet identified in the same mask (filler object) is shown.

are quoted. The models for this lens are still not fully constrained in the regions where the two $z = 6.4$ are detected, both due to proximity to the critical curves, and, lack of multiple-images constraints nearby.

We also acknowledge the possibility that the two sources presented here are actually counter images of a single background galaxy, as some of the models above predict counter images within few, to dozen arcseconds, from the location of the other $z = 6.4$ object. We did not detect, however, any additional counter images where the models predict them (although possibly, due to lesser magnification where other images are predicted). As not all models predict counter images, and predicted counter images were not identified in the data, it cannot be unambiguously determined if indeed the two objects are images of the same source. What is relevant here, though, is the agreement among the different models that the sources are strongly magnified ($\mu > 5$), and the single or double nature does not alter our findings on the derived physical properties. In the following, to be most conservative, we derive rest-frame quantities by adopting $\mu = 5$ for both sources, and express the results in terms of $\mu_5 = \mu/5$.

2. DATA AND SAMPLE SELECTION

2.1. Spectroscopic observations with LBT/MODS1

The spectroscopic observations have been performed in dual mode with the MODS1 instrument at the LBT, that exploits the two red (5800-10300Å) and blue (3200-6000Å) channels, yielding a total spectra coverage from 3200 to $\sim 10300\text{\AA}$ on source. The red G670L and blue G400L grisms with a slit width of $1''$ have been adopted, providing a spectral resolution of $R \simeq 1500$ for both. Science frames of 1200s have been acquired with a dithering pattern of $1.5''$ shift along the slit for a total integration time of 16800s for 859, and 11200s for 1730. The average seeing conditions were $\simeq 1.0''$. Data reduction has been performed with the MODS1 spectroscopic reduction pipeline based on VIPGI tasks (Scodreggio et al. 2005).⁴ In the two slits located on 859 and 1730, two emission lines are clearly detected at 8980Å and 8981Å, respectively, with observed fluxes of $1.4 \times 10^{-17} \text{ergs}^{-1} \text{cm}^{-2}$ (with $S/N = 15$) and $\simeq 1.0 \times 10^{-17} \text{ergs}^{-1} \text{cm}^{-2}$ (with $S/N = 9$), respectively (see Figures 2 and 3).

⁴ <http://lbt-spectro.iasf-milano.inaf.it/pipelines/Info/>

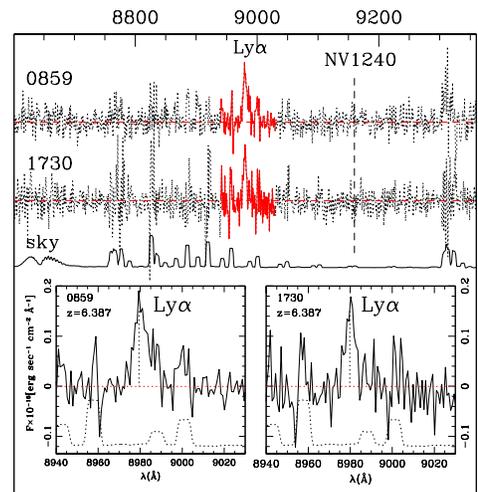


FIG. 3.— One dimensional spectra of 859 and 1730 (top dotted). The highlighted red parts of the spectra are zoomed in the bottom panels with the sky spectrum (dotted). The position of the NV1240 line is also shown with a vertical dashed line.

3. RESULTS

- *Nature of the lines:* The large spectral coverage (3200 – 10300Å) allows us to exclude low redshift solutions like $H\alpha$ at $z = 0.37$ or [OIII]5007 at $z = 0.79$, that would be in contrast with the single line detection. The only possible degeneracy is among [OII]3727 and $Ly\alpha$. However, [OII]3727 can be reliably excluded because of the following reasons: (1) the doublet [OII]3726-3729 is resolved in the present observations (see an example in Figure 2, panel C) and (2) the observed equivalent width (see below) of the lines is not compatible with the typical values observed at $z < 1.5$, i.e., they are too large (e.g., Vanzella et al. 2009 and their Fig. 12). Moreover, source 859 shows an asymmetric line profile toward the red wavelengths (Figure 3), that is typical of this transition at high redshift. The spectrum of 1730 is slightly shallower (11200s) and noisier than 859 (close to the edge of the slit), and prevents us from detecting the asymmetric shape, but the line width and the equivalent width are not consistent with the [OII]3727 doublet.

Therefore we conclude that the two emission lines are

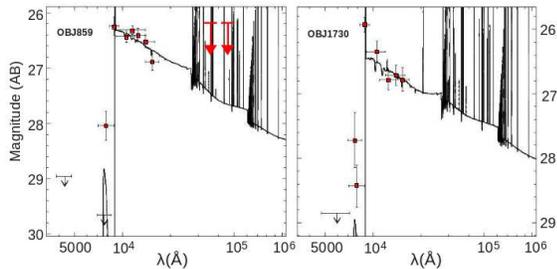


FIG. 4.— SED fits with BC03 templates are shown. Nebular emission lines are included in the fit. The two arrows for 859 are 1-sigma lower limits of IRAC 3.6 μ m and 4.5 μ m channels.

$Ly\alpha$ at the same redshift 6.387 ± 0.002 . The striking accordance of the two redshifts may add support to the hypothesis that these two objects are multiple images of the same background source. If confirmed, this could provide further constraints to the lens model and therefore deserves future investigation and lens remodeling, which is out of the scope of the present work. In the following we assume that these are two individual objects and look at the properties of each of them separately.

- *Rest frame UV continuum luminosity at 1500Å*: As mentioned above the wide spread on the magnifications allow us to identify an interval of possible luminosities. Given the observed Y105 magnitudes ($\simeq 1500\text{\AA}$) of 26.42 ± 0.11 for 859 and 26.34 ± 0.16 for 1730, the two sources have unlensed luminosities of $L_{1500} \simeq 0.25 \mu_5^{-1} L_{z=6}^*$, adopting $L_{z=6}^*$ from Bouwens et al. (2007). It is worth noting that if $\mu > 10$, these galaxies would have $L_{1500} < 0.07 L_{z=6}^*$. However even in the more conservative case ($\mu > 5$), these are among the faintest ones spectroscopically confirmed at these redshifts with such a high signal to noise (Balestra et al. 2013; Bradac et al. 2012; Schenker et al. 2012).

- *Equivalent widths and FWHM of the lines*: The continuum is not detected in the spectra. Therefore, we derive the continuum level under the $Ly\alpha$ transition by using the closest HST band not including the line (Y105), and correcting for the UV slope β (see below). The rest-frame EWs of 859 and 1730 are $45 \pm 7\text{\AA}$ and $32 \pm 10\text{\AA}$, respectively. These are typical values if compared with those observed at similar redshifts among Lyman break galaxies or $Ly\alpha$ emitters (Stark et al. 2011). The FWHM of the lines is also quite modest, after correcting for the instrumental profile and doubling them for IGM absorption (e.g., Laursen et al. 2011), it turns out to be $\lesssim 200\text{ km s}^{-1}$.

- *Ultraviolet spectral slope β ($F_\lambda = \lambda^\beta$)*: Following Castellano et al. (2012) and Bouwens et al. (2013), ultraviolet spectral slopes have been estimated by fitting the near infrared WFC3 magnitudes redward the $Ly\alpha$ line, using the Y105, J125, F140W and H160 bands (for 859 the F110W band was also available and has been included in the fit). Being achromatic, strong lensing is not affecting the colors of the sources. The emerging slopes for 859 and 1730 are very steep, $\beta = -3.02 \pm 0.37$ and $\beta = -3.01 \pm 0.56$. Interestingly, this similarity is consistent with the option that these two objects are multiple images. While the source 1730 is close to a bright galaxy and its photometry has to be taken with caution, source 859 is isolated and with reliable colors (Figure 2).

TABLE 1. OBSERVED AND PHYSICAL PARAMETERS FOR 859 AND 1730.

Quantity	macs0717_0859	macs0717_1730
R.A. (J2000)	07:17:38.18	+37:45:16.9
Decl. (J2000)	07:17:37.85	+37:44:33.7
Redshift	$6.387(\pm 0.002)$	$6.387(\pm 0.003)$
Y105(observed)	$26.42(\pm 0.11)$	$26.34(\pm 0.16)$
H160(observed)	$26.88(\pm 0.15)$	$26.78(\pm 0.18)$
H160(unlensed)	$28.63+2.5\text{Log}_{10}(\mu_5)$	$28.53+2.5\text{Log}_{10}(\mu_5)$
β_{UV}	$-3.02(\pm 0.37)$	$-3.01(\pm 0.56)$
$SFR_{UV}(M_\odot/\text{yr})$	$1.6 [1-3] \mu_5^{-1}$	$2 [1-5] \mu_5^{-1}$
$SFR_{Ly\alpha}(M_\odot/\text{yr})$	$1.2 [1.0-1.4] \mu_5^{-1}$	$1 [0.8-1.2] \mu_5^{-1}$
$M_*(\times 10^7 M_\odot)$	$4 [2-12] \mu_5^{-1}$	$2 [2-22] \mu_5^{-1}$
E(B-V)	$0.0 [0.0-0.06]$	$0.0 [0.0-0.1]$
age (Myr)	$25 [10-100]$	$10 [10-250]$
Met. (Z)	0.02	0.2
$\times L_{z=6}^*$	$0.25 \mu_5^{-1}$	$0.27 \mu_5^{-1}$
Area (sq.kpc)	$0.8 \mu_5^{-1}$	$0.7 \mu_5^{-1}$
FWHM($Ly\alpha$)(km/s)	$100 [70-130]$	$140 [100-180]$
$EW_{rest}(Ly\alpha)(\text{\AA})$	$45 [38-53]$	$32 [22-42]$
flux($Ly\alpha$) $\times 10^{-18}$	$2.8 [2.6-3.0] \mu_5^{-1}$	$2.0 [1.8-2.3] \mu_5^{-1}$

NOTE. — $Ly\alpha$ fluxes are in units of $\text{ergs}^{-1}\text{cm}^{-2}$. Physical properties refer to BC03 models with nebular emission. Quantities related to $Ly\alpha$ do not include possible IGM absorption. In parentheses the 68% confidence interval (statistical) is reported and $\mu_5 = 1$ corresponds to $\mu = 5$.

As noted by Z13, the possible presence of the 2175Å dust feature may interfere with the estimate of the UV slope. However, we tend to exclude this possibility on the basis of the redshift and the $Ly\alpha$ emission lines that favor a low dust attenuation.

- *Size of the sources*: As reported in B13, the two sources are resolved in the HST/WFC3 images. Their isophotal areas (provided by SExtractor) converted into physical units are $0.8\mu_5^{-1}$ and $0.7\mu_5^{-1}$ sq.kpc. If they are two distinct objects, we estimate a proper separation of ~ 30 kpc in the source plane at $z = 6.387$.

- *AGN activity*: At $z = 6.387$ the expected NV1240 line is not detected (see Figure 3). The 1-sigma upper limit $NV1240/Ly\alpha$ is < 0.07 (typical value for AGNs is 10%, Alexandroff et al. 2013). Considering the above ratio, that only 5% of high redshift LAEs are possible AGNs (Malhotra et al. 2003), and that they are spatially resolved, we conclude the $Ly\alpha$ emission is due to SF activity.

4. DISCUSSION AND CONCLUSIONS

As described above, the two discovered sources (or a single one in the case of multiple images) are the faintest galaxies at $z > 6$ ever observed with a spectroscopic redshift confirmation and well detected $Ly\alpha$ lines and SEDs. The investigation of *new* luminosity regimes through the strong-lensing magnification gives the opportunity to explore possible new physical conditions.

4.1. Nature of the stellar populations

We examine their rest-frame properties through a SED analysis. We first derive physical parameters assuming ordinary stellar populations, i.e. by comparing the observed SED with a set of Bruzual & Charlot (2003) templates (BC03), assuming Salpeter IMF, metallicity of 0.02, 0.2, 1.0 Z/Z_\odot , and E(B-V) spanning the range [0.0 – 1.0]. The current 1- σ lower limits from IRAC

($3.6\mu\text{m}$ and $4.5\mu\text{m}$ channels) for 859 are $\simeq 26.1\text{AB}$, too shallow to provide solid constraints on $[\text{OIII}]5007+H\beta$ and $H\alpha$ nebular emissions. The other source 1730 is contaminated by close brighter sources. The SED fitting with BC03 includes nebular line and continuum emission following Schaerer & de Barros (2009) (see Castellano et al. 2014 for further details). The output of this exercise is listed in Table 1. Regardless of the adopted μ , the two sources turn out to be very small ($\lesssim 1$ sq.kpc), with low SFRs ($\simeq 1 - 2M_{\odot}/\text{yr}$) and low stellar masses of $< 10^8 M_{\odot}$. The properties related to colors (i.e., independent from the magnification μ), such as dust attenuation, age, and metallicity, are consistent with newborn objects (see Table 1).

It is interesting to explore also the possible presence of extremely metal poor (EMP, $Z \simeq 1/2000Z_{\odot}$) and PopIII stars ($Z = 0$). For this purpose we consider the SED fitting and the predicted HST/WFC3 colors provided by Inoue et al. (2011) and Zackrisson et al. (2013), (I11 and Z13, respectively), that also include nebular contribution. The observed UV slope is compatible either with very young, but still standard (PopII) stellar populations (BC03), or with EMP/PopIII stars. In particular, 859 (with the most reliable photometry) has a $(J - H)$ color of -0.48 ± 0.19 that is consistent with an age $< 10\text{Myr}$ and $\log_{10}(Z/Z_{\odot})$ even smaller than -6 (as shown in I11). Similarly, compared with the models of Z13 and Raiter et al. (2010), the slope and colors are compatible with ages $< 10\text{Myr}$ and very metal poor stars. Conversely, the modest $Ly\alpha$ EWs would suggest that we are dealing with standard stellar populations, given that PopIII stars are often associated with $Ly\alpha$ EW $\sim 500 - 1500\text{\AA}$ rest-frame (Schaerer et al. 2003; Raiter et al. 2010; I11). Another possibility is that the $Ly\alpha$ EW could be lowered for extremely metal poor ($Z/Z_{\odot} < 10^{-4}$) and even Pop III ($Z/Z_{\odot} = 0$) galaxies if $f_{esc} > 0$ (Zackrisson et al. 2013). For example, I11 found a $Ly\alpha$ EW of $\simeq 65\text{\AA}$ for $Z = 0$ and 10Myr constant star formation, when $f_{esc} = 0.9$.

Another property of the sources is that $\text{SFR}(Ly\alpha) \gtrsim \text{SFR}(UV)$, adopting the standard Kennicutt conversions (Kennicutt 1998). As discussed in Verhamme et al. (2008), this would be in line with an $E(B-V) \simeq 0$ and $f_{esc}(Ly\alpha)$ close to unity (see Atek et al. 2008, 2013). However, the Kennicutt relations may not be valid here, i.e., in the case of a young burst or constant SF for short timescales and/or the assumed IMF. In fact the ratio $\text{SFR}(Ly\alpha)/\text{SFR}(UV)$ can be as high as 4 in the very beginning, 1-50 Myr (Verhamme et al. 2008 Fig.15, based on synthesis models of Schaerer 2003), so that the IGM can play an important role to lower the above ratio (damping part of the $Ly\alpha$ line).

To summarize, such blue UV colors in these $z \sim 6.5$, sub-luminous galaxies are in line with the general trend of bluer UV colors with decreasing luminosity observed at high-redshift (e.g. Vanzella et al. 2009; Reddy et al. 2011; Balestra et al. 2013). While it is hard to make definitive statements about the populations content of these galaxies given the limited information we have about them, we observe that a relative low-efficiency star formation activity in these low-mass systems, and its corresponding low time scale for chemical enrichment compared to their more massive counterparts, could ex-

plain their very blue UV colors even without having to invoke PopIII stars, although we certainly cannot exclude their presence in the stellar populations of these sources (e.g. Finkelstein et al. 2010). These kind of galaxies could be examples of very low chemical enrichment, dust-free systems, barely higher than the pristine gas that is probably still feeding their activity of star formation.

4.2. Cosmic reionization

Regardless of the nature of the stellar populations, the potential role these sources have in the framework of cosmic reionization is intriguing. It is believed that the abundant, fainter galaxies ($M_{\star} < 10^9 M_{\odot}$) could significantly contribute to, or even be the dominant populations in, providing the ionizing radiation (e.g., Fontanot et al. 2013; Ferrara & Loeb 2013; Yajima et al. 2011; Razoumov & Sommer-Larsen 2010; but see Gnedin et al. 2008).

The direct measure of the escape fraction of ionizing radiation (f_{esc}) is in principle possible at $z < 4$. At higher redshifts the measurement is unfeasible due to the complete IGM attenuation of the Lyman continuum. Nonetheless it is worthwhile to investigate each of the main components that build the f_{esc} quantity. As discussed in Vanzella et al. (2012), the f_{esc} parameter is the product of the gas opacity $T_{900}^{HI} = \exp(-\tau_{900})$ and the dust attenuation $T_{dust} = 10^{-0.4 \times A_{900}}$. To first order, Lyman continuum emitters should have both low dust content A_{900} and low optical depth τ_{900} . Interestingly, the two sources described in this letter could match such requirements. First, given the very steep UV continuum the term related to dust attenuation is higher than zero ($T_{dust} = 10^{-0.4 \times A_{900}} > 0$, i.e., $E(B-V)=0$). Second, even though addressing the gas attenuation in the interstellar medium with current data is admittedly less reliable, it is worth noting that the presence of $Ly\alpha$ emission would not be in contrast with a $f_{esc} > 0$. As discussed in Nakajima & Ouchi (2013), the EW($Ly\alpha$) remains almost unchanged if f_{esc} is $\simeq 0-0.8$. In addition, taking advantage of radiative transfer models of Schaerer et al. (2011), the relatively small FWHM of the $Ly\alpha$ lines ($\lesssim 300\text{km s}^{-1}$, see Table 1) would be in line with the expected low mass and a possible low HI column density in front of the stars. The reason is that $Ly\alpha$ resonance scattering is less effective if N_{HI} is low, and photons escape easily along the shorter path, decreasing the velocity dispersion. This, could, in turn, be the telltale of efficient feedback in these systems, capable of either sweeping away or ionize a significant fraction of the gas surrounding the stars, a mechanism advocated by theoretical models in low-mass halos to self-regulate star formation. As a consequence, a proportionally higher fraction of ionizing radiation could be leaking out of these systems compared to their more massive counterparts and be available to keep the IGM ionized. Even though the derivation of the gas opacity of the interstellar medium is highly uncertain here, the two objects (or the single one, if they are multiple images) are still consistent with the possibility that $f_{esc} > 0$.

Regarding the single or multiple nature of the sources, the similarity in the physical and observed characteristic we derived in this work, would support they are multiple images of a single background $z = 6.4$ galaxy, but the different mass models we examined remain inconclusive regarding this option.

As discussed in Zackrisson et al. (2012) and Z13, such galaxies represent the ideal candidates for future near- and mid-infrared spectroscopic observations, especially in the investigation of the interplay between the UV slopes and the equivalent width of $H\beta$ lines, and its relation to the f_{esc} parameter. Future facilities such as JWST and extremely large telescopes will address this issues.

We acknowledge the support from the LBT-Italian Coordination Facility for the execution of observations, data distribution and reduction. We thank Gianni Zamorani, Marco Mignoli and Francesco Calura for useful discus-

sions. This work utilizes gravitational lensing models produced by PIs Bradac, Ebeling, Zitrin & Merten, Sharon, and Williams funded as part of the HST Frontier Fields program conducted by STScI. STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555. The lens models were obtained from the Mikulski Archive for Space Telescopes (MAST). Support for AZ is provided by NASA through Hubble Fellowship grant #HST-HF-51334.01-A awarded by STScI. AF acknowledges the contribution of the EC FP7 SPACE project ASTRODEEP (Ref.No: 312725). LI is partially supported by CATA-Basal, Conicyt.

REFERENCES

- Alexandroff, R., Strauss, M. A., Greene, J. E., Zakamska, N. L., Ross, N. P., Brandt, W. N., Liu, G.; Smith, P. S.; Ge, J., et al., 2013, MNRAS, 435,3306
- Atek, H., Kunth, D., Hayes, M., Ostlin, G., & Mas-Hesse, J. M. 2008, A&A, 488, 491
- Atek, H., Kunth, D., Schaerer, D., Mas-Hesse, J. M., Hayes, M., Ostlin, G., Kneib, J.-P., 2013, (arXiv1308.6577), ApJ, in press
- Balestra, I., Vanzella, E., Rosati, P., Monna, A., Grillo, C., Nonino, M., Mercurio, A., Biviano, A., Bradley, L., Coe, D., et al., 2013, A&A, 559, 9
- Bouwens, R. J., Illingworth, G. D., Franx, M., Ford, H., 2007, ApJ, 670, 928
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Labbe, I., van Dokkum, P. G., Trenti, M., Franx, M., Smit, R., Gonzalez, V., Magee, D., 2013, (arXiv1306.2950), ApJ, in press
- Bradac, M., Vanzella, E., Hall, N., Treu, T., Fontana, A., Gonzalez, A. H., Clowe, D., Zaritsky, D., Stiavelli, M., Clement, B., 2012, ApJ, 755, 7
- Bradley, L. D., Zitrin, A., Coe, D., Bouwens, R., Postman, M., Balestra, I., Grillo, C., Monna, A., Rosati, P., et al., (arXiv:1308.1692), ApJ, in press
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Castellano, M., Fontana, A., Grazian, A., Pentericci, L., Santini, P., Koekemoer, A., Cristiani, S., Galametz, A., et al., 2012, A&A, 540, 39
- Ferrara, A., Loeb, A., 2013, MNRAS, 431, 2826
- Finkelstein, S. L.; Papovich, C.; Giavalisco, M., Reddy, N. A., Ferguson, H. C., Koekemoer, A. M.; Dickinson, M., 2010, ApJ, 719, 1250
- Fontanot, F., Cristiani, S., Pfrommer, C., Cupani, G., Vanzella, E., 2013, (arXiv1312.0615), MNRAS, accepted
- Gnedin, N. Y., Kravtsov, A. V., Chen, H-W, 2008, ApJ, 672, 765
- Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189
- Inoue, Akio K., 2011, MNRAS, 415, 2920
- Limousin, M., Ebeling, H., Richard, J., Swinbank, A. M., Smith, G. P., Jauzac, M., Rodionov, S., Ma, C.-J., et al., 2012, A&A, 544, 71
- Laursen, P., Sommer-Larsen, J., Razoumov, Alexei O., 2011, ApJ, 728, 52
- Malhotra, S., Wang, J. X., Rhoads, J. E., Heckman, T. M., Norman, C. A., 2003, ApJ, 585, 25
- Medezinski, E., Umetsu, K., Nonino, M., Merten, J., Zitrin, A., Broadhurst, T., Donahue, M., Sayers, J., Waizmann, J.-C., et al., 2013, ApJ, 777, 43
- Nakajima, K., Ouchi, M., 2013, (arXiv1309.0207), MNRAS, in press
- Postman, M., Coe, D., Bentez, N., Bradley, L., Broadhurst, T., Donahue, M., Ford, H., Graur, O., Graves, G., et al., 2012, ApJS, 199, 25P
- Raiter, A., Schaerer, D., Fosbury, R. A. E., 2010, A&A, 523, 64
- Razoumov, Alexei O.; Sommer-Larsen, Jesper, 2010, ApJ, 710, 1239
- Reddy, Naveen A., 2011, Nature, 469, 479
- Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J., Stark, D. P., 2010, Nature, 468, 49
- Robertson, B. E., Furlanetto, S. R., Schneider, E., Charlot, S., Ellis, R. S., Stark, D. P., McLure, R. J., Dunlop, J. S., et al., 2013, ApJ, 768, 71
- Schaerer, D., 2003, A&A, 397, 527
- Schaerer, D. & de Barros, S., 2009, A&A, 502, 423
- Schaerer, D., Hayes, M., Verhamme, A., Teyssier, R., 2011, A&A, 531, 12
- Schenker, M. A., Stark, D. P., Ellis, R. S., Robertson, B. E., Dunlop, J. S., McLure, R. J., Kneib, J.-P., Richard, J., 2012, ApJ, 744, 179
- Scodeggio, M., Franzetti, P., Garilli, B., Zanichelli, A., Paltani, S., Maccagni, D., Bottini, D., Le Brun, V., et al., 2005, PASP, 117, 1284
- Stark, D. P., Ellis, R. S., Ouchi, M., 2011, ApJ, 728, 2
- Vanzella, E., Giavalisco, M., Dickinson, M., Cristiani, S., Nonino, M., Kuntschner, H., Popesso, P., Rosati, P., et al., 2009, ApJ, 695, 1163
- Vanzella, E., Pentericci, L., Fontana, A., Grazian, A., Castellano, M., Boutsia, K., Cristiani, S., Dickinson, M., et al., 2011, ApJ, 730, 35
- Vanzella, E., Guo, Y., Giavalisco, M., et al. 2012, ApJ, 751, 70
- Yajima, H., Choi, J-H, Nagamine, K., 2011, MNRAS, 412, 411
- Zackrisson, E., Inoue, A K., Jensen, H., 2013, ApJ, 777, 39
- Zackrisson, E., Zitrin, A., Trenti, M., Rydberg, C-E, Guaita, L., Schaerer, D., Broadhurst, T., Östlin, G., Ström, T., 2012, MNRAS, 427, 2212
- Zitrin, A., Broadhurst, T., Rephaeli, Y., Sadeh, S., 2009, ApJ, 707, 102