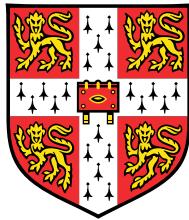


**PHOTOMETRIC REDSHIFTS
AND LUMINOUS
HIGH-REDSHIFT GALAXIES IN
THE DEEP FIELDS OF THE
DARK ENERGY SURVEY**



Anna Lidewij Schooneveld

Supervisors: Dr. M. Banerji, Prof. P.C. Hewett

Institute of Astronomy
University of Cambridge

This dissertation is submitted for the degree of
Master of Science

Magdalene College

May 2021

Abstract

In this thesis I explore photometric redshifts and high-redshift galaxies using multiwavelength datasets, in particular the combination of optical and near-infrared broadband photometry.

Firstly, I merge optical imaging from the deep fields of the Dark Energy Survey (DES) with near-infrared data from the Vista Deep Extragalactic Observations (VIDEO) survey, producing a 10-band ($grizZYJHK_s$) multiwavelength catalogue. I achieve this via forced photometry, where the deeper DES data is applied to determine detections which are used to extract flux measurements from the VIDEO imaging. I demonstrate that this technique increases the total number of objects with VIDEO data by 22% compared to simple position matching, and by as much as 34% in a sub-region where DES is especially deep compared to VIDEO. The final combined catalogue contains 2 152 575 unique objects over a footprint of 12.1 deg^2 , with 10.8 deg^2 reaching 5σ limiting magnitudes of at least $z_{5\sigma} = 24.9$, of which 4.3 deg^2 reaches at least $z_{5\sigma} = 26.2$.

Secondly, I compute photometric redshifts for the DES+VIDEO catalogue. I test 18 configurations for the template fitting code **LePHARE** (2 different template sets and 9 different options for each) and assess their performance compared to spectroscopic redshifts, measuring the bias, scatter, outlier fractions, and high-redshift accuracy. I find that the most optimal configuration at high redshifts adds emission lines to the templates, incorporates adaptive offsets, applies an $N(z)$ prior, uses fixed aperture photometry, and includes both optical and near-IR fluxes. I also study the implications of the photometric redshift tests for general (low-redshift dominated) applications. There, the most important result is that adding near-infrared fluxes strongly improves photometric redshifts, with lower scatter (σ lower by 1.1% and 25.5% for the two template sets, and σ_{68} lower by 45.9% and 56.4%) and lower outlier fractions ($f_{2\sigma}$ lower by 12.9% and 25.5%, and $f_{3\sigma}$ lower by 14.3% and 33.5%). I also determine a suitable star-galaxy separation method for the DES+VIDEO catalogue based on SED fitting. For this separation, I demonstrate that near-infrared data is also of crucial importance: without VIDEO data the stellar contamination rate increases by a factor of 22-35.

Lastly, I use the DES+VIDEO photometric redshifts to find bright ($m_{AB} < 25.0$) high-redshift ($z \gtrsim 5$) Lyman-break galaxies. Using a series of targeted cuts, I identify 34 $z \sim 5$ candidates at $z \sim 5$ and 36 candidates at $z \sim 6$.

Contents

List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 An overview of astronomical surveys	1
1.1.1 A brief history of astronomical surveys	1
1.1.2 Technological advancements: towards a data revolution	3
1.1.3 Deep-field surveys	6
1.1.3.1 Brief interlude on distances, redshifts and cosmic time	6
1.1.3.2 Small area space-based surveys	9
1.1.3.3 Intermediate-area ground-based surveys . . .	12
1.1.4 Multiwavelength datasets	14
1.1.4.1 Uncovering galaxy properties with multiwavelength data	15
1.1.4.2 The benefits of multiwavelength data for estimating redshifts	16
1.1.5 Conclusion on surveys	17
1.2 Photometric redshifts	17
1.2.1 Historical overview	19
1.2.2 Methods	20
1.2.2.1 Empirical methods	20
1.2.2.2 Template methods	22
1.2.2.3 Comparison	24
1.2.3 Applications	25

1.2.3.1	Cosmology	25
1.2.3.2	Galaxy evolution	26
1.2.4	Conclusion on photometric redshifts	26
1.3	High-redshift galaxies	27
1.3.1	Selection methods	28
1.3.1.1	Narrowband searches for Lyman- α emitters	28
1.3.1.2	Lyman-break selection	29
1.3.1.3	Quiescent and dusty high-redshift galaxies	32
1.3.2	Lyman-break galaxies at high redshifts	33
1.3.2.1	Historical overview and state of the art	33
1.3.2.2	The importance of Lyman-break galaxies for high-redshift galaxy evolution studies	36
1.3.3	Conclusion on high-redshift galaxies	40
1.4	Thesis outline	41
2	Combined DES+VIDEO catalogue	43
2.1	Data	44
2.1.1	The Dark Energy Survey (DES)	44
2.1.1.1	Overview	44
2.1.1.2	Data collection	44
2.1.1.3	Data reduction	48
2.1.1.4	Catalogues	50
2.1.1.5	Data quality	51
2.1.2	The Vista Extragalactic Deep Observations Survey (VIDEO)	58
2.1.2.1	Overview	58
2.1.2.2	Data collection	59
2.1.2.3	Data reduction	61
2.1.2.4	Data quality	62
2.1.3	Spectroscopic redshifts	65
2.2	Forced photometry and catalogue production	67
2.2.1	Motivation	67
2.2.2	Method	70
2.2.2.1	Image resampling and preparation	70
2.2.2.2	Source extraction	71
2.2.3	Source extraction method validation	74

2.2.4	Correction of the VIDEO photometric uncertainties	76
2.2.4.1	Aperture flux errors	78
2.2.4.2	Auto flux errors	82
2.2.4.3	Verification of the corrected errors	84
2.2.5	Combined DES+VIDEO catalogue production pipeline .	84
2.2.5.1	Duplicates	86
2.2.5.2	Note on aperture corrections	87
2.3	Catalogue verification and results	88
2.3.1	Astrometry verification	88
2.3.2	Photometry verification	89
2.3.3	Forced photometry verification and results	90
2.3.4	Catalogue results	95
3	Photometric redshifts	97
3.1	Motivation	97
3.2	Description and implementation of the photometric redshift algorithm	99
3.2.1	Choice of code	99
3.2.1.1	χ^2 minimisation	100
3.2.1.2	Procedure for missing data	102
3.2.1.3	Reduced χ^2	103
3.2.2	$N(z)$ prior	103
3.2.3	Adaptive offsets	105
3.3	Templates	107
3.3.1	Galaxy templates	107
3.3.1.1	AVEROI_NEW templates	108
3.3.1.2	COSMOS templates	108
3.3.2	Stellar and QSO templates	110
3.3.3	Extinction	111
3.3.4	Emission lines	112
3.4	Photometric redshift configurations	114
3.4.1	The default configurations	115
3.4.1.1	Templates	115
3.4.1.2	Photometric redshift code options for the de-fault configurations	116

3.4.1.3	Input photometry for the default configurations	118
3.4.2	Alternative configurations	118
3.5	Configuration testing	120
3.5.1	Photometric redshift computation method	120
3.5.2	Accuracy metrics for assessing photo-z performance . .	122
3.5.3	Results	125
3.6	Discussion on configuration testing	132
3.6.1	Templates	132
3.6.1.1	Template sets	132
3.6.1.2	Extinction	134
3.6.1.3	Emission lines	136
3.6.2	Photo-z code options	137
3.6.2.1	Prior	137
3.6.2.2	Adaptive offsets	138
3.6.3	Input photometry	139
3.6.3.1	Auto vs aperture fluxes	139
3.6.3.2	Near-IR data	141
3.6.3.3	Template sets revisited	148
3.7	Full DES+VIDEO catalogue	149
3.7.1	Star-galaxy separation	149
3.7.1.1	Background	149
3.7.1.2	A χ^2_v -based method	150
3.7.1.3	Method validation	151
3.7.1.4	The effect of near-IR data	154
3.7.2	Total galaxy redshift distribution	155
4	High-redshift Candidates	157
4.1	Motivation	157
4.2	High-redshift galaxy candidate selection process	158
4.2.1	Brief overview	158
4.2.2	Template choice	158
4.2.3	Broad initial cuts	159
4.2.3.1	Weight flag cut	160

4.2.3.2	Redshift selection cut: definition of $z \sim 5$ and $z \sim 6$ samples	161
4.2.3.3	Star-galaxy separation	163
4.2.3.4	Galaxy fit	164
4.2.3.5	Magnitude cut	164
4.2.3.6	Conclusion of first round cuts	167
4.2.4	Visual inspection	167
4.2.4.1	Contamination from transient emission and imaging artefacts	168
4.2.4.2	Contamination from low-redshift galaxies and QSOs	172
4.2.4.3	Contamination from stars	175
4.2.5	Second round of cuts	177
4.2.5.1	Artefacts and transients	177
4.2.5.2	Low-redshift galaxies and QSOs	178
4.2.5.3	Stars	180
4.2.5.4	Manual removal	181
4.2.5.5	Summary and verification of the second selec- tion round	181
4.3	Final sample	184
4.3.1	Comparison with previous studies	191
4.3.1.1	Number of expected candidates	191
4.3.1.2	Comparison with Willott et al. (2013)	193
4.4	Attempted spectroscopic follow-up	194
5	Summary and conclusion	197
5.1	Summary of Chapter 2: combined DES+VIDEO catalogue	197
5.1.1	Suggestions for future work	199
5.2	Summary of Chapter 3: photometric redshifts	200
5.2.1	Suggestions for future work	202
5.3	Summary of Chapter 4: high-redshift galaxies	203
5.3.1	Suggestions for future work	204
Bibliography		205

Appendix A Configuration files	217
A.1 S Warp configuration file	217
A.2 S Extractor configuration file	220
A.3 L e P HARE configuration file	223
Appendix B False-colour image code	231

List of Figures

1.1	Relation between redshift and cosmic time	8
1.2	Selection of Lyman-break galaxies	31
1.3	Luminosity functions at $4 < z < 10$	38
2.1	DECam focal plane	45
2.2	DES and VIDEO filter transmission curves	46
2.3	Map of the DES footprint	47
2.4	Depths in the ELAIS-S1 field	54
2.5	Depths in the XMM-LSS field	55
2.6	Depths in the CDF-S field	55
2.7	Area as a function of 5σ depth for the total survey footprint . .	57
2.8	VIRCAM focal plane	59
2.9	Example weight map	64
2.10	z_{spec} distribution of available spectroscopic redshifts	67
2.11	Forced photometry	69
2.12	Verification of the photometry method: number of source detections	76
2.13	Verification of the photometry method: photometry	77
2.14	Segmentation image	79
2.15	Error estimation procedure	80
2.16	Verification of the astrometry	89
2.17	Verification of the photometry	91
2.18	Number counts in the K_s -band	93
2.19	Number counts in the i -band and z -band	95
3.1	Galaxy template SEDs	109

3.2	Dust extinction curves	113
3.3	z_{spec} vs z_{phot} distributions of the test configurations	127
3.3	<i>continued</i>	128
3.3	<i>continued</i>	129
3.3	<i>continued</i>	130
3.4	Accuracy metrics as a function of photometric redshift between $0 < z_{\text{phot}} < 6.5$	145
3.5	Accuracy metrics as a function of photometric redshift between $0 < z_{\text{phot}} < 1.5$	146
3.6	Star-galaxy separation verification	153
3.7	Total photometric redshift distribution	155
4.1	Area as a function of 7.5σ depth for the DES+VIDEO footprint .	165
4.2	Example of contamination from a transient	169
4.3	Example of contamination from an imaging artefact	170
4.4	Example of low-redshift contamination due to g -band and r -band noise	173
4.5	Example of a low-redshift contaminant	175
4.6	Example of a stellar contaminant	176
4.7	False-colour images of $z \sim 5$ candidates	189
4.8	False-colour images of $z \sim 6$ candidates	190

List of Tables

1.1	Small-scale space-based surveys	11
1.2	Intermediate-scale ground-based surveys	13
2.1	Wavelength coverage of broadband filters	46
2.2	Coordinates of the DES deep fields	48
2.3	Range of depths in the DES deep fields	56
2.4	Coordinates of the VIDEO tiles	60
2.5	Minimum and maximum depths in the VIDEO data	65
2.6	Spectra and source surveys	66
2.7	Aperture diameters	74
2.8	Comparison of photometric uncertainties	83
3.1	Photometric accuracy test results	126
4.1	Selection criteria for the first round of cuts	166
4.2	Selection criteria for the second round of cuts	182
4.3	Final $z \sim 5$ candidate sample	186
4.4	Final $z \sim 6$ candidate sample	188

Chapter 1

Introduction

1.1 An overview of astronomical surveys

Astronomy, the study of the universe and its contents, is one of the oldest natural sciences. Across its rich history, from the first naked eye observations of the solar system to probing the edge of the observable universe, astronomical research has progressed through a combination of two methods. The first approach is to study an individual object, such as a star, planet or galaxy in detail. But the number of objects in the night sky is too large, and their variety too great, to investigate them all in this manner. The second approach, therefore, is to survey the sky to collect large samples. These samples can be used as a way to find new objects for detailed follow-up, or for statistical studies of populations. Generally speaking, progress in astronomy has often been pushed forward by the development of new observational tools (Djorgovski et al. 2013). With each new invention, a new frontier for science opens up. When this happens, surveys are often the first step in the exploration of the newly accessible regimes.

1.1.1 A brief history of astronomical surveys

Following Wynn-Williams (2016), an insightful way to describe the history of astronomy is to focus on new technologies as a trigger for progress. With the discovery of each major observational toolkit, a new era of discovery opens up. Wynn-Williams (2016) divides astronomy into five distinct historical epochs:

The Naked Eye Era From BC to the invention of the telescope in the 17th century the sky was catalogued using only naked eye observations. This period saw the first record of what could be considered an astronomical survey: a star catalogue of 71 stars and constellations made by the Babylonians around the 8th century BC.

The Telescope Era Galileo's invention of the telescope in 1609 launched the beginning of a new form of astronomy. Telescopes allowed for much deeper observations, but objects still had to be recorded by hand. The Messier (1781) catalogue and star counts by the Herschel family in the late 18th century are famous examples of important works from this period.

The Photographic Era The third era started with the introduction of astrophotography around 1880. Using photographic plates, astronomers could make permanent records of their observations, and long exposures allowed them to probe much fainter objects. The Astrographic Catalogue and the Carte du Ciel were ambitious projects undertaken by a collaboration of 20 observatories between 1891 and 1950. Their purpose was to make a photographic map of the entire celestial sphere. Together with the Harvard Plate Collection (1885–1993), which repeatedly photographed the sky in search of variable stars, these could be considered the first instances of a survey in the modern sense of the word. Of all the photographic surveys throughout the epoch, the Palomar Observatory Sky Survey (POSS-I; 1949–1958; see Minkowski & Abell 1963), was perhaps the most impactful. This project, covering two thirds of the sky, was made possible by the invention of Schmidt telescopes, which were especially well suited to taking the wide field exposures required for efficient surveying.

The Electromagnetic Era Up until the 1930s, all astronomy was carried out using visible light (with small extensions into the ultraviolet and infrared using special photographic plates). From the 1930s onwards, special detectors opened up the radio, X-ray, ultraviolet (UV), and infrared (IR) wavebands. Of particular interest to this thesis is the development of infrared astronomy, which studies radiation roughly from $1 \mu\text{m}$ to 1mm . The range from $1 \mu\text{m}$ to $2.5 \mu\text{m}$ is

referred to¹ as the near-infrared², $2.5\text{ }\mu\text{m}$ to $30\text{ }\mu\text{m}$ as the mid-infrared, $30\text{ }\mu\text{m}$ to $450\text{ }\mu\text{m}$ as the far-infrared, and $450\text{ }\mu\text{m}$ to 1 mm as the submillimeter. The IR branch of astronomy took off in the 1960s with the development of sensitive solid-state IR detectors. The Two Micron Sky Survey (TMSS; Neugebauer & Leighton 1969) was the first all-sky IR survey, imaging the whole sky in the near-IR at $2.2\text{ }\mu\text{m}$ (i.e. the present-day K -band).

The CCD Era Several breakthroughs in digital technology and computing triggered the beginning of a fifth era of astronomy in the 1990s. Thanks to the enormous progress facilitated by these inventions, we are now firmly in a time where digital technology enables enormous surveys discovering sources by the billions. Several important developments in this period will be the primary focus of the current chapter.

1.1.2 Technological advancements: towards a data revolution

With every new invention in each epoch, the capacity for generating observational data increased. The scientific potential of surveys was, however, often held back by limited ability to process all this data. For example, optical catalogues in the photographic era had to be compiled by hand, which restricted their size to at most several tens of thousands of objects. It was not until the 1990s, when the technology for scanning and digitising images became available, that the full potential of photographic surveys could be unlocked (Lasker et al. 1990; Djorgovski et al. 1998). To give an example, after the POSS-II (the successor to the POSS; 1980s-1990s; Reid et al. 1991) had been digitised, the scanned images contained 3 TB of pixel data. These generated a catalogue of ~ 50 million galaxies and ~ 500 million stars to limiting magnitudes of ~ 22 mag (Djorgovski et al. 1998) – a vast increase over the several tens of thousands of galaxies that could be recorded in catalogues compiled by hand.

Infrared astronomy was similarly also initially held back by limited data recording and processing capacity. The aforementioned TMSS from the 1960s il-

¹It must be emphasised that these definitions can differ between publications, and that the conventions listed here are the ones used in this thesis.

²In physics the infrared region conventionally starts at $0.7\mu\text{m}$, just beyond the reddest light visible with the naked eye; but in astronomy light between $0.7\mu\text{m}$ – $1\mu\text{m}$ is generally imaged with the same instruments as visible light and is therefore often regarded as part of the optical waveband.

lustrates these challenges very well. Signals for the eight lead sulfite detectors that made up the TMSS camera were recorded on multi-channel medical strip chart recorders, and source detection and measurements were carried out by eye. Another example comes from other experiments in the 1970s and 1980s with mid- and far-IR telescopes on sub-orbital rockets, which recorded 25 MB of data during 4 minutes of operation. The fact that this data took 5 hours to process, illustrates the enormous computing progress that needed to be made to enable the digital revolution of the CCD era.

By the end of the 1980s, however, a digital all-sky IR survey started to become technologically possible. Crucial factors were the release from military classification of solid-state infrared detectors suitable for astronomical research, as well as the availability of cheap computing to process the huge resulting data streams. The first two major digital IR surveys were the Two Micron All Sky Survey (2MASS; 1996–2001; Skrutskie et al. 2006) and the Deep Near Infrared Survey of the Southern Sky (DENIS; 1997–2001; Epchtein et al. 1997). They covered the entire sky and the southern hemisphere respectively, both using cameras made up of an array of 256×256 detectors (i.e. pixels). Every one of these 56 536 detectors was more sensitive than each one of the eight TMSS sensors, resulting in an 11 mag increase in depth over the TMSS (final survey depths were $K \sim 14.3$ for 2MASS and $K \sim 14.0$ for DENIS). The increased sensitivity also facilitated a vast increase in the number of source detections: 470 000 000 and 355 000 000 for 2MASS and DENIS respectively vs. 5600 for the TMSS.

Meanwhile, optical astronomy surveys continued to rely largely on photographic plates. Nevertheless, by the early 1990s, the technology for a large digital project was finally beginning to converge in the optical regime as well, thanks to the invention of large-scale digital cameras with charged coupled device (CCD) chips, together with the development of cheap computing capabilities. In contrast to photographic surveys, which generally used several photographic plates sensitive to various colours, CCD cameras can image different colours in an automated fashion. This is achieved by applying a series of broadband filters (also called bands) that only transmit light within certain wavelength limits (this technique had already been employed in several digital IR surveys in the 1990s as well). The first major project to capitalise on all this new digital optical technology was the Sloan Digital Sky Survey (SDSS; 2000-ongoing; York et al. 2000), which has since covered just over a third of the entire sky to depths of $r \sim 22.2$.

Its 120 megapixel camera consists of an array of 30 CCDs of 2048×2048 pixels each, which has thus far been used to assemble an enormous catalogue of nearly one billion unique objects. As of 2019, the total data volume of the SDSS is over 170 TB (Aguado et al. 2019).

Following on from these early successes, the CCD era of astronomy blossomed. With ever bigger and more advanced projects, astronomy is now well into the petascale data stream regime. For example, the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; 2010-ongoing; Kaiser et al. 2002, Kaiser et al. 2010) currently produces several TB of data every night, and its most recent 1.6 PB data release from January 2019 is the largest in astronomical history. Using a 1.4 gigapixel camera, Pan-STARRS aims to image three quarters of the sky to $r \sim 23.2$ mag. A second example of a large modern optical survey is the Dark Energy Survey (DES; 2012–2019; Flaugher 2005). This project currently constitutes the deepest large-scale survey in the southern hemisphere, having observed 5000 deg^2 to depths of $r \sim 24.3$ with its 570 megapixel camera. After data processing will have been completed, its final total volume is expected to be around 2 PB (Mohr et al. 2012).

In the near-infrared, current state-of-the-art instruments include the United Kingdom Infrared Telescope (UKIRT) and the Visible and Infrared Survey Telescope for Astronomy (VISTA). With 16 megapixel and 67 megapixel cameras respectively, both telescopes have carried out several large-scale surveys, ranging from footprints of approximately 4000 deg^2 to a full hemisphere (Lawrence et al. 2007; McMahon et al. 2013; Dye et al. 2018). The depths of these surveys reach up to $K \sim 18.2$ mag Vega (UKIDSS on UKIRT) and $K_s \sim 18.1$ mag Vega (VHS on VISTA). Such substantial increases in depths and camera sizes over 2MASS and DENIS has allowed these newest surveys to collect far larger numbers of objects. For instance, the wide galactic and extragalactic components of UKIDSS found 84 million galaxies and nearly a billion stars, an enormous gain compared to the $\lesssim 500$ million sources for 2MASS and DENIS.

The data explosion of the last 30 years presently shows no signs of slowing down. Further telescopes are currently planned or under construction, the most prominent of which is probably the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008). When completed, its 3.2 gigapixel camera will be the largest in the world, and it will aim to cover half the sky to a limiting magnitude of $r \sim 27.7$ by taking many exposures in short succession. Every night LSST is expected to

generate around 20 TB of data, which means that it would take just nine nights to surpass the entire SDSS data output. At the end of the project, the final volume of raw images will amount to 60 PB, and the corresponding 15 PB source catalogue is expected to contain 10 billion galaxies and a similar number of stars.

It is clear that the digital revolution of the CCD era has changed astronomy from a relatively data-poor to an intensely data-rich discipline (Djorgovski et al. 2013). This transformation brings many new opportunities for research on ever larger and deeper datasets. The science generated by all these surveys so far has been very broad, and has covered cosmological research such as measurements of large-scale structure, dark matter, and dark energy; as well as galaxy evolution studies. On a smaller scale, research has included investigations of the oldest stars, brown dwarfs, the structure of the Milky Way, and dwarf galaxies.

1.1.3 Deep-field surveys

So far we have seen the amazing technological progress that has revolutionised astronomy by generating very large wide-area surveys. But sometimes, one can also discover a great deal from observing a smaller patch of the sky in detail. The main advantage of such narrower surveys is that they can dedicate much more exposure time to a specific region, which means that observations can go several magnitudes deeper. In some form, small but deep surveys have always existed. Indeed, nearly every telescope throughout history has at some point been pushed to its observational limit to see what lies at the edge. But the field of small-scale surveys truly took off with the space-based observations of the *Hubble Space Telescope (HST)*. Its great success prompted a vast range of surveys from space, and later also from the ground. These projects, sometimes referred to as ‘deep fields’, are almost always aimed at galaxy formation and evolution studies. Their observed area is too small for studies of large-scale cosmology, stars, and galactic structure, but they are ideal for finding distant galaxies, as deep imaging is needed to find these faint sources.

1.1.3.1 Brief interlude on distances, redshifts and cosmic time

It is interesting to pause and reflect briefly on the scientific value of distant galaxies. Such objects are of primary interest to the study of galaxy evolution, because they can provide a window into the cosmic past. This is due to the finite speed

of light, which causes the light from any astronomical source to take a non-zero amount of time to reach an observer. The observer therefore sees the object as it was in the past³. This concept implies a result that is deeply entrenched in modern astronomy, yet in essence highly profound: *in studying galaxies with increasing distance one can probe increasingly far into the cosmic past*. Because this connection is of crucial importance to this thesis, the current section presents a number of important related principles and quantities.

First of all, *redshifts* provide a means of measuring distances to faraway galaxies. Due to the expansion of the universe, objects that are moving away from an observer solely due to this expansion (i.e. moving with the so-called Hubble flow) recede with increasing relative velocity the more distant they are (Hubble 1929). The light emitted by such an object is Doppler-shifted, and this shift can be captured by a dimensionless quantity called redshift. The redshift z of such a source is defined as:

$$z \equiv \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}} = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} - 1, \quad (1.1)$$

where λ_{em} is the wavelength emitted in the rest frame of the moving object, and λ_{obs} is the wavelength in the rest frame of the observer.

Via the standard model of cosmology, the redshift of an object resulting from cosmic expansion can be linked to the lookback time t_L , which is the difference between the age of the universe at the point of observation (i.e. now), and the age of the universe when the object emitted the observed photons (Hogg 1999 following Peebles 1993) :

$$t_L(z) = \frac{1}{H_0} \int_0^z \frac{dz'}{(1+z')\sqrt{\Omega_R(1+z')^4 + \Omega_M(1+z')^3 + \Omega_k(1+z')^2 + \Omega_\Lambda}}. \quad (1.2)$$

Here Ω_R is the present-day radiation density, Ω_M is the present-day total matter density, Ω_Λ is the present-day dark energy density, $\Omega_k = 1 - \Omega_M - \Omega_\Lambda$ is the curvature, and H_0 is the Hubble constant. Throughout this thesis, all calculations

³A note from the author: I believe that this point ought to inspire a deep appreciation in anyone with a love for astronomy. The fact that looking further equates to looking earlier in time, is one of the most scientifically useful outcomes of the laws of physics. And yet it is not a necessary truth that the universe is this way; it is a lucky coincidence of sorts, that the universe we live in happens to be one with a finite light speed. – A. L. S.

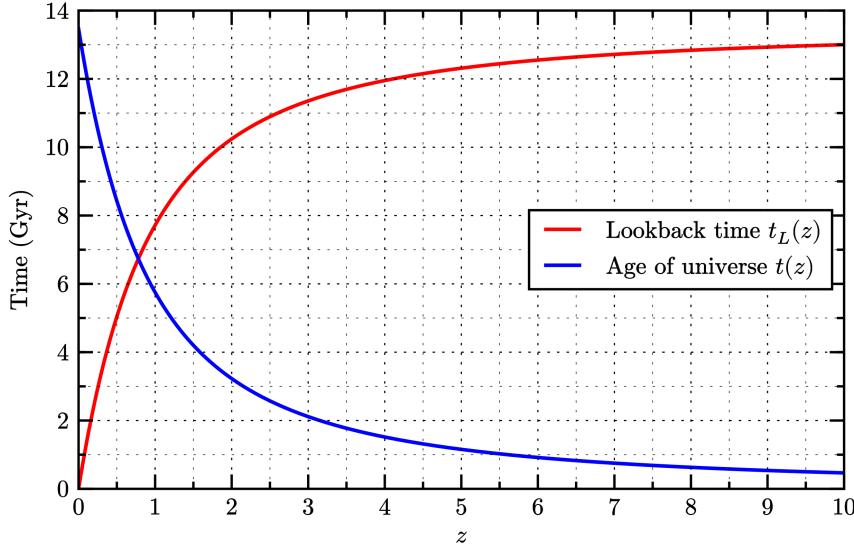


Figure 1.1 The lookback time $t_L(z)$ and age of the universe $t(z)$ for an object observed at a redshift z , provided that its redshift is purely caused by cosmic expansion.

will assume⁴ a cosmology of $\Omega_M = 0.3$, $\Omega_\Lambda = 0.70$, $\Omega_R = 0$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. In addition to the lookback time, the age of the universe t when the light from an object at redshift z was emitted can be found by changing the integration boundaries in Equation 1.2 (Frieman et al. 2008):

$$t(z) = \frac{1}{H_0} \int_z^\infty \frac{dz'}{(1+z')\sqrt{\Omega_R(1+z')^4 + \Omega_M(1+z')^3 + \Omega_k(1+z')^2 + \Omega_\Lambda}}. \quad (1.3)$$

To visualise the relation between cosmic time and redshift, the lookback time and age of the universe⁵ are both plotted as a function of redshift in Figure 1.1.

⁴Even though the assumption of $\Omega_R = 0$ is not strictly true, it is an excellent approximation in the matter-dominated era with which this thesis is concerned, as measurements have shown that $\Omega_R \simeq 0.8 \times 10^{-4}$ (Frieman et al. 2008).

⁵A note from the author: it is interesting to note that the lookback time increases incrementally slowly with increasing redshift. The fact that the laws of physics cause this behaviour is highly serendipitous; it results in an extremely useful set of circumstances in which the history of the universe can be studied with increasing resolution further back in time. In this way, the redshifts of the very earliest galaxies provide a very precise handle to study the build-up of cosmic structure.
— A. L. S.

Measuring redshifts

The equations above illustrate why redshifts are so important to galaxy evolution studies: if it is possible to obtain redshift measurements, galaxies can be sorted by cosmic epoch, so that changing properties can be studied statistically as a function of time. As implied by Equation 1.1, redshifts can be found by comparing the wavelengths of certain features in an observed, i.e. redshifted, galaxy spectral energy distribution (SED) to their corresponding expected rest-frame wavelengths. The most precise way to do this involves spectroscopy, comparing the observed wavelengths of certain emission or absorption lines in a high-resolution spectrum to the wavelengths of these transitions as measured in a laboratory. Some astronomical surveys, such as the SDSS, incorporate this technique via a spectroscopic component (Bolton et al. 2012). There have also been other projects that are exclusively spectroscopic, targeting galaxy candidates that have been selected from other imaging surveys (e.g. Colless et al. 2001; Weiner et al. 2005; Le Fèvre et al. 2005).

However, as will be discussed in more detail in Section 1.2, obtaining spectra is time-consuming, and unfeasible for the huge catalogues produced by modern surveys. Nevertheless, even without a high-resolution spectrum that exhibits individual absorption or emission lines, it is often still possible to use colours from broadband filters to obtain an approximate redshift estimate. Examples of this method include photometric redshifts (see Section 1.2) and colour cuts (exclusively for objects at $z \gtrsim 3$; see Section 1.3.1.2). In this way, the redshifts of galaxies in the enormous surveys of the CCD era can be estimated using only the survey data itself.

1.1.3.2 Small area space-based surveys

With the power of distance measurements provided by redshift estimates, deep-field surveys can be an extremely powerful tool for galaxy evolution studies. The current section briefly reviews the developments in this fast-moving field of research.

The prototype for the modern concept of extremely deep fields is the Hubble Deep Field (HDF; Williams et al. 1996), imaged by the *HST*. The space-borne nature of this instrument makes it especially well suited to extremely deep imaging. Because it avoids blurring by the Earth’s atmosphere, images can be more

SMALL-SCALE SPACE-BASED OPTICAL AND NEAR-INFRARED SURVEYS

Survey	Type	Epoch	Area	Bands, depth	mag AB	Nº galaxies
GSS ⁱ	Visible	1994	127 arcmin ²	V_{606} I_{814} 26.1 25.4	5 σ	11 000
HDF ⁱⁱ	Visible	1995	6.25 arcmin ²	U_{300} B_{450} V_{606} I_{814} 27.0 27.9 28.2 27.6	10 σ	3000
HDF-S ⁱⁱⁱ	Visible	1998	6.25 arcmin ²	U_{300} B_{450} V_{606} I_{814} 26.8 27.7 28.3 27.7	10 σ	2700
GOODS ^{iv}	Visible ^b	2002–2003	320 arcmin ²	B_{435} V_{606} i_{775} z_{850} 27.8 27.8 27.1 26.6	10 σ	73 000
HUDF ^v	Visible	2003–2004	11 arcmin ²	B_{435} V_{606} i_{775} z_{850} 29.1 29.5 29.4 28.7	10 σ	10 000
NICMOS UDF ^{vi}	NIR	2003	5.8 arcmin ²	J_{110} H_{160} 27.7 27.7	5 σ	1300
HXDF ^{vii,a}	Vis-NIR	2002–2013	11 arcmin ²	B_{435} V_{606} i_{775} I_{814} z_{850} 29.8 30.3 30.3 29.1 29.4 Y_{105} J_{125} JH_{140} H_{160} 30.1 29.8 29.8 29.8	5 σ	14 000
COSMOS ^{viii}	Visible ^c	2003–2005	1.8 deg ²	g_{475} I_{814} 27.9 27.5	5 σ	1 700 000
	Visible ^c		1.07 deg ²	U_{300} B_{450} 24.8 26.7		
	NIR ^c		330 arcmin ²	H_{160} 25.9		21 000
AEGIS ^{ix}	Visible ^c	2004–2005	700 arcmin ²	V_{606} I_{814} 28.1 27.5	5 σ	80 000
	NIR ^c		46 arcmin ²	J_{110} H_{160} 25.7 25.5	10 σ	4000
CANDELS ^x	Vis-NIR	2010–2013			5 σ	250 000
<i>Wide</i>			770 arcmin ²	V_{606} I_{814} 27.6 27.5 Y_{105} J_{125} H_{160} 26.6 26.4 26.5		
<i>Deep</i>			120 arcmin ²	B_{435} V_{606} i_{775} I_{814} z_{850} 27.8 28.2 27.5 28.4 27.3 Y_{105} J_{125} H_{160} 27.3 27.4 27.2		

continued

 SMALL-SCALE SPACE-BASED OPTICAL AND NEAR-INFRARED SURVEYS

References:

- ⁱ Groth et al. (1994) ⁱⁱ Williams et al. (1996) ⁱⁱⁱ Casertano et al. (2000)
^{iv} Giavalisco et al. (2004a) ^v Beckwith et al. (2006) ^{vi} Thompson et al. (2005)
^{vii} Illingworth et al. (2013) ^{viii} Scoville et al. (2007) ^{ix} Davis et al. (2007)
^x Grogan et al. 2011; Koekemoer et al. 2011

Other comments:

- ^a The HXDF is assembled out of all Hubble data in the HUDF region.
Almost all fields in this table have been studied in several wavelengths since being released, but the following projects contained further multiwavelength data as part of their core science goals:
^b Also contains X-ray, mid-IR, far-IR data; and ground-based optical, near-IR data.
^c Also contains X-ray, UV, mid-IR, far-IR data; and ground-based optical, near-IR, radio data.
-

Table 1.1 Overview of major small-scale space-based surveys over optical ('visible'/'vis) and near-infrared ('NIR') wavelengths.

detailed and deeper, as light is spread out over fewer pixels. The HDF contains 3000 galaxies with magnitudes of up to $V \sim 30$ mag for the faintest sources, and several objects with redshifts of $z > 5$. The Hubble Ultra Deep Field (HUDF; Beckwith et al. 2006) and the Hubble Extreme Deep Field (HXDF; Illingworth et al. 2013) imaged the same region of the sky as the original HDF, over larger areas and with newer cameras. These surveys achieved even higher depths and pushed the discovery of high-redshift galaxies to $z \sim 7$ and $z \sim 12$ respectively, reaching the very limit of what is possible with current technology. Table 1.1 presents a summary of some core properties of the Hubble Deep Field surveys.

After the success of the Hubble Deep Fields, further deep *HST* surveys followed. Progress was mainly made in the direction of larger areas and more wavelength coverage. This helped to build up a more comprehensive picture of galaxy evolution, as the larger size of the observed fields allowed the assembly of larger, statistically significant samples to study changes in galaxy populations. Table 1.1 summarises several noteworthy examples of these surveys. Each one specialised in a specific redshift regime, but together they cover cosmic time between roughly $0.5 \lesssim z \lesssim 8$. Examples of galaxy properties that have been studied with these datasets include mass assembly, the coupled evolution of galaxies, active galactic nuclei (AGN), and large-scale structure. The bigger sample sizes also facilitated applications in cosmology, such as studies of dark matter through gravitational lensing.

1.1.3.3 Intermediate-area ground-based surveys

A major downside to space-based surveys is that the main optical and near-IR instrument is the *HST*, which has a very restricted amount of available observing time and a limited field of view compared to ground-based telescopes. On top of that, ground-based telescopes are more cost-effective for larger surveys, especially for areas over $\sim 1 \text{ deg}^2$. To capitalise on these advantages, astronomers have started to conduct intermediate-area deep surveys from the ground, often utilising the sophisticated terrestrial instruments developed during the CCD revolution (described in Section 1.1.2). Such intermediate surveys cover several square degrees, to much higher depths than the large-scale wide-area projects from Section 1.1.2. In this way, they can cover much more area than space-based ones, although they are generally not quite as deep and have much lower image resolution due to atmospheric blurring.

Their position between the shallow wide-area ground-based surveys and the ultra-deep narrow space-based surveys makes intermediate-scale surveys ideal for intermediate applications. As will be illustrated in more detail below, they are often used to collect larger samples than possible with space-based surveys, in order to improve sample statistics. Furthermore, they are more likely to contain rare objects that require a larger search area. Intermediate-scale surveys can also unveil cosmological structures larger than the few Mpc scales probed by space-based surveys, and can reduce biases introduced by cosmic variance. On top of that, they offer a significant increase in lookback time over wide-area surveys. Their major downside is that they are not able to reach quite as high redshifts as the space-based datasets, so they generally cannot be used to find the utmost distant sources.

Table 1.2 lists several important examples of intermediate-scale surveys. For many of these, a larger observational footprint has offered significant scientific advantages. A common research goal has been to collect large samples for improving population statistics. For instance, the Subaru Deep Field (SDF) survey revealed statistically significant samples of Lyman-break galaxies at $z \approx 4\text{--}5$ and Lyman- α emitters at $z \approx 5\text{--}6$. Similarly, the UKIDSS Deep Extragalactic Survey (UKIDSS-DXS) aimed to facilitate a statistical study of galaxy properties at $z \sim 1$, to compare with the $z \sim 0$ picture from local surveys such as SDSS. Furthermore, UltraVISTA has enabled a representative study of the peak of star formation at $2 < z < 3$, and the VISTA Deep Extragalactic Observations (VIDEO) survey has

INTERMEDIATE-SCALE GROUND-BASED OPTICAL AND NEAR-INFRARED SURVEYS

Survey	Type	Epoch	Area	Bands, depth	mag AB	Nº galaxies
NDWFS ⁱ	Vis-NIR	1997–1999	9 deg ²	B_w R I K 27.2 26.3 25.9 21.0	5σ	3 200 000 ^a
DLS ⁱⁱ	Visible	2001–2006	20 deg ²	B V R z' 26.6 26.5 26.9 24.8	5σ	5 000 000
SDF ⁱⁱⁱ	Visible	2002–2004	920 arcmin ²	B V R i' z' 28.5 27.7 27.8 27.4 26.6	3σ	210 000
SDXF ^{iv}	Visible	2002–2005	1.2 deg ²	B V R_c i' z' 28.4 27.8 27.7 27.7 26.6	3σ	900 000 ^a
CFHTLS-D ^v	Visible	2003–2009	4 deg ²	$u*$ g' r' i' z' 27.5 27.9 27.9 27.7 26.5	5σ	> 300 000
UKIDSS-UDS ^{vi}	NIR	2005–2012	0.77 deg ²	J H K 26.3 26.0 25.6	5σ	250 000
UKIDSS-DXS ^{vi}	NIR	2005–2012	35 deg ²	J K 23.4 22.9	5σ	3 000 000 ^a
			5 deg ²	H 23.4		
VIDEO ^{vii}	NIR	2009–2018	12 deg ²	Z Y J H K_s 25.7 24.5 24.4 24.1 23.8	5σ	1 700 000 ^a
UltraVISTA ^{viii}	NIR	2009–2018			5σ	200 000 ^a
<i>Deep</i>			1.5 deg ²	Y J H K_s 24.8 24.5 24.1 23.8		
<i>Ultra-Deep</i>			0.73 deg ²	Y J H K_s 25.7 25.4 25.1 24.1		
DES Deep ^{ix}	Visible	2013–2019			5σ	2 000 000
<i>Shallow</i>			24 deg ²	g r i z 27.7 26.5 26.8 26.6		
<i>Deep</i>			6 deg ²	g r i z 28.0 28.2 27.9 27.7		

References:

- ⁱ Jannuzzi & Dey (1999) ⁱⁱ Wittman et al. (2002) ⁱⁱⁱ Kashikawa et al. (2004) ^{iv} Furusawa et al. (2008) ^v Gwyn (2012) ^{vi} Lawrence et al. (2007) ^{vii} Jarvis et al. (2013) ^{viii} McCracken et al. (2012) ^{ix} Bernstein et al. (2012); Dark Energy Survey Collaboration et al. (2016)

Other comments:

^a Because galaxy counts were not provided, these are total source counts.

^b These are the expected final depths for point sources from the full five years of observations. They differ from the depths for the DES data presented later in this thesis (Table 2.3 and Figure 2.7), which are for a 1.95 arcsec aperture using only the first year of observations.

Table 1.2 Overview of major intermediate-scale ground-based surveys over optical ('visible'/'vis') and near-infrared ('NIR') wavelengths.

been targeted at providing a comprehensive view of the evolution of galaxies with their environment out to $z \approx 4$. A second commonly occurring goal of the surveys in question has often involved finding uncommon objects. Rare object searches have included AGN searches in the VIDEO survey, the deep component of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS-D), and the NOAO Deep Wide Field Survey (NDWFS). Additionally, the most luminous galaxies can be studied at $3 < z < 5$ in the CFHTLS-D, at $z > 4$ in the NDWFS, and out to $z \approx 7$ in UltraVISTA and the UKIDSS Ultra-Deep Survey (UKIDSS-UDS). Ultravista was also designed to uncover galaxies at the highest redshifts ($z \approx 6.5--9$). A third aim of some ground-based surveys, such as the Dark Energy Survey Deep Fields and CFHTLS-D, has been to look for Type Ia supernovae to constrain dark energy. Lastly, the fourth major application of intermediate-scale surveys concerns studies of large-scale structure for the purpose of constraining cosmological parameters. Examples include the Subaru Deep XMM-Newton Survey (SDXF), NDWFS, and the Deep Lens Survey (DLS), the latter of which specifically focused on weak gravitational lensing.

1.1.4 Multiwavelength datasets

Many of the astronomical surveys that have been discussed often cover a particular wavelength regime, generally optical or near-infrared. Their scientific successes show that each such dataset can be incredibly powerful on its own. However, in many cases, combining observations from surveys over multiple wavebands can open up further opportunities for interesting discoveries.

Historically, the combination of surveys over a wide range of the electromagnetic spectrum has been very fruitful. Perhaps the oldest and simplest application of such a multiwavelength approach is the use of optical data for the identification of sources that were discovered in other frequencies. For instance, in the early ages of infrared and radio astronomy, the physicists that pioneered these fields often turned to the optical POSS-I survey (see Section 1.1.1) for this purpose. Other historical examples where combining datasets from different wavelengths has revealed results that could not be recognised in a single dataset alone include the discovery of quasars (QSOs) and ultraluminous starbursts, and the interpretation of γ -ray bursts (see Djorgovski et al. 2013).

To capitalise on the opportunities offered by multiwavelength astronomy, contemporary observations are often coordinated between instruments special-

ising in different wavelength regimes. In particular, several of the small and intermediate area surveys from Tables 1.1 and 1.2 have been designed as part of a collaboration of different instruments, and some others have later been followed up at other wavelengths. These developments are moving astronomy into a regime where the energetic output of a source can be traced throughout large parts of the electromagnetic spectrum. The study of galaxy evolution is one research area that particularly benefits from the great scientific power of panchromatic datasets, and the following two sections will present two examples.

1.1.4.1 Uncovering galaxy properties with multiwavelength data

The first asset of a multiwavelength approach is that different frequencies trace different galaxy properties. To name a few examples, optical imaging unveils their positions and morphologies, as well as possible environmental disruptions such as mergers with neighbouring objects (Darg et al. 2010; Willett et al. 2017). Near-infrared data can be valuable as well, for instance for measuring stellar masses (Bell et al. 2003). Furthermore, X-ray, mid-infrared, and radio observations can detect AGN, including those obscured by dust (Hickox & Alexander 2018). X-ray data can also be used to find hot intergalactic gas around galaxy mergers and clusters (Cox et al. 2006; Burchett et al. 2018). Finally, information from the ultraviolet, mid-IR, and radio wavebands can track the rate of active star formation (Brown et al. 2017).

The wavelength regimes listed above generally apply to the rest-frame emission. When objects are redshifted, the wavebands required to trace certain galaxy properties will become redder. For instance, as will be discussed in detail in Section 1.3.1.2, galaxies generally show a sharp break at (rest-frame) wavelengths of $\lambda = 912\text{--}1216\text{ \AA}$, emitting very little flux below this cutoff. At $z > 6$, this cutoff is redshifted to $\lambda > 8500\text{ \AA}$, so sources at these distances will emit almost all their flux in the observed near-infrared and beyond. Therefore, imaging at $\lambda > 8500\text{ \AA}$ will be needed to observe them at all. Furthermore, properties which are locally measured at UV wavelengths, such as the star formation rate, are traced by observed-frame near-IR imaging at these high redshifts.

1.1.4.2 The benefits of multiwavelength data for estimating redshifts

The second core strength of multiwavelength datasets involves improved redshift estimates from broadband colours. Such measurements generally rely on several strong spectral features in galaxy rest-frame UV and optical emission (most importantly the Lyman limit at $\lambda = 912 \text{ \AA}$, the $\lambda = 1216 \text{ \AA}$ break, the Balmer break at $\lambda = 3646 \text{ \AA}$, and the $\lambda = 4000 \text{ \AA}$ break). However, with increasing redshift, these features will trivially shift towards longer wavelengths, so that redder imaging becomes more and more important. For instance, at $z > 1.5$ the 4000 \AA break has redshifted out of the optical and into the near-IR. Secondly, as mentioned earlier, beyond $z > 6$ the observed emission from galaxies — and hence all their important spectral features — lies almost entirely in the near-IR and beyond. Therefore, near-infrared imaging is almost always necessary to determine redshifts for $z > 6$ galaxies (McLure et al. 2009; Bowler et al. 2014). This point will be revisited later in Section 1.3.2.1.

A particularly common method for estimating redshifts in modern astronomy is the so-called photometric redshift technique. An in-depth discussion of this method will follow shortly in Section 1.2, where it will be shown that this is a highly effective technique for obtaining redshift estimates from broadband fluxes in large surveys. The photometric redshift approach essentially computes the redshift of a given object by using its broadband colours as a very low resolution spectrum. Simple statistics *prima facie* dictate that adding data in more filters will improve the resulting redshift estimates, simply by increasing the amount of available information. The most striking example of this is perhaps the photometric redshift catalogue for the COSMOS field (Ilbert et al. 2009). With 30 broadband colours covering the UV, optical, near-IR, and mid-IR, the accuracy of this panchromatic dataset approaches that of spectroscopic redshifts (when compared with spectroscopic redshifts $\Delta z = z_{\text{spec}} - z_{\text{phot}}$, the dispersion is $\sigma_{\Delta z}/(1+z_{\text{spec}}) = 0.007$ for sources brighter than $i_{\text{AB}}^+ < 22.5$). Furthermore, several simulations (Banerji et al. 2008; Abdalla et al. 2008) and observational studies (Ilbert et al. 2009; Jarvis et al. 2013; Banerji et al. 2015) have indicated that photometric redshift estimates benefit from including infrared data in particular. Altogether, it is therefore clear that combining surveys from multiple wavebands is beneficial to galaxy evolution studies (as well as many areas of astronomy in general).

1.1.5 Conclusion on surveys

We have now reached the end of our birds-eye-view journey through survey astronomy, which forms the first part of the introduction to this thesis. Along the way, we have seen how advances in technology and instrumentation have given rise to a situation where celestial sources can now be discovered electronically millions or billions at a time. Thanks to this data-driven revolution, the last 30 years can undoubtedly be characterised as the CCD era of astronomy. This period has witnessed large-scale wide surveys imaging substantial parts of the sky, as well as smaller deeper surveys unveiling increasingly distant cosmic structures. We have also seen how combining observations over multiple wavelength regimes can often increase the scientific potential of surveys even further. The enormous datasets resulting from all of these developments have led to many new discoveries across several branches of astronomy, and there are doubtlessly many more to come.

But the explosion of data also comes with certain obstacles. Some of these, such as the infrastructure needed to process the gigantic data streams, lie within the realm of technology and computer science. Other challenges are computational and astrophysical in nature, and concern the extraction of physically meaningful quantities from survey data. It is aspects from the latter category that the current text will address.

As part of this, the first part of this thesis will focus on the task of combining optical and near-infrared data from two deep intermediate-area ground-based surveys to create a single, more powerful multiwavelength catalogue. This new merged data product can harness the potential of panchromatic datasets, creating a useful resource for the rest of this thesis as well as for the scientific community. The next part of the thesis will then describe the process of acquiring distance estimates for all objects in the catalogue. This is achieved through photometric redshifts, and this computational method forms the topic of the next part of this introduction.

1.2 Photometric redshifts

Many branches of research in modern extragalactic astronomy rely crucially on distance measurements. As mentioned previously in Section 1.1.3.1, spectroscopic redshifts generally offer the most accurate way of doing this. However, for

a multitude of reasons that will be discussed imminently, it is not always feasible to measure redshifts spectroscopically. To circumvent this problem, photometric redshifts (photo-zs) can be measured from broadband magnitudes only. The idea is simple: multiband photometry is essentially a very low resolution spectrum, and can therefore be used to estimate redshifts (Yee 1998). Primarily, this tactic relies on broad spectral features that exist in the emission of most galaxies (predominantly the Lyman break at $\lambda = 912\text{--}1216 \text{ \AA}$ and the Balmer/4000 \AA break). The photometric redshift approach has two major strengths:

1. Photometric redshifts can be computed for a far larger number of objects than is feasible with spectroscopy. The catalogues of modern surveys in the CCD era can contain millions or billions of galaxies. Despite the efficiency of modern multi-object spectrographs, it is only possible to measure meaningful spectra for a few percent of these sources, as spectroscopy is relatively time-consuming (Le Fèvre et al. 2005; Salvato et al. 2019). On the other hand, broadband magnitudes can be extracted directly from the imaging, which is obtained for all objects in the field of view simultaneously. The remaining problem of estimating the photo-zs is computational in nature, but the time required for this is minimal compared to the duration of spectroscopy. The latter issue is especially pressing for high-redshift research, as even the most luminous galaxies at redshifts $z \sim 6$ are so faint that they require integration times of several hours (e.g. Willott et al. 2013). This makes spectroscopy an unfeasible method to search for such distant sources, which are extremely rare compared to the abundance of low-redshift galaxies.
2. The photometric redshift method can be used on sources for which conventional spectroscopy is not possible. As it relies only on broadband magnitudes, the technique can be applied to all galaxies in a given source catalogue, including objects with very low apparent brightness all the way out to a survey's imaging limits (Yee 1998; Salvato et al. 2019). These faint sources that lie outside the current realm of spectroscopy include very small galaxies or galaxies at extremely high redshifts.

These two advantages have made the use of photometric redshifts ubiquitous for large datasets, and the method has become an integral part of data processing in galaxy surveys (Dahlen et al. 2013). The following sections will give a historical

overview of the development of photo-zs and discuss specific techniques available in the literature.

1.2.1 Historical overview

Baum (1962) was the first ever author to use broadband photometry for estimating galaxy redshifts. He computed photo-zs by comparing magnitudes in nine photometric filters to spectra from galaxies in the Virgo cluster. After that, the next occurrence of the photometric method was not until the 1980s, when it was once again employed by several authors. Butchins (1981, 1983), Puschell et al. (1982), and Koo (1985) all estimated photometric redshifts by comparing colours from photographic plates to those derived from template galaxy SEDs. Midway through the decade, Loh & Spillar (1986) first pioneered the use of CCD photometry. After these initial developments, the photometric redshift method remained relatively unused for the next 10 years, but from the middle of the 1990s it experienced a resurgence. The main reasons for this revival were twofold:

- As the data revolution of the CCD era took off in the 1990s, broadband photometry could suddenly be obtained efficiently for huge samples of galaxies. The object counts of the resulting source catalogues were far too high to perform spectroscopy for every object, so it became necessary to use a more efficient method for distance measurements. In order to meet this need, many authors turned to photometric redshifts.
- The publication of the HDF (see Section 1.1.3.2) provided a dataset containing galaxies with magnitudes out to $V \sim 30$ mag AB. Such sources are far too faint for conventional spectroscopy, necessitating the use of photometric redshifts.

With these factors as impetuses for further development and refinement of the photo-z approach, many papers followed soon after (e.g. Connolly et al. 1995; Brunner et al. 1997; Wang et al. 1998; Lanzetta et al. 1996; Gwyn & Hartwick 1996; Pello et al. 1996). As the number of existing surveys and their source counts continued to rise while the CCD era advanced, the use of photometric redshifts grew even more widespread. More codes became available in the literature, and the available methods became more sophisticated. These developments will shortly be detailed further in Section 1.2.2, which will describe in

detail the variety of codes that have been created over the last three decades. As the use of photometric redshifts became standard practice, many scientists started to make their photo-z methods publicly available. The first such public code was **HyperZ** (Bolzonella et al. 2000). Many others soon followed, including **BPZ** (Benítez 2000), **ANNz** (Collister & Lahav 2004), **ZEBRA** (Feldmann et al. 2006), **LePHARE** (Ilbert et al. 2006; Arnouts & Ilbert 2011), **EAZY** (Brammer et al. 2008), **ArborZ** (Gerdes et al. 2010), **TPZ** (Carrasco Kind & Brunner 2013; Carliles et al. 2010) and **GPz** (Almosallam et al. 2016).

1.2.2 Methods

Walcher et al. (2011) and Yee (1998) identify two main photometric redshift strategies commonly used in the literature. *Empirical methods* employ a subset of objects with measured spectroscopic redshifts as a training set. This subset is used to determine an empirical relation between (spectroscopic) redshift and galaxy properties such as colour and magnitude. Photometric redshifts are then inferred by applying this relation to the entire dataset. *Template methods* use pre-existing galaxy spectral energy distributions — also known as templates — from models or libraries of observed spectra. These SEDs can be stretched to any redshift, and predicted magnitudes can be computed by convolving the shifted SEDs with the transmission curves used in the relevant survey. For each object, all the predicted magnitudes are then compared to the observed photometry to find the best-fitting redshift and template combination. The following two sections will describe both empirical and template methods in more detail.

1.2.2.1 Empirical methods

As mentioned above, empirical methods rely on a training sample to calibrate a relation $z = z(C, m_0)$ between redshift z and observed colours C or magnitudes m_0 (Benítez 2000). The phrase empirical refers to the fact that this relation is derived empirically from data in the training sample only.

The first empirical methods were developed when photometric redshifts rose to prominence in the mid-1990s. These early approaches were simple yet powerful. Even fitting functions $z = z(C, m_0)$ where z is a linear function or simple polynomial proved effective at estimating redshifts when appropriately tuned (Connolly et al. 1995). For example, Wang et al. (1998) computed photo-zs via a

linear relationship between redshift and three photometric colours ($U - B$, $B - V$, and $V - I$). The simple polynomial approach was applied successfully by e.g. Connolly et al. (1995), who fitted a second order polynomial to U , B_J , R_F and I_N magnitudes, and by Brunner et al. (1997), who employed an iterative process with several quadratic functions of U , B , R , and I magnitudes.

Early such forms of the empirical method proved superior to template methods (Walcher et al. 2011) for a number of reasons. Firstly, because the training set consists of real galaxies, there can be no issues with inaccurate templates. On top of that, the empirical relation is calibrated using magnitude measurements from survey data, so it inherently accounts for the effects of filter bandpasses and flux calibrations. Empirical methods can also easily provide realistic estimates of redshift uncertainties (Walcher et al. 2011), and they generally outperform template methods in terms of speed as well (Vanzella et al. 2004; Salvato et al. 2019).

A disadvantage of the empirical approach is that the derived relation $z = z(C, m_0)$ is only valid for the full survey if the training set is representative, that is, if the training set has similar observables C and m_0 to the full survey. This can be an issue when calculating photo-zs for objects much fainter than the training set (such as at high redshifts), because the relation has to be extrapolated in those cases. Another problem is that galaxies at redshifts beyond those in the training sample may have different intrinsic SEDs. A further, related drawback is that the sample also needs to be large enough to cover all colours, magnitudes, galaxy types, and redshifts to a statistically significant degree, so that the inferred empirical relation is statistically sound.

Empirical photo-z codes have become more advanced since the early methods described before, making use of sophisticated computational techniques which have developed alongside the field of machine learning. The core idea of these approaches is still the same as for the simple early methods: through machine learning, a training set is used to build up a relationship between survey observables and redshifts. Examples of specific algorithms include decision tree classification (e.g. **ArborZ**), random forests (e.g. **TPZ**), support vector machines (e.g. Wadadekar 2005), Gaussian processes (e.g. **GPz**), and Neural Networks (e.g. **ANNz**). Machine learning techniques inherit all the strengths and weaknesses of the early empirical methods listed above. However, they also enjoy the additional advantage of being able to include other observables on top of the

standard photometry. Examples include the bulge-to-total flux ratio (Sarajedini et al. 1999), surface brightness (Kurtz et al. 2007), Petrosian flux radius (Collister & Lahav 2004), and concentration index (Collister & Lahav 2004).

1.2.2.2 Template methods

Unlike the empirical strategy, which makes use of a spectroscopic training set, the template method computes photometric redshifts by directly comparing the observed photometry to a library of pre-existing SEDs. These SEDs can be empirical templates (widely used template sets include those by Coleman et al. 1980 and Kinney et al. 1996), or theoretical models (such as Bruzual & Charlot 2003).

All of the earliest photo-z implementations discussed in Section 1.2.1 were variations on the template fitting method. The pioneering work by Baum (1962), and the studies by Butchins (1981, 1983) and Loh & Spillar (1986) all used empirical spectra. In the mid-1980s, theoretical templates were first introduced by Koo (1985). That decade also saw the first implementation of a χ^2 minimisation procedure for template fitting (Puschell et al. 1982; see Section 3.2.1.1 for a detailed example of how χ^2 minimisation works in the context of a specific photo-z code). Such χ^2 algorithms, or variations thereof, have since become the most popular computational technique for contemporary template-based studies, and they are used in all subsequent papers and codes mentioned in this section.

Following on from this early period, template methods remained in frequent use when photometric redshifts became increasingly popular during the start of the CCD revolution in the 1990s. Significant works from this period include studies by Lanzetta et al. (1996), Gwyn & Hartwick (1996) and Fernández-Soto et al. (1999), who all capitalised on the newly available deep photometry from the Hubble Deep Field. Soon after, a large number of authors created many more template fitting codes, which were often made available for public use. Prominent examples are **HyperZ**, **BPZ**, **ZEBRA**, **LePHARE**, and **EAZY**, and many of these still remain in frequent use to the current day.

Compared to the empirical approach, a major asset of the template method is that it works even if there is no sufficient training set available. For this reason, template fitting codes are the preferred option when exploring new regimes, such as in new survey datasets that do not yet have large training sets available, or at high redshifts and faint magnitudes beyond the spectroscopic limit (Walcher et al. 2011; Leistedt & Hogg 2017). Additionally, a further advantage of template fitting

is that it can be performed in a fully probabilistic fashion, with explicit priors over redshifts and types of galaxies (Benítez 2000; Leistedt & Hogg 2017). The outputs of some machine learning codes could also be interpreted in probabilistic terms, but their priors are often complicated and implicitly constructed from training data, rather than specified explicitly. Furthermore, the template method allows additional physical properties such as ages or star formation rates to be obtained directly from the best-fit SEDs (Walcher et al. 2011; Leistedt & Hogg 2017). It must be noted, though, that some machine learning codes can predict these properties as well, given a suitable training set (see e.g. Ball et al. 2004).

The primary issues with the template fitting method lie with the template library, which needs to be representative of all galaxies in the survey. This is particularly an issue for theoretical model SEDs, which may not be entirely correct, and could be affected by specific details in the treatment of properties such as emission lines or dust reddening (Walcher et al. 2011). It is also important that the template set is complete, i.e. that it spans the entire colour space of galaxies in the survey. Empirical templates are most susceptible to this problem; firstly because empirical libraries contain a limited number of SEDs, and secondly because the spectra they are measured from are local and thus potentially intrinsically different from higher redshift galaxies at a different evolutionary stage (Walcher et al. 2011).

Another drawback of the template fitting technique is that it cannot fully capture flux complexities such as biases or error underestimation (Bonfield et al. 2010; Leistedt & Hogg 2017). These are exactly the types of situations where the empirical photo-z methods discussed in the previous section excel, as they can easily adapt to imperfect flux measurements. Some template fitting methods such as **LePHARE**, **EAZY**, and **GOODZ** (Dahlen et al. 2010) attempt to deal with this problem by adding correction terms to existing templates, or by adopting very flexible templates. For instance, **LePHARE** and **GOODZ** incorporate some of the strengths of empirical methods to mitigate the imperfect flux problem: they use a training set of galaxies to derive corrections to zero-points (**LePHARE**) or template SEDs (**GOODZ**) that minimise the scatter between photometric and spectroscopic redshifts. Nonetheless, these fixes often do not achieve the accuracy level attained by empirical codes. They can also be computationally expensive (Leistedt & Hogg 2017), although it must be remarked that the latter is not a prob-

lem when using small- or intermediate-scale surveys, for instance when hunting for high-redshift galaxies in deep surveys over relatively small areas.

Furthermore, colour-redshift degeneracies, which occur when the colours of a galaxy can be well fitted by SEDs at multiple redshifts, also present a challenge for photometric redshifts in general. They can be particularly problematic for template fitting codes, especially if the template set contains too many SEDs. Empirical codes can mostly circumvent the issue by breaking the degeneracy with additional information from magnitudes (or other observables for more recent machine-learning type codes). However, template methods have also found a way to deal with the problem via the use of Bayesian statistics. Through this framework, extra information can be included in the form of prior probabilities, which quantify the known likelihood of observed properties like galaxy shapes, angular sizes or magnitudes as a function of (measured spectroscopic) redshift. This extra information can break the degeneracy, because after incorporating prior probabilities one redshift solution will generally have a higher posterior probability than the others. In some cases, this could potentially even give template fitting codes an edge over the empirical approach, since it is possible to determine priors independently from the survey or field for which the photo-zs are to be computed. In fact, many codes include pre-computed priors from previous, well-established spectroscopic fields. Lastly, a Bayesian interpretation of SED fitting can also retrieve redshift probability functions. Template methods that utilise those can produce accurate photo-z uncertainties, allowing such codes to catch up with empirical codes regarding realistic errors. Bayesian statistics and prior probabilities were first implemented in **BPZ** by Bolzonella et al. (2000), and have since been included in many other codes (out of the ones mentioned in this chapter, specifically **ZEBRA**, **LePHARE**, **EAZY**, and **GOODZ**).

1.2.2.3 Comparison

Whether an empirical or template fitting approach would be more suitable for a given dataset depends on the specific application and the availability of spectroscopic redshifts. In those cases where a complete and representative sample of spectroscopic redshifts is available, empirical codes perform better (Abdalla et al. 2011; Salvato et al. 2019). Empirical codes are also not subject to photometry biases, and are therefore better if the aim is to reduce the impact of such biases, provided of course that a good sample of spectra is available. Finally, the fact

that empirical methods are computationally faster provides a distinct advantage for the most massive datasets. This can be a strong asset for the enormous contemporary large-scale surveys. On the other hand, in deep surveys over small areas which extend out to very high redshifts (where there is insufficient spectroscopic coverage), template fitting codes should be favoured (Hildebrandt et al. 2010). Regarding the performance of specific codes, Hildebrandt et al. (2010) and Abdalla et al. (2011) provide a detailed comparison of several publicly available photo-z methods, both template and empirical – although it is interesting to note that these two studies failed to identify a single best-performing code in either category.

1.2.3 Applications

The main drawback to the speed and efficiency of the photometric redshift strategy is a reduction in accuracy compared to spectroscopy. Photo-zs have larger errors on the redshift measurements, and they are subject to a greater risk of gross misestimation for a small number of individual objects (so-called catastrophic outliers). This can be a problem in areas where accuracy or certainty are of the utmost importance. However, since quantity is a quality of its own, for many scientific applications the reduced accuracy is an acceptable trade-off. Subsequently, photometric redshifts have become a key tool in many branches of extragalactic astronomy. This introduction cannot possibly do justice to all the existing applications in this extremely rich and fast-moving field, but some key examples will be presented briefly below.

1.2.3.1 Cosmology

Because they facilitate the statistical analysis of large bodies of galaxies, photometric redshifts have become essential in large parts of observational cosmology (Brammer et al. 2008). By mapping out the positions of galaxies through broad regions of space across cosmic time, photo-zs provide valuable information about cosmological large-scale structure (Peacock et al. 2006; Blake et al. 2007; Stanszewski et al. 2009; Weinberg et al. 2013; de Simoni et al. 2013; Abbott et al. 2018a; Elvin-Poole et al. 2018; Hikage et al. 2019). This large-scale structure, in turn, can provide a handle on dark energy and dark matter, and provide a better understanding of the geometry, composition, evolution, and fate, of the universe.

The digital revolution of the CCD era has played a crucial role in facilitating this research, and will continue to do so for the foreseeable future. Several current and future large-scale modern cosmology surveys, such as the SDSS, DES, Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018), and LSST are dedicated towards collecting broadband photometry for billions of objects, from which photometric redshifts can derive cosmological parameters.

1.2.3.2 Galaxy evolution

As suggested previously in Section 1.1.3.1, redshift estimates can also be used to select samples of galaxies in specific cosmic epochs, so that changes in their properties can be traced throughout the history of the universe. In this way, the photo-zs of galaxies in large multiwavelength datasets have long been used to study evolving galaxy populations (e.g. Fontana et al. 2000; Finkelstein et al. 2012; Bradley et al. 2014). The time evolution of galaxy sizes (Giallongo et al. 2000), their UV spectral slopes (Finkelstein et al. 2012), and their star formation rates (Daddi et al. 2007; Ilbert et al. 2013) are examples of the types of properties that have been traced in this way. The luminosity function (LF), which gives the number density of galaxies per unit of comoving volume per unit luminosity (Dunlop 2013), is a common method of summarising changing galaxy demographics. Photometric redshifts can be used to determine the LFs at different redshifts, which, for instance, can be used to study changes in the abundances of bright or faint galaxies (e.g. Poli et al. 2003; Ramos et al. 2011; McLure et al. 2010). By comparing these observational luminosity functions to results obtained from models, it is also possible to obtain insights into the physical processes that drive this evolution. The scientific value of the LF will be illustrated further in Section 1.3.2.2, specifically in the context of galaxies at $z \gtrsim 5$.

1.2.4 Conclusion on photometric redshifts

Altogether, we have thus far witnessed clearly how, thanks to their computational efficiency, photometric redshifts have become an excellent way to extract important information out of datasets from modern surveys. Thanks to that valuable quality, this way of estimating cosmic distances now forms a crucial tool for many types of extragalactic astronomy. A core part of this thesis therefore con-

cerns computing photo-zs for the optical+near-IR catalogue outlined earlier in Section 1.1.5.

The objective of this work is threefold. Firstly, since photometric redshifts have become so vital to astronomy, it is valuable to study how to optimise their application. For instance, it is important to investigate the effect of the many different code options that we have seen in Section 1.2.2. The optical+near-IR catalogue from this thesis can provide a good testing ground for such tests. Notably, the catalogue also enables an investigation of the extent to which photometric redshifts improve with the addition of near-infrared data, a result suggested by several theoretical (e.g. Banerji et al. 2008, Abdalla et al. 2008) and observational studies (Ilbert et al. 2009, Jarvis et al. 2013, Banerji et al. 2015). With projects such as LSST and the *James Webb Space Telescope* (*JWST*; Gardner et al. 2006) set to continue the progression in survey area and depth for at least the next 10 years, photo-zs will remain an extremely important area of study in the coming decades. With regard to the future, insights on how to perfect their accuracy are thus highly valuable scientifically. Secondly, because of their many applications in astronomy, providing photometric redshifts for the aforementioned optical+near-IR catalogue further increases the value of this dataset for the scientific community in general. Lastly, another core aim of this thesis, which will be introduced in the next part of this introduction, is to search for $z \geq 5$ galaxies in the new optical+near-IR dataset. Because a photometric redshift strategy is ideally suited to identifying such objects (which we have previously noted often lie beyond the spectroscopic limit), photo-zs form a vital part of this endeavour.

1.3 High-redshift galaxies

Returning briefly to the topic of modern surveys, Sections 1.1.3 and 1.1.4 have described how the impressive advance of deep field imaging over multiple wavelength regimes has opened the distant universe up to astronomers. Throughout this thesis, unless otherwise stated, the phrase ‘high-redshift’ is used to refer to far-away sources at $z \geq 5$, and closer objects are denoted as ‘low-redshift’. By observing sources at those high redshifts, astronomers are beginning to catch a glimpse of the formation of galaxies in the first billion years of cosmic time. The fact that we are now able to probe such very early times could perhaps be

considered one of the most impressive achievements of survey astronomy. The current part of this chapter will now introduce this rapidly advancing field.

1.3.1 Selection methods

Compared to the general population of celestial objects, high-redshift galaxies are faint, which makes it extremely time-consuming to measure their spectra, if they are bright enough that this is possible at all. Their faintness, compounded with the fact that they are also extremely rare, rules out spectroscopy as a feasible way to find $z \gtrsim 5$ sources. To overcome this obstacle, an arsenal of several alternative search strategies has been developed in the literature. As Section 1.2 suggested at the start, photometric redshifts are indeed one possibility. The following sections will present several other search methods, and place the photo-z approach within this general scientific landscape.

1.3.1.1 Narrowband searches for Lyman- α emitters

At high redshifts, an important – and frequently strong – spectral line for determining redshifts is Lyman- α at $\lambda_{\text{em}} = 1216 \text{ \AA}$, which corresponds to the transition between the ground state and first excited state of a neutral hydrogen atom. Although this line is undeniably best observed via spectroscopy, in the absence of spectra it is frequently still possible to detect it via the use of narrowband imaging. With this technique, large photometric catalogues which include such imaging can be used to search for so-called Lyman- α emitters (LAEs). These sources can be selected via the distinguishing property that they are brighter in a given narrowband filter compared to a broadband filter centred around the same frequency. This is because if the Lyman- α line falls within a narrowband filter, the bandpass-averaged flux of a LAE in this filter will be higher than its continuum measured by the wider broadband filter. The redshift of the LAE can then be estimated directly from the effective wavelength of the narrow bandpass.

Narrowband selection of LAEs was proposed early on by Partridge & Peebles (1967), and one of the first LAEs was found two decades later (Djorgovski et al. 1985). But it was not until the emergence of large-aperture telescopes and wide-field optical imagers that larger samples were discovered. Over the last two decades, thousands of LAEs have been found at $2 < z < 8$ (e.g. Cowie & Hu 1998; Rhoads et al. 2000; Ouchi et al. 2008; Tilvi et al. 2010), of which several hun-

dred at $z \gtrsim 5$ (e.g. Rhoads & Malhotra 2001; Ouchi et al. 2008, 2010; Hibon et al. 2010, 2011; Matthee et al. 2015). When studying these galaxies, it was found that on average, high-redshift LAEs selected via narrowband imaging are significantly fainter than galaxies selected through their continuum emission (as will be explained shortly, this is generally done via colour selection methods or photometric redshifts). The known samples of LAEs therefore form a complementary population to continuum-selected galaxies (Finkelstein 2016). Whether they are in essence a separate type of object, or whether narrowband searches are simply able to retrieve lower-mass galaxies from the same general population as the continuum-selected sources, is still a subject of active debate (see Dunlop 2013 for a comprehensive review of this topic).

1.3.1.2 Lyman-break selection

It is also possible to identify high-redshift galaxies using broadband filters alone. This method makes use of a distinctive spectral break that is present in the ultraviolet colours of young star-forming galaxies. Due to an abundance of young blue stars, these galaxies emit large amounts of UV continuum radiation (Madau et al. 1998). This UV emission displays a break shortward of the Lyman limit at $\lambda = 912 \text{ \AA}$, which corresponds to the minimum energy that can ionise a hydrogen atom in its ground state. This drop, often referred to as the Lyman break, originates from the hydrogen ionisation edge in the stellar atmosphere of massive stars, and is further accentuated by photoelectric absorption from interstellar neutral hydrogen (HI) gas. Since massive stars and HI gas are abundant in young star-forming galaxies, their 912 \AA Lyman break is generally very strong (typically $\gtrsim 2.5$ mag AB; Leitherer et al. 1995). At higher redshifts, the ultraviolet continuum between $\lambda = 912 \text{ \AA}$ and the Lyman- α transition at $\lambda = 1216 \text{ \AA}$ is also dimmed, this time by Lyman- α absorption from pockets of neutral hydrogen along the line of sight. This phenomenon, known as the Lyman- α forest⁶, causes a second drop at 1216 \AA . Since the number of intervening HI clouds generally increases with redshift, this secondary break becomes increasingly strong the

⁶Photons emitted by a distant galaxy at wavelengths between 912 \AA and 1216 \AA (as well as any remaining ones below 912 \AA), are redshifted to longer wavelengths by the expansion of the universe as they pass through space on their trajectory to earth. If they are redshifted to 1216 \AA while passing through an HI cloud, these photons will excite a Lyman- α transition and be absorbed. As the light travels through different pockets of neutral hydrogen at different redshifts, this process causes a series of absorption lines below 1216 \AA that is referred to as the Lyman- α forest.

further away a galaxy lies. At $z \sim 3$, it typically causes a drop in magnitude of roughly ~ 0.5 mag AB when measured through broadband filters (Madau 1995; Songaila 2004), and at $z \sim 6$ the decrease in magnitude has risen to $\sim 2\text{--}3$ mag AB (Songaila 2004). At high redshifts, the 1216 Å break hence overtakes the 912 Å break as the main distinguishing UV spectral feature. For this reason, the phrase ‘Lyman break’ often refers to the 1216 Å Lyman- α break in the context of high-redshift galaxies.

For many star-forming galaxies, the Lyman break is strong enough to be observable in broadband imaging, especially if it falls between two filters. Such a galaxy will be visible in all filters redder than the Lyman break, while it will be undetected (it will ‘drop out’) in all bluer filters. Because of this behaviour, such objects are often referred to as *dropout* galaxies, and the first filter longwards of the Lyman break is often referred to as the *detection* filter. Objects with a Lyman break between the *u*-band and *g*-band (or equivalently the *B*-band in other filter sets) are labelled *u*-dropouts. As their observed Lyman break lies between the *u* and *g* filter (roughly at $\lambda_{\text{obs}} \simeq 3500$ Å, depending on the filter set), their redshift is approximately $z \sim 3$. Using the exact same reasoning, *g*-dropouts (sometimes called *B*-dropouts) lie at $z \sim 4$, *r*-dropouts (or *V*-dropouts) at $z \sim 5$, and *i*-dropouts at $z \sim 6$. The dropout mechanism is depicted in Figure 1.2, which illustrates the appearance of Lyman-break galaxies at various redshifts.

The simplest way to select galaxies via their Lyman break involves using colour cuts to select galaxies that show a red colour between the dropout band and the detection band (to confirm a strong Lyman break) and a blue colour between this detection filter and the next reddest band (to select blue star-forming galaxies rather than low-redshift galaxies or stars with intrinsically red colours). This technique is demonstrated in the top right of Figure 1.2.

Naturally, it is also possible to find dropout galaxies in a specific redshift range using photometric redshifts. At high redshifts, such a photo-z approach can often essentially be viewed as a form of Lyman-break selection. This is because at $z \gtrsim 5$, the Balmer/4000 Å break — which is generally the second prominent spectral break — has redshifted beyond the near-infrared. When using only broadband optical and near-IR data, the Lyman break is therefore the only strong spectral feature that photo-z algorithms can use to estimate the redshift. Throughout this thesis, any galaxy selected via its broadband rest-frame UV col-

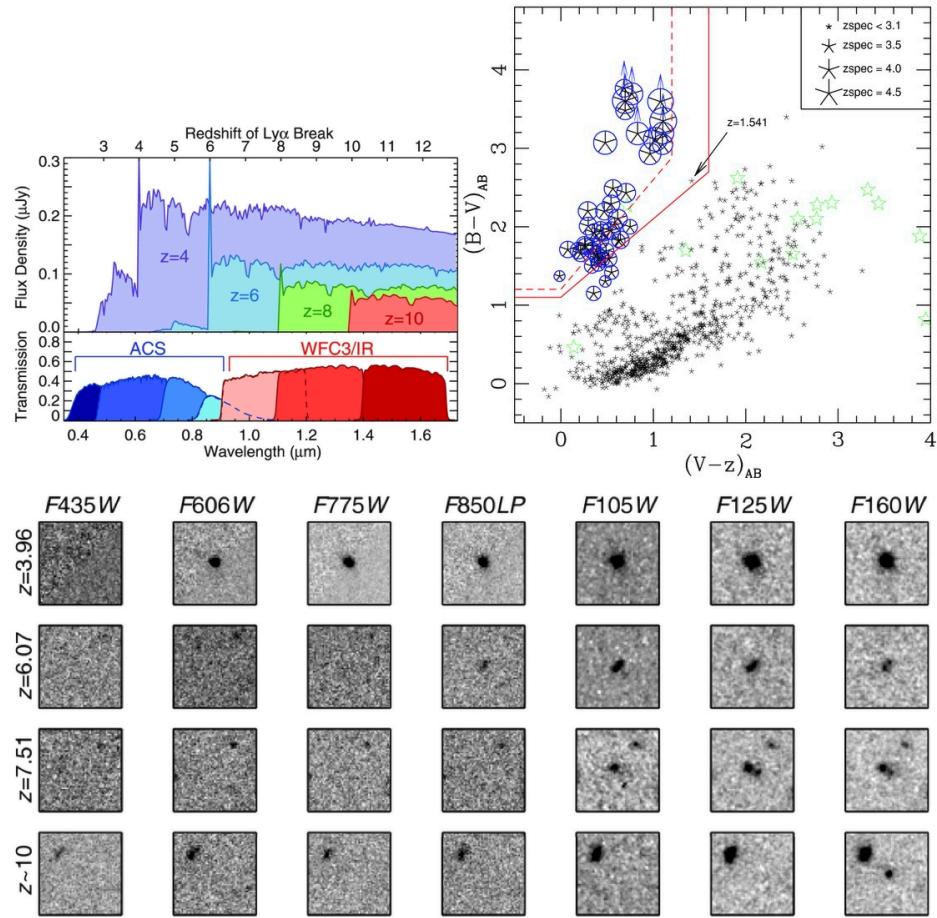


Figure 1.2 Illustration of the selection process for Lyman-break galaxies. The top left part is taken from Finkelstein (2016), and shows model spectra of star-forming galaxies at $z = 4$, $z = 6$, $z = 8$, and $z = 10$ together with the transmission curves for seven *HST* filters. The bottom part, also adapted from Finkelstein (2016), consists of cutouts of *HST* imaging showing real candidates in each of the aforementioned filters. It is clear that if the Lyman break lies between two filters, the galaxies show no flux ('drop out') in filters bluer than the Lyman break, but do appear in all redder filters. The top right figure is reproduced from Vanzella et al. (2009), and shows their colour-colour selection of B -dropouts (which lie at $3.1 < z < 4.4$) using the same filterset as the other two subfigures. The black markers indicate spectroscopically confirmed galaxies, with larger symbols denoting higher redshifts up to $z = 4.5$. The green \star symbols denote galactic stars, and the red box delineates the colour-colour cuts that were imposed to select B -drops. It is noteworthy that the only contamination inside the box consists of a single low-redshift galaxy (flagged by the arrow), thus demonstrating that the colour cuts lead to highly efficient selection.

ours will be referred to as a *Lyman-break galaxy* (LBG). This includes objects selected via colour-colour cuts as well as via photometric redshifts.

Colour selection methods and photometric redshifts each have their own merits. The colour-cut method is conceptually simple, computationally fast, not affected by specifics in the photo-z code (such as the template set), and highly reproducible. These advantages also make it more straightforward to model contamination and completeness rates for selected LBG samples, which is necessary for some statistical studies (e.g. the luminosity function introduced in Section 1.2.3.2 and to be discussed further in Section 1.3.2.2). On the other hand, colour cuts only make use of data in at most three filters, whereas the photo-z approach makes use of the full available dataset. Additionally, as indicated in Section 1.2.2, some photometric redshift methods can also constrain certain galaxy properties, as well as output a redshift probability distribution for each source. The latter estimates the probability of potential low-redshift solutions, and also helps to increase the redshift precision (with photometric redshifts the precision at $z \sim 6$ is around $\Delta z \sim \pm 0.2\text{--}0.3$, as opposed to $\Delta z \sim 0.5$ for colour-cut techniques; Finkelstein 2016). The history and current status of observational searches for LBGs will be discussed in Section 1.3.2.1, where it will be shown that both photo-zs and colour cuts remain competitive up to the current day. The choice of method therefore largely depends on the research goals and available filtersets for each LBG study, and on the personal preference of the authors.

1.3.1.3 Quiescent and dusty high-redshift galaxies

The selection of LAEs and LBGs described above is based on their rest-frame UV emission, which means that these methods are only sensitive to galaxies with active star formation that is not significantly attenuated by dust. Therefore, these strategies will miss extremely dusty galaxies, as well as passive quiescent galaxies that no longer form stars.

However, an alternative search method that makes use of submillimeter imaging has revealed several highly dust-attenuated galaxies at $z \gtrsim 5$ (see e.g. Riechers et al. 2010; Walter et al. 2012). The UV radiation produced by the intense star formation in these extremely reddened objects is almost entirely absorbed by the vast amounts of dust in their interstellar media. This process heats the dust, which then re-radiates its energy at (rest-frame) far-infrared wavelengths. At high redshifts, this emission is shifted into the submillimeter regime, where

it can be detected by instruments such as the *Herschell Space Observatory* and the James Clerk Maxwell Telescope. An in-depth discussion of extremely dusty galaxies is outside the scope of this thesis, but Casey et al. (2014) provide an excellent review for those who are interested in more information.

Regarding quiescent galaxies, there are currently no known secure candidates⁷ of this population at $z \gtrsim 5$. While photometric redshifts and colour-colour cuts directed at the Balmer break⁸ have yielded a handful of plausible candidates at $z \sim 4.5$ (Nayyeri et al. 2014; Merlin et al. 2018), none of these are spectroscopically confirmed. The fact that searches have thus far not returned any convincing results is in fact not unexpected; since galaxies at $z > 5$ would have needed to reach quiescence within 1.2 Gyr of the Big Bang, a large population of such objects is very unlikely. If they do indeed exist, the deep near-IR imaging that will soon be available with *JWST* should be able to uncover them (Finkelstein 2016; Mawatari et al. 2016).

1.3.2 Lyman-break galaxies at high redshifts

Out of all the different types of high-redshift galaxies discussed above, this thesis exclusively concerns itself with Lyman-break sources, for reasons that will be substantiated in Section 4.1. The rest of the current review of high-redshift science will therefore focus exclusively on this topic. The discussion will cover a brief overview of $z \gtrsim 5$ LBG searches in the literature, as well as an outline of several interesting quantities and physical probes that can be studied via these objects.

1.3.2.1 Historical overview and state of the art

The first attempts to identify distant galaxies by their Lyman break began at the start of the CCD era in the early 1990s (Guhathakurta et al. 1990; Steidel & Hamilton 1993). These early studies used colour-colour cuts to select u -dropouts around $z \sim 3$, and the first such candidate was soon spectroscopically confirmed (Giavalisco et al. 1994). By the end of the decade, small samples of galaxies were

⁷Although it must be noted that Mawatari et al. (2016) have found a number of uncertain contenders.

⁸This method works analogously to the Lyman-break colour selection described in Section 1.3.1.2.

starting to be found around $z \sim 5$, with the rapidly developing photo-z techniques providing a valuable tool in this endeavour (Lanzetta et al. 1996; Fernández-Soto et al. 1999). A handful of these candidates quickly received spectroscopic confirmation (Spinrad et al. 1998; Weymann et al. 1998). Nevertheless, larger samples of high-redshift galaxies could not be identified until the development of better red-sensitive instrumentation. After this need was met by new *HST* instruments in the mid-2000s, space-based surveys such as GOODS and the HUDF (see Table 1.1 for both) started to unveil samples of more than a hundred $z > 5$ galaxy candidates, discovered through either colour-cut or photometric redshift methods (Stanway et al. 2003; Giavalisco et al. 2004b; Grazian et al. 2006; Bouwens et al. 2006). One of the first spectroscopic confirmations of a $z \sim 6$ candidate followed soon after (Dickinson et al. 2004). Around the same time, parallel improvements in terrestrial telescopes enabled significant progress from the ground. Surveys such as the SDF, SDXF (both optical, see Table 1.2), and UKIDSS (near-IR, idem) prompted the discovery of large samples of high-redshift LBGs from ground-based data, again using both colour cuts and photo-z selection. Although only just over twenty candidates were found at $z > 7$ (Ouchi et al. 2009), more than a hundred were discovered at $z \sim 5$ and $z \sim 6$ (Nagao et al. 2004; Ouchi et al. 2004; McLure et al. 2006; Yoshida et al. 2006). These ground-based searches covered areas of $\sim 1 \text{ deg}^2$ – more than a two-fold increase over GOODS, the largest space-based project at the time. Since larger search areas are more likely to contain rare objects, these ground-based datasets complemented space-based imaging by revealing bright galaxies (<26.0 mag AB), which are too uncommon for the relatively narrow surveys from space.

The last 10 years brought even further advancements in the selection of LBG samples. From space, improvements in instrumentation led to deeper and wider surveys⁹ such as CANDELS (see Table 1.1), HUDF09 (PI: Illingworth, described in e.g. Bouwens et al. 2011), and HUDF12 (Koekemoer et al. 2013), which together yielded around 3000 $z \sim 5$ and 800 $z \sim 6$ LBG candidates (Finkelstein et al. 2015b; Bouwens et al. 2015). The extreme depths also allowed the selection of several hundred secure candidates out to $z \sim 7\text{--}8$ (Finkelstein et al. 2015b; Bouwens

⁹The CANDELS, HUDF09 and HUDF12 data have also been combined into the HXDF, which was introduced in Section 1.1.3.2.

et al. 2015), and roughly a few dozen¹⁰ objects at $z \gtrsim 9$ (Oesch et al. 2013, 2014; Bouwens et al. 2015). Meanwhile, there similarly were significant advances from the ground, with surveys such as CFHTLS, UltraVISTA, and VIDEO (see Table 1.2 for all) covering footprints of $> 1 \text{ deg}^2$. Those larger search areas opened up the study of the brightest ($z_{\text{AB}} < 25.0$) high-redshift candidates, which will be discussed in more detail shortly. They also revealed several dozen $z \sim 7$ candidates (Castellano et al. 2010; Bowler et al. 2014), but the vast majority of ground-based objects were discovered at $z \sim 5\text{--}6$, with a combined total of hundreds of plausible LBGs at $z \sim 6$ (Willott et al. 2013; Bowler et al. 2015; McLure et al. 2009) and thousands at $z \sim 5$ (McLure et al. 2009; Van der Burg et al. 2010). The reason why so many more $z \sim 5$ candidates have been found compared to $z \sim 6$, besides an intrinsic drop in number density with increasing redshift, lies partly in the available datasets. The underlying cause is that secure redshift estimates require detections in at least 2 bands redwards of the Lyman break¹¹. Since many optical datasets include i - and z -band data, r -dropouts at $z \sim 5$ can be found from optical data alone, since they can be detected in both i and z . However, additional redder filters are required for i -dropouts at $z \sim 6$, which means that these searches are restricted by the availability of deep near-IR data.

Up until the current day, colour-cut methods and photometric redshift selection both remain competitive, and the studies mentioned above have used both approaches in roughly equal numbers. Out of all the LBG candidates that all those searches have identified, more than a hundred $z \gtrsim 5$ galaxies have so far been spectroscopically confirmed, mostly at $z < 7$ (e.g. Malhotra et al. 2005; Vanzella et al. 2009; Stark et al. 2010; Jiang et al. 2011; Curtis-Lake et al. 2012; Pentericci et al. 2014; Kakiuchi et al. 2018).

The picture that emerges from this collective body of research is rather impressive. Astronomers are now able to catch a first glimpse of the $z \gtrsim 9$ universe, and statistically secure samples of hundreds of $z \sim 7\text{--}8$ galaxies are beginning to

¹⁰ Among these is an object spectroscopically confirmed at $z = 11.1$ (Oesch et al. 2016). This galaxy is currently the most distant known source in the universe. As this implies that we are now able to observe a glimpse of the universe just 400Myr after the Big Bang, its discovery can certainly be regarded as a great triumph of extragalactic astronomy.

¹¹ This is necessary to select galaxies with blue intrinsic colours, in order to remove contamination from intrinsically red objects such as dusty galaxies and low mass stars. This requirement applies to both photometric redshift methods and colour cuts, although in the case of photometric redshifts, more detections are additionally useful for providing a better constraint on the fit.

emerge. The state of the art at $z \sim 5$ and $z \sim 6$ also contains a wealth of data, with over 1000 and 10 000 candidates respectively. Nevertheless, despite all this progress, number counts of the brightest ($z_{\text{AB}} < 25.0$) LBGs at $z \sim 6$ are still relatively low. At the time of writing¹², there are 14 such candidates from a 4 deg^2 ground-based study by Willott et al. (2013), 9 from a 1.65 deg^2 ground-based search by Bowler et al. (2015), and approximately¹³ 13 from 0.26 deg^2 from a predominantly space-based investigation by Bouwens et al. (2015). After accounting for duplicates, those studies contain a relatively modest total of 32 unique $z_{\text{AB}} < 25.0$ candidates. A combination of deep optical and near-IR imaging over a large area of several square degrees would be required to improve these numbers.

1.3.2.2 The importance of Lyman-break galaxies for high-redshift galaxy evolution studies

Samples of high-redshift Lyman-break galaxies from the last 20 years have contributed enormously to the study of galaxy evolution at early times. This is an extremely rich field with a vast quantity of ongoing research, and a detailed summary of recent developments can be found in the reviews by Dunlop (2013) and Finkelstein (2016). However, in order to provide a brief illustration of the interesting science that can be achieved with high-redshift LBGs, this section highlights¹⁴ a number of interesting galaxy properties that can be investigated through these objects.

UV luminosity function

One of the most straightforward quantities to measure from samples of Lyman-break galaxies is the UV luminosity function (LF; briefly introduced in Section 1.2.3.2 earlier), which quantifies the number density of galaxies per unit of comoving volume per unit of absolute rest-frame UV luminosity (usually $M_{1600 \text{ \AA}}$ or

¹²It is possible that the study by Finkelstein et al. (2015b) also contains some candidates, but they did not publish apparent magnitudes, and the footprint of their study is almost entirely contained within Bouwens et al. (2015), so these results are omitted.

¹³It must be noted that these galaxies are classified as bright due to their $Y_{105} < 25.0$ magnitude, as Bouwens et al. (2015) did not publish z -band magnitudes. Therefore it is possible that the actual number of $z < 25.0$ slightly deviates from this number.

¹⁴Since the target of this thesis is bright LBGs (for reasons that will be presented in Section 1.3.3 shortly), the narrative will focus predominantly on those objects. More information on all the science that can be achieved with fainter sources can be found in e.g. Dunlop (2013) and Finkelstein (2016).

$M_{1500 \text{ \AA}}$). The reason why LBG LFs are usually measured in the ultraviolet is because the deepest survey data generally exists at optical and near-IR frequencies, which for LBG redshifts of $z \gtrsim 3$ translates to the rest-frame UV.

The UV luminosity function is an extremely useful measure, as it provides insights into physical processes in the universe. For instance, it can be used to estimate the total number of ionizing UV photons produced by galaxies, and thus whether galaxies could have been the main contributor to the reionisation of the intergalactic medium (Robertson et al. 2010, 2013). Furthermore, because it tracks light from newly formed blue stars, the UV luminosity function can also be converted into the cosmic star formation rate density, which quantifies the total amount of star formation occurring in the universe at a given redshift. Studying the variation of the cosmic star formation rate with redshift has started to provide some insights into the growth of stellar mass, and when its build-up commenced (Ellis et al. 2013; Madau & Dickinson 2014; Bouwens et al. 2015; Finkelstein et al. 2015b). Lastly, because it can be computed fairly easily from both theory and observations, the LF is a useful probe for some of the physical processes that drive galaxy evolution. A comparison between the shape of the luminosity function (from observations) and the halo mass function of the underlying dark matter haloes (from simulations) can provide clues as to the physical mechanisms at play. At both the bright and the faint end, the luminosity function at most redshifts has generally been found to lie below what would be expected on the basis of galaxy halo masses. This suggests that the conversion of gas (the abundance of which roughly scales with the dark matter halo mass) into starlight is less efficient in both the brightest and the faintest galaxies. It is believed that this phenomenon is caused primarily by processes that heat the interstellar medium and expel gas from galaxies, which subsequently suppresses their star formation. The primary candidates for causing this disruption include supernova feedback and stellar winds at the faint end, and AGN feedback at the bright end (Benson et al. 2003; Somerville et al. 2008; Somerville & Davé 2015).

To study the changes in all these physical processes over time, UV luminosity functions are usually computed from samples of LBGs at various redshifts. In order to be able to reach secure conclusions, it is important that these functions are well measured. To illustrate to what extent current data provide adequate constraints, Figure 1.3 shows a compilation of LF data points from recent observational studies at $4 < z < 10$. It is clear that the measurements at $z \sim 4$ provide

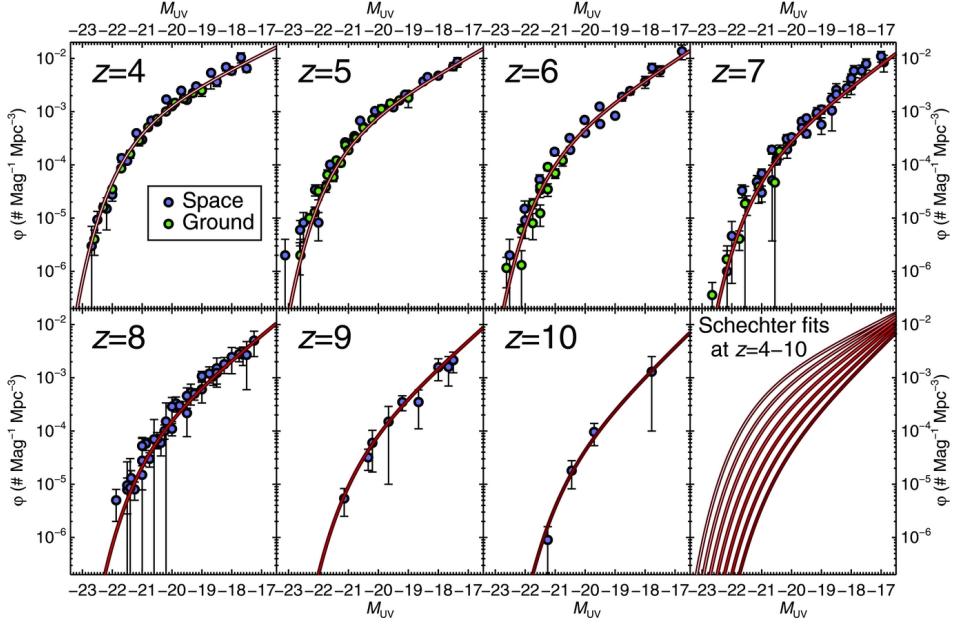


Figure 1.3 A compilation of UV luminosity functions created by Finkelstein (2016), constructed from recent samples of LBGs in various redshift bins. The dots are data points from several space-based and ground-based studies (sources are listed in Finkelstein 2016), and the red curves are reference Schechter function (Schechter 1976) fits to these data points (computed as described in Finkelstein 2016).

fairly solid constraints, thanks to the generous samples of LBGs used to obtain these data points. Similarly, it is evident that the same largely applies at $z \sim 5$ (with the exception of the extreme bright end, which shows some large error bars mainly due to space-based studies having low number counts in this regime). And conversely, it is also apparent that the results at $z \gtrsim 7$ are generally rather uncertain, primarily due to the small sample sizes. The data points at $z \sim 6$, on the other hand, present a more mixed situation. At the faint end, the uncertainties are small, although there does exist some scatter (likely originating from systematic differences in the LF computation between studies; Finkelstein 2016). The bright end likewise suggests some tension between studies, but critically shows much larger error bars in addition to that. Those big uncertainties are predominantly caused by the small number of known bright galaxies, an issue which has previously been noted in Section 1.3.2.1. Better constraints in this regime would be valuable, since the bright end around $z \sim 6$ also happens to be especially interesting scientifically. This is because there are indications that

the aforementioned drop in bright sources compared to the halo mass function, which is known to exist at lower redshifts (e.g. Van der Burg et al. 2010), may have commenced around this time (Bowler et al. 2014, 2015). This could indicate the start of processes that suppress star formation (such as the beginning of AGN activity), or the onset of dust build-up that obscures some of the stellar UV radiation (see Finkelstein et al. 2012; Bouwens et al. 2014; Rogers et al. 2014).

Other galaxy properties

In addition to the UV luminosity function, there are several other important physical attributes that can be studied with the help of high-redshift LBGs. A major one of these is their stellar masses, which can be probed if mid-IR *Spitzer Space Telescope* data is available (Yan et al. 2006; Stark et al. 2009; Finkelstein et al. 2010; González et al. 2011). Such imaging captures the rest-frame optical emission, which, in contrast to the rest-frame UV, is sensitive to lower-mass stars that may make up the bulk of galaxy mass. As well as for its own sake, stellar mass has been used to study the amount of star formation in a galaxy per unit of stellar mass (the specific star formation rate; Stark et al. 2013; González et al. 2014; Salmon et al. 2015) and the dependence of galaxy stellar masses on the underlying dark matter halo mass (McLure et al. 2009; Behroozi et al. 2013; Finkelstein et al. 2015a). Dust attenuation is another interesting property that is being studied, as there is a reasonable amount of evidence that out to $z \sim 7$ the brightest galaxies already contain an appreciable amount of dust (displaying reddening of $E(B - V) \sim 0.1\text{--}0.15$; Bouwens et al. 2012, 2014; Finkelstein et al. 2012; Rogers et al. 2014; Finkelstein et al. 2015a). Finally, some authors have also begun to make small steps to measure the evolution of galaxy sizes (Oesch et al. 2010a; Huang et al. 2013; Willott et al. 2013; Bowler et al. 2014).

Spectroscopic follow-up of high-redshift LBG candidates can also provide interesting physical insights, since spectra can be used to measure the possible presence and strength of the Lyman- α emission line. This can provide constraints on reionisation: if galaxies beyond a certain redshift show an average reduction in their Lyman- α flux, this could signal the presence of increased neutral hydrogen in the IGM (and hence incomplete reionisation). Recent studies indicate that such a drop in Lyman- α has indeed been observed around $z > 6$ (Stark et al. 2010; Pentericci et al. 2011; Finkelstein et al. 2013; Pentericci et al. 2014), which agrees with the current consensus that reionisation ended around $z \approx 6$.

The specifics of how the reduction in Lyman- α at $z > 6$ can be converted into particular details about the progression of reionisation are currently the topic of active debate, and there are some tensions between theory and observation (see e.g. the reviews by Dijkstra 2014 and Finkelstein 2016, and references therein). While spectroscopy at $z > 6.5$ would provide the most useful constraints, this is challenging with most instruments due to the dramatic reduction in the sensitivity of optical spectrographs at the extreme red end. However, spectra of $z \approx 6$ galaxies are easier to obtain and would also be useful. For instance, galaxies at $z \approx 6$ can be used to extrapolate the Lyman- α equivalent width and escape fraction at $z > 6.5$ in the absence of neutral hydrogen (Stark et al. 2009, 2011). This baseline can then be compared to the observed signature from real $z > 6.5$ galaxies to estimate how much (not-yet-ionised) neutral hydrogen gas is present in the IGM. There is evidence that the Ly α escape fraction and equivalent width decrease with luminosity, but for the brightest sources the uncertainties on this hypothesis are still large due to the small number counts (Stark et al. 2009, 2011; Curtis-Lake et al. 2012). More spectra of luminous $z \approx 6$ objects would therefore be useful. At these great distances, bright galaxies are also more likely to yield usable spectra, making them very suitable targets.

1.3.3 Conclusion on high-redshift galaxies

In the review above, which focused primarily on Lyman-break galaxies, we have witnessed some key developments in recent high-redshift galaxy evolution studies. Along the way, the discussion presented a strong case in favour of collecting larger samples of the brightest ($z_{\text{AB}} < 25.0$) galaxies at $z \sim 6$ in particular. For these objects, it was found that the current sample sizes are still small, which leads to large uncertainties on statistical properties. We observed several areas where bright LBG candidates would be valuable for future science. First and foremost, they are required to better constrain the bright end of the luminosity function, which can provide interesting clues as to the onset of AGN activity or dust build-up in the early universe. On top of that, they can also be used to investigate other physical properties, such as galaxy stellar mass and dust content. Finally, if spectra can be obtained, the Lyman- α emission of bright sources can help to provide further constraints on reionisation.

The issue with the small number counts of known bright galaxies is largely caused by the relatively modest footprints of previous optical+near-IR surveys,

which did not cover a large enough deep area to find substantial numbers of these rare sources. However, with an effective search area of 4.9 deg^2 (7.5σ depths of $z_{AB} < 25.0$), the new optical+near-IR catalogue presented in this thesis offers the largest deep search area to date. Since its footprint only partially overlaps with the previous largest 4 deg^2 study by Willott et al. (2013), it can provide an appreciable increase in potential sources. Armed with this new dataset, the third part of this thesis thus tackles the important task of identifying additional bright high-redshift Lyman-break galaxies. The largest scientific potential lies in sources at $z \sim 6$; but since the dataset is also suitable for objects at $z \sim 5$ those sources are also investigated, as we have seen that the extreme bright end of the LF at that redshift does still show considerable uncertainties. The high-redshift samples from these searches will be very valuable for future studies, for instance to help address some of the open questions regarding galaxy properties mentioned in the previous paragraph and Section 1.3.2.2.

1.4 Thesis outline

Altogether, the current chapter has covered three important and related research areas in contemporary astronomy: surveys, photometric redshifts, and high-redshift galaxies. Each of these topics corresponds to one chapter of this thesis, and the main aims of each part have already been identified in the intermediary conclusions of each section (i.e. Sections 1.1.5, 1.2.4, and 1.3.3). A more detailed structural outline of this thesis is given below:

- **Chapter 2** will address the creation of a **combined optical and near-infrared source catalogue**. The first section of this chapter will introduce the datasets for this thesis, which consist of the optical (*grizY*) Dark Energy Survey (DES) deep fields, the near-IR (*ZYJHK_s*) Vista Deep Extragalactic Observations (VIDEO) survey, and a catalogue of publicly available spectroscopic redshifts for a subset of these sources. The second section will then describe the process of combining the DES and VIDEO datasets into a single multiwavelength catalogue. This is achieved by creating a pipeline that extracts VIDEO forced photometry directly from the imaging and merges it with pre-existing DES collaboration catalogues. The third section will address accuracy verification for the resulting combined DES+

VIDEO catalogue, and conclude with an overview of the final photometric data product.

- **Chapter 3** will discuss the computation of **photometric redshifts for the ($grizZYJHK_s$) DES+VIDEO catalogue**. The first part will present the template fitting method **LePHARE** as the choice of code, and describe the working of this algorithm together with its various options and the relevant template SEDs. The second part will then introduce various **LePHARE** configurations for testing different templates, **LePHARE** options, and photometry. Each setup is used to compute photo-zs for a subset of DES+VIDEO objects with spectroscopic redshifts, and the photo-z performance of these test runs is assessed via a number of accuracy metrics that capture the difference between photometric and spectroscopic redshift. The primary purpose of this analysis is to identify the most suitable configuration to compute redshifts for the high-redshift search in Chapter 4 (see below), but the discussion will also address the implication of the test results for general photo-z research. The final part of the chapter will then focus on applying the chosen configurations to the full DES+VIDEO catalogue. This section will also discuss finding an appropriate star-galaxy separation strategy to filter out nonsensical photo-zs from stars, before finally presenting the final redshifts for all galaxies in the DES+VIDEO dataset.
- **Chapter 4** will describe the **search for bright $z \sim 5$ and $z \sim 6$ Lyman-break galaxies in the DES+VIDEO catalogue**. The first part will discuss identifying a suitable strategy to select a sample of $m_{AB} < 25.0$ candidate LBGs at $z \sim 5$ and $z \sim 6$, detailing the three selection rounds used in this search. The second part will then present the resulting final LBG sample, and compare the outcomes to other studies in the literature.

Chapter 2

Combined DES+VIDEO catalogue

Astronomy is undoubtedly in a data driven golden age of surveys. Chapter 1 described how, spurred by technological advances in imaging and data processing technology, this revolution has enabled the efficient discovery of millions of galaxies in deep fields. Moreover, the observation of these sources over many parts of the electromagnetic spectrum has opened up a wide range of tracers for galaxy activity. With studies starting to constrain galaxy properties all the way into the first billion years of cosmic evolution, further opportunities for progress now lie in the direction of deeper surveys and larger areas. With a combination of optical and near-infrared imaging over the largest deep ($z_{\text{AB}} < 24.9$, $K_s < 23.8$) area to date, this thesis tackles the latter strategy.

The process of merging optical data from the Dark Energy Survey (DES) and near-infrared data from the VISTA Deep Extragalactic Observations (VIDEO) survey into a combined DES+VIDEO catalogue forms the topic of the current chapter. The first section will introduce the core DES and VIDEO datasets, and present the spectra that are used as auxiliary data in later chapters. After that, the discourse will turn to the forced photometry process that combines the DES and VIDEO data, as well as various data checks to confirm the validity of the resulting combined catalogue.

2.1 Data

2.1.1 The Dark Energy Survey (DES)

2.1.1.1 Overview

The Dark Energy Survey (DES; Flaugher 2005; Abbott et al. 2018b) is an optical photometric survey with the primary scientific goal of probing the nature of dark energy by measuring the history of cosmic expansion to high precision. It achieves this through four independent probes: supernova distances, galaxy cluster counts, weak gravitational lensing and galaxy angular clustering. The last three quantities are measured by the so-called *wide-field* survey, which covers 5000 deg^2 in five optical *grizY* filters. This imaging provides the required morphological information and number counts, which can be tracked over cosmic time via photometric redshifts, similarly computed from the broadband data. The *time-domain* survey contains the supernova part of DES. This transient component covers 27 deg^2 imaged at regular seven-night intervals in *griz* filters, in order to detect Type Ia supernovæ and measure their lightcurves. Because Type Ia supernovæ are standard candles, measuring their light output provides an independent distance measurement to track the expansion of the universe, and thus constrain dark energy.

2.1.1.2 Data collection

Imaging for DES has been obtained with the specially constructed Dark Energy Camera (DECam; Flaugher et al. 2015), mounted on the Blanco 4-m telescope at the Cerro Tololo Inter-American Observatory. Containing the latest wide-field imaging technology, DECam has a 3 deg^2 field of view and a 570 megapixel CCD camera, made up of 62 arrays arranged in a hexagonal pattern as per Figure 2.1. The camera has a resolution of $0.263 \text{ arcsec pix}^{-1}$. Observations for DES were conducted using the five main DECam broadband filters *grizY*, of which the transmission curves are plotted in Figure 2.2 and wavelength coverage provided in Table 2.1.

The survey saw first light in September 2012, and observations for the full survey were carried out between August 2013 and January 2019. Over these six years, the 5000 deg^2 wide-field area was covered several times in all five *grizY* filters to create the footprint shown in Figure 2.3. In the wide component, each

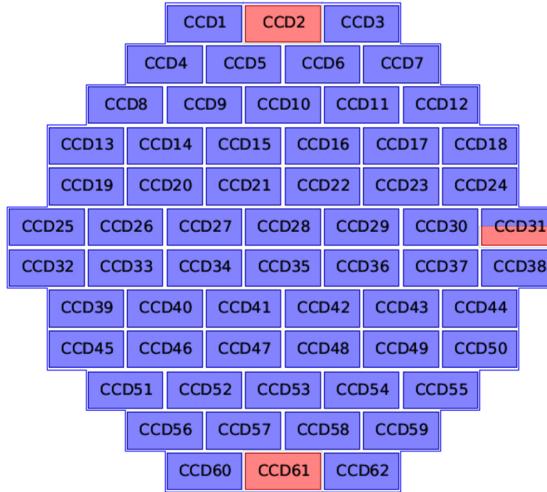


Figure 2.1 Focal plane of the Dark Energy Camera (DECam), consisting of 62 CCDs arranged in a hexagonal pattern covering a 3 deg^2 field-of-view. North is displayed at the top and east is on the right. The red sections (CCD 2, CCD 61, and half of CCD 31) were not functional for most of the Year 1 observations used in this catalogue, and data from these regions is not used in this thesis. *Image credit:* Abbott et al. (2018b).

(hexagonal) exposure in the same field was offset by roughly half the focal plane radius, such that objects were observed by different CCDs in each visit. This dithering strategy minimises the effects from inhomogeneities in the DECam apparatus and enhances the photometric calibration. Every year, the exposures collected up to that point are co-added to produce deep imaging (see Abbott et al. 2018b and Morganson et al. 2018 for details), which will reach the final survey depths after the full five years' worth of data have been processed.

The time-domain survey has employed a different observing strategy. In this component, the target fields were observed frequently using *griz* filters only, at regular intervals approximately seven nights apart (Kessler et al. 2015). Similarly to the wide-field survey, these exposures are then co-added. These co-added images constitute the so-called DES deep fields, interchangeably referred to as supernova fields. As they contain a substantially higher number of exposures per unit area, these fields reach $\sim 1\text{--}2$ mag AB deeper in *griz* compared to the rest of DES. Y-band coverage in the supernova fields is provided by the (shallower) wide-field data. In contrast with the main survey, where subsequent exposures in the same field were dithered to achieve maximum data uniformity, each supernova pointing in a given field is the same to within a few arcseconds (Kessler

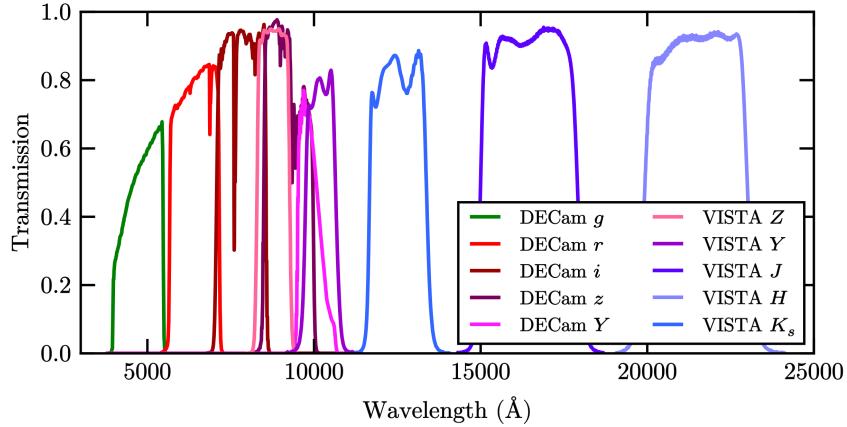


Figure 2.2 Transmission curves for the optical to near-infrared filters used in this project from DES (DECam) and VIDEO (VISTA). These plots show the relative total throughput, including the effects of the atmosphere and quantum efficiency of the detector.

WAVELENGTH COVERAGE OF FILTERS

Filter	DECAM			Filter	VISTA		
	λ_{cent} (Å)	λ_{min} (Å)	λ_{max} (Å)		λ_{cent} (Å)	λ_{min} (Å)	λ_{max} (Å)
<i>g</i>	4730	3980	5480	<i>Z</i>	8780	8285	9255
<i>r</i>	6420	5680	7160	<i>Y</i>	10200	9735	10665
<i>i</i>	7840	7100	8570	<i>J</i>	12520	11660	13380
<i>z</i>	9260	8500	10020	<i>H</i>	16450	14995	17905
<i>Y</i>	10090	9530	10650	<i>K_s</i>	21470	19925	23015

Table 2.1 The central wavelengths λ_{cent} , and left and right edges λ_{min} and λ_{max} for all DECam and VISTA filters used in this thesis. λ_{min} and λ_{max} are defined as the wavelength at which 50% relative transmission occurs. Each central wavelength is a simple interpolation between λ_{min} and λ_{max} .

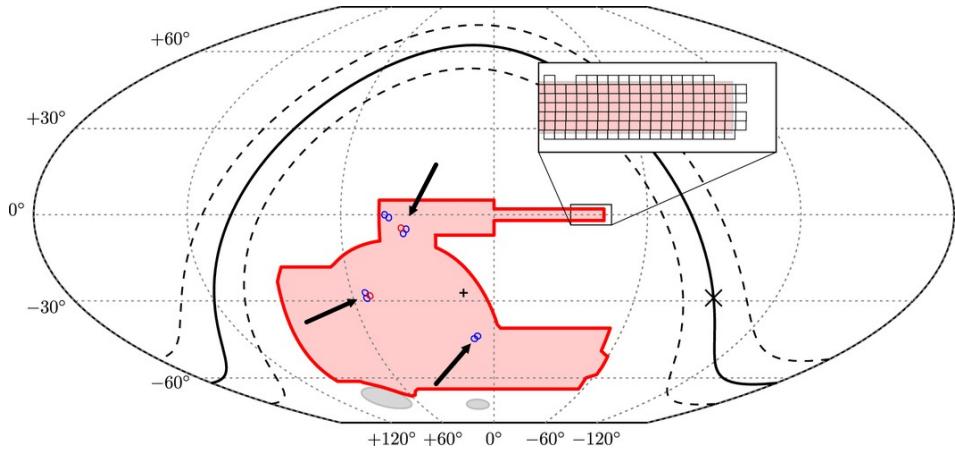


Figure 2.3 Map of the Dark Energy Survey (DES) area in celestial coordinates. The 5000 deg^2 wide-field survey is shown in red. The eight shallow supernova (SN) fields are displayed as blue circles, and the two deep SN fields are displayed as red circles. The catalogue created in this thesis is based on the eight most southern SN fields, denoted by the black arrows. The other two SN fields around the celestial equator are not included in this work, because they do not overlap with any near-infrared data of a suitable depth. The black line presents the Milky Way plane, and the Galactic center and Galactic pole are marked by a cross (\times) and a plus ($+$) respectively. The inlay shows the way the wide-field survey is cut up into co-add tiles by the DESDM processing pipeline. *Image credit:* Abbott et al. (2018b)

et al. 2015). While this is useful for supernova science, since it makes it easy to identify if an object has changed in brightness between pointings, an unfortunate side effect of this observing strategy is the presence of chip gaps in the co-added images. These chip gaps originate from the arrangement of the DECam CCDs, which, as demonstrated in Figure 2.1, are not perfectly adjacent (see Figures 2.4, 2.5, and 2.6 for a visible imprint of these gaps in the data). To compensate for the lack of dithering, the supernova survey used a special photometric calibration strategy, detailed in Kessler et al. (2015).

The DES deep fields, referred to as the **DFULL** dataset by the DES collaboration, form the optical component of the DES+VIDEO catalogue presented in this thesis. To illustrate the positions of the fields, their locations are superimposed on the wide-area DES footprint in Figure 2.3. It is useful to emphasise that, even though DES has observed a total of ten supernova fields, this thesis includes only eight fields (E1, E2, X1, X2, X3, C1, C2 and C3). The other two (S1 and S2, overlapping the SDSS Stripe 82; Adelman-McCarthy et al. 2007; Abazajian et al. 2009) were disregarded due to a lack of available deep near-infrared data. Cent-

ral coordinates for the eight fields in this thesis are provided in Table 2.2. The ‘E’ fields overlap the ELAIS-S1 field (Gruppioni et al. 1999), the ‘X’ fields overlap the *XMM–Newton* Large-Scale Structure field (XMM-LSS; Pierre et al. 2004), and the ‘C’ fields overlap the *Chandra* Deep Field South (CDF-S; Xue et al. 2011). Since each supernova field extends over approximately 3 deg^2 (the field-of-view of a single pointing), the area coverage of the optical data available for this thesis totals around 24 deg^2 .

DES DEEP (SUPERNOVA) FIELDS				
DES name	Field	R.A. (J2000)	DEC (J2000)	Type
E1	ELAIS-S1	00:31:29.86	-43:00:34.6	shallow
E2	ELAIS-S1	00:38:00.00	-43:59:52.8	shallow
X1	XMM-LSS	02:17:54.17	-04:55:46.2	shallow
X2	XMM-LSS	02:22:39.48	-06:24:43.6	shallow
X3	XMM-LSS	02:25:48.00	-04:36:00.0	deep
C1	CDF-S	03:37:05.83	-27:06:41.8	shallow
C2	CDF-S	03:37:05.83	-29:05:18.2	shallow
C3	CDF-S	03:30:35.62	-28:06:00.0	deep

Table 2.2 Central coordinates of the eight 3 deg^2 DES deep fields, also referred to as supernova fields. The two so-called deep supernova fields X3 and C3 contain more exposures and are $\sim 1 \text{ mag AB}$ deeper than the shallow fields.

It is important to point out that this thesis exclusively uses of data from the first year of observations (see Diehl et al. 2014). This dataset was released internally to the DES collaboration as **Y1A1** (Drlica-Wagner et al. 2018) and is a subset of the data published in the first public data release **DR1** (which covers all the first three years; Abbott et al. 2018b). The reason for including only the **Y1A1** data is that at the time the work in this thesis was carried out, data from later years was not yet available due to processing delays. Fortunately, the DES+VIDEO catalogue production pipeline presented in this thesis largely uses automated processes, so it would be straightforward to extend this work to incorporate data from any later releases.

2.1.1.3 Data reduction

After the Year 1 exposures had been taken, the data was processed and combined into deep co-added images. Before this could occur, all DES exposures were required to pass preliminary quality checks on the point spread function (PSF)

and signal-to-noise ratio, detailed in Drlica-Wagner et al. (2018). The exposures that passed these quality control standards were then processed by the DES data management system (DESDM; Sevilla et al. 2011; Mohr et al. 2012; Morganson et al. 2018), with the aim of producing science ready imaging. The details of this procedure are described at length in Drlica-Wagner et al. (2018) for the **Y1A1** data used in this thesis. During data processing, the DESDM pipeline removed instrumental signatures from the telescope and DECam setup, and performed masking of artefacts including bright stars, cosmic rays, and fast-moving transients such as satellites and meteors. The DESDM algorithm also created weight images. These weight planes were generated for each individual exposure based on the inverse variance of a flat field image, as this variance captures the Poisson noise and read noise of the instrument. Weight values were then set to zero for all saturated pixels in bright stars, and for masked pixels from artefacts and transients.

The processed exposures then underwent a second round of quality control cuts, intended to blacklist and remove individual CCD images with severe imaging artefacts from phenomena such as scattered light from bright stars, or airplane and meteor trails. For **Y1A1**, only 1% of CCD images were discarded in this way, and the details of this procedure can once again be found in Drlica-Wagner et al. (2018). DESDM then produced image co-adds of all approved images, obtained from the weighted average of all overlapping images (using the weight maps described above). The co-addition was performed using **SWarp** (Bertin et al. 2002; Bertin 2010), which remapped all pixels onto a uniform grid with a pixel scale of $0.263 \text{ arcsec pix}^{-1}$. This grid was then cut up into co-add tiles of $10\,000 \times 10\,000$ pixels (covering $0.73 \text{ deg} \times 0.73 \text{ deg}$), as illustrated in the inlay of Figure 2.3. Tiles overlap at the edges by about 1 arcmin, so that for every observed source there exists a tile where it is at least 0.5 arcmin from the edge. One co-added tile was produced for each *grizY* filter, in addition to a **CHI_MEAN** combination of the $r + i + z$ co-adds intended as a detection image for extracting photometry (Szalay et al. 1999; Bertin 2010; Drlica-Wagner et al. 2018). The *g*-band was omitted from this combined image due to the large PSF and noisy sky background in this filter, and the *Y*-band was rejected because of its limited depth and irregular PSF (Morganson et al. 2018). In addition to these science co-adds, co-added weight images for all five filters and the combined detection image were produced from the individual exposure weight plane images described

above. Conceptually, the idea behind these co-add weight images is to assign a higher weight to pixels with more exposures, and a lower weight in areas where the data is poor.

2.1.1.4 Catalogues

As part of the data reduction pipeline, DESDM also generated source catalogues from the DES co-adds. The philosophy at the core of this assembly was to create the most inclusive selection of objects, while maintaining low contamination rates (Drlica-Wagner et al. 2018). Such a wide selection of sources suits the aims of this thesis nicely, as the eventual selection of high-redshift objects in Chapter 4 contains its own specific cuts to narrow down the objects. Furthermore, because of the extensive testing that the DES collaboration has applied to the DESDM pipeline, the resulting data products can be trusted to adhere to high quality standards. For these reasons, this thesis has adopted the DESDM catalogues for its optical photometry.

These DESDM catalogues were created with **SExtractor** (Bertin & Arnouts 1996), according to a procedure detailed in Drlica-Wagner et al. (2018). Objects were detected in the aforementioned combined $r + i + z$ image, and photometry was extracted in each $grizY$ filter based on the detected positions in the combined image. This technique of using a different image for detection and source extraction is commonly referred to as ‘forced photometry’, and will be revisited in Section 2.2.

Despite the great diligence with which the DES collaboration has optimised photometric calibration, there nevertheless exist some colour non-uniformities in the DESDM images and catalogues used as the optical backbone for this thesis (Drlica-Wagner et al. 2018). These discrepancies are due to position-dependent variations in the photometric calibration, together with varying levels of Galactic reddening over the survey footprint. The DES collaboration has resolved these issues by determining magnitude corrections, derived via the so-called stellar locus regression (SLR) method (Ivezić et al. 2004; High et al. 2009; Desai et al. 2012; Kelly et al. 2014). Using a code by Kelly et al. (2014), Drlica-Wagner et al. (2018) employed this technique to calibrate the proper photometric zero-points from the particular shape of the stellar locus in colour-colour space. Using the DES colours of star samples in $\sim 0.1 \text{ deg}^2$ segments, the appropriate SLR correction was determined empirically for each segment of the total footprint. For

the data in this thesis, these corrections are later applied by the author at the catalogue level, as will be described in Section 2.2.5.

2.1.1.5 Data quality

In order to provide an overview of the quality of the optical data in this thesis, this section will outline several important properties of the DES deep field data, specifically the seeing, depths, and weight flags.

Seeing

The DES observing protocol dictated that the supernova fields had to be observed (1) when the seeing was above 1.1 arcsec and thus deemed inadequate for the precision cosmology requirements of the wide-field survey, or (2) when a particular supernova field had not been observed for 7 days. As a result, the seeing in the DES deep fields varies considerably. To eliminate any data with a PSF that is too poor for many scientific applications, the preliminary quality checks described in Section 2.1.1.3 imposed a quality cut of $\text{FWHM} < 2 \text{ arcsec}$ on any exposures that were included in the co-adds. All in all, these various constraints on the seeing have resulted in an average PSF width of $\sim 1.0 \text{ arcsec}$ in the DES deep fields (Kessler et al. 2015).

Depths

The observing strategy described in Section 2.1.1.2 created inhomogeneous coverage over the DES footprint. Because each field contains regions with different numbers of exposures, the survey depths are non-uniform both across and inside tiles. The heterogeneous nature of the data is increased further by the effect of dead CCDs, chips gaps, and masked objects (e.g. stars and satellite or airplane streaks). With this in mind, the DES collaboration has provided a rigorous analysis of the coverage and depth variation, by creating depth maps that chart the limiting magnitude as a function of sky position. The paper by Drlica-Wagner et al. (2018) describes the creation of these maps in detail. In essence, their approach used the `mangle` code (Hamilton & Tegmark 2004; Swanson et al. 2008) to capture the DES survey coverage, firstly masking out bad regions and dividing the DES data into segments. For each segment, the 10σ limiting magnitude in a 1.95 arcsec aperture in a given filter (i.e. $m_{10\sigma}$) was then calculated based

on the **mangle** co-add weight maps (created from the DESDM weight images¹ mentioned in Section 2.1.1.3). These calculations used the following equation:

$$m_{10\sigma} = m_{\text{zp}} - 2.5 \log 10 \sqrt{\pi \frac{(D/2)^2}{\omega_{\text{pix}}^2} \frac{1}{w_{\text{tot}}}}, \quad (2.1)$$

where $m_{\text{zp}} = 30$ is the photometric zero-point for the DES tiles, $D = 1.95$ arcsec is the aperture diameter, $\omega_{\text{pix}} = 0.263$ arcsec pix⁻¹ is the pixel scale and w_{tot} is the total co-added weight in a particular **mangle** segment. The resulting map of 10σ limiting magnitudes was then pixelised and stored as a **HEALpix** map, which was distributed to the DES collaboration.

The depths from this map are later added to the DES+VIDEO catalogue by the author, together with depths for other values of σ (as will be detailed in Section 2.2.5). The procedure used to calculate these other depths is as follows. Following Rykoff et al. (2015), the 10σ limiting magnitudes in the **HEALpix** map are firstly converted to other depths via the effective exposure time t_{eff} , which for each filter can be parameterised as a function of limiting magnitude:

$$t_{\text{eff}} = \exp(a + b(m_{10\sigma} - 21.0)). \quad (2.2)$$

In this formalism $m_{10\sigma}$ is the 10σ limiting magnitude, 21.0 is a convenient pivot point, and a and b are parameters that were provided by the DES collaboration (obtained by fitting to the DES data). The effective exposure time is related to the flux noise F_{noise} via

$$F_{\text{noise}} = \frac{F_{10\sigma}^2 t_{\text{eff}}}{10^2} - F_{10\sigma}, \quad (2.3)$$

and the 10σ limiting flux $F_{10\sigma}$ required to compute this flux noise can be obtained straightforwardly from the aforementioned 10σ limiting magnitudes:

$$F_{10\sigma} = 10^{(m_{10\sigma} - m_{\text{zp}})/2.5}. \quad (2.4)$$

¹To be precise, **mangle** produced its own co-add weight maps from the individual exposure weights rather than the final co-add weight maps.

Here $m_{\text{zp}} = 30$ once more represents the photometric zero-point². Combining Equations 2.2, 2.3, and 2.4 yields F_{noise} , which can then be translated into limiting fluxes $F_{N\sigma}$ for other values of σ denoted by N (in this formalism $N = 5$ yields 5σ depths):

$$F_{N\sigma} = \frac{N^2 + \sqrt{(N^2)^2 + 4t_{\text{eff}}N^2F_{\text{noise}}}}{2t_{\text{eff}}}. \quad (2.5)$$

Finally, these limiting fluxes are converted back to magnitudes to obtain the final limiting magnitude $m_{N\sigma}$ for an arbitrary number N :

$$m_{N\sigma} = m_{\text{zp}} - 2.5 \log F_{N\sigma}. \quad (2.6)$$

Via this method, 7.5σ , 5σ , 3σ , and 1σ depths can be computed for every DES filter. Section 2.2.5 will describe how these values are eventually added to the DES+VIDEO catalogue, so that the final data product contains $m_{10\sigma}$, $m_{7.5\sigma}$, $m_{5\sigma}$, $m_{3\sigma}$, $m_{2\sigma}$ and $m_{1\sigma}$ limiting magnitudes in the *grizY*-bands for each object.

To visualise the depths in the eventual DES+VIDEO catalogue, the *z*-band 5σ depth maps (obtained from the 10σ DES maps as described above) are plotted in Figures 2.4, 2.5 and 2.6 for the ELAIS-S1, XMM-LSS, and CDF-S fields respectively. It must be noted that these maps contain only the DES data included in the combined DES+VIDEO catalogue, and that they thus display only regions that overlap the VIDEO survey. The plots clearly show the heterogenous nature of the DES data, as well as the imprint of the hexagonal DECam CCD array and chip gaps in between CCDs. They also illustrate the varying coverage in the DES supernova fields, clearly demonstrating the higher depths in the deepest regions X3 and C3 (see Table 2.2). An overview of the depth variation in all DES filters is given in Table 2.3, which shows the minimum, maximum and median values of the 5σ depths for the full *grizY* filterset.

²The DES collaboration wiki pages mention that this zeropoint ought to be 22.5. However, the author believes that this was likely a mistake, since the paper by Drlica-Wagner et al. (2018) clearly states that the tile zero-point value used in the $m_{\text{lim},10\sigma}$ computations is 30.0. The reason why the wiki pages list a value of 22.5 is likely that the equations on these pages were copied directly from Rykoff et al. (2015). That paper was written based on SDSS data, where the zero-point is indeed 22.5. The author has calculated the difference in depths between either of these zero-points, and fortunately determined that they were of the order of ~ 0.02 mag, so even if the assumption of $m_{\text{zp}} = 30.0$ in this thesis is in fact not correct, the effect of this choice is very small.

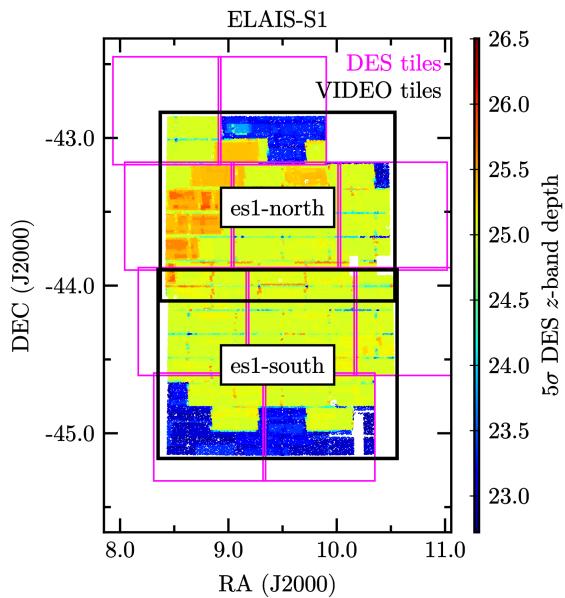


Figure 2.4 z -band 5σ limiting magnitudes in 1.95 arcsec apertures for the DES data that overlaps the VIDEO footprint in the ELAIS-S1 field. These depths have been derived from the 10σ depth maps provided by the DES collaboration as detailed in the text. The DES collaboration has included the SLR corrections described in Section 2.1.1.4 in their 10σ depth estimates. The magenta squares show the edges of the DES co-add tiles. VIDEO tiles are outlined in black.

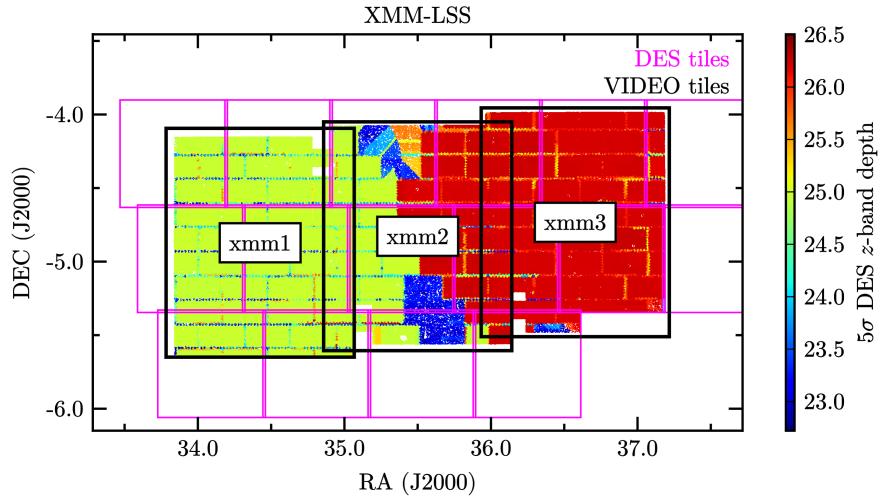


Figure 2.5 z -band 5σ limiting magnitudes in 1.95 arcsec apertures for the XMM-LSS field. The image description is otherwise identical to that of Figure 2.4. The red regions correspond to the deep supernova field X3.

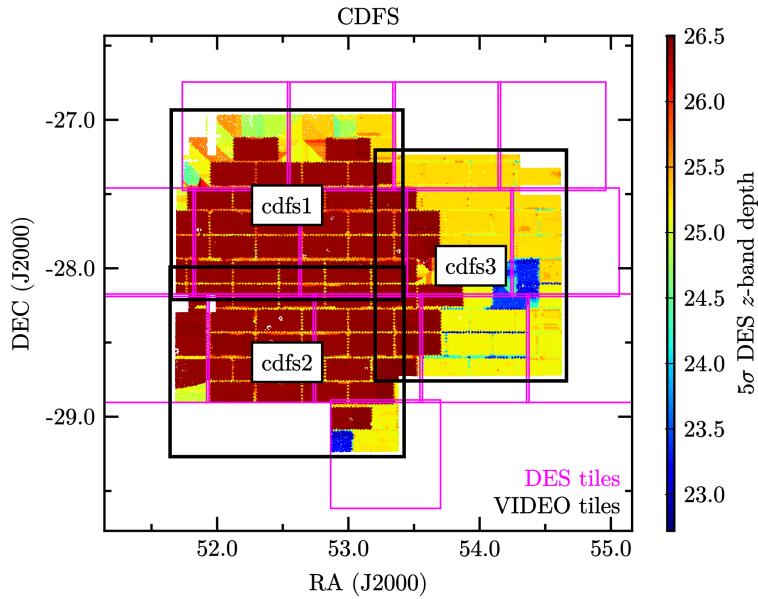


Figure 2.6 z -band 5σ limiting magnitudes in 1.95 arcsec apertures for the CDF-S field. The image description is otherwise identical to that of Figure 2.4. The red regions correspond to the deep supernova field C3.

5 σ LIMITING MAGNITUDES IN THE DES DEEP FIELDS		
Filter	5 σ depth range (mag AB)	Median 5 σ depth (mag AB)
<i>g</i>	24.2 - 26.9	25.9
<i>r</i>	23.8 - 26.9	25.6
<i>i</i>	23.3 - 26.6	25.4
<i>z</i>	22.7 - 26.5	25.2
<i>Y</i>	21.2 - 23.5	23.1

Table 2.3 Minimum, maximum, and median (averaged over area) values of the 5 σ limiting magnitudes in 1.95 arcsec apertures, for the DES deep field data overlapping the VIDEO footprint. Values were obtained from the 5 σ depth maps derived as per the text, based on the 10 σ limiting magnitude maps provided by the DES collaboration. The significantly lower *Y*-band depths are explained by the fact that *Y*-band coverage is provided by the wide-area survey, which is considerably shallower.

To measure the effective area available for specific science goals, it is useful to consider the total DES+VIDEO area that is available to a given depth. This quantity is explored in Figure 2.7, which for each of the *grizY* filters shows the relationship between a given 5 σ limiting magnitude and the total area with that limiting magnitude or higher. Out of the total combined footprint of 12.1 deg², the largest part (\approx 11 deg² in *griz*) comes from the supernova fields. The remaining coverage (\approx 1 deg² of the *griz*-bands and all of the *Y*-band) is from the wide-field survey. The effect of the shallow and deep supernova fields (see Table 2.2) can be seen in the two ‘steps’ of the *griz* distributions. In the *z*-band, for instance, there is 10.8 deg² available to 5 σ depths of $z_{5\sigma} = 24.9$, originating from all eight supernova fields (deep and shallow). A sub-area of 4.3 deg² extends even further to $z_{5\sigma} = 26.2$; this data comes from the deep supernova fields X3 and C3.

Although the 10.8 deg² area with $z_{5\sigma} = 24.9$ that remains after considering VIDEO overlap is much less than the total ~ 24 deg² of DES data available, a comparison with similar surveys in Table 1.2 demonstrates that it is still among the largest footprints to these depths. What is more, the $z_{5\sigma} = 26.2$ limiting magnitudes in the deepest 4.3 deg² regions are comparable to those of the CFHTLS deep fields, which currently comprise the deepest optical imaging over ~ 4 deg². The area-depth coverage of the catalogue in this thesis turns out to be even more significant when also considering the VIDEO limiting magnitudes, because together the DES+VIDEO combined footprint covers the largest area among existing optical+near-IR datasets with comparable near-infrared depths. This point

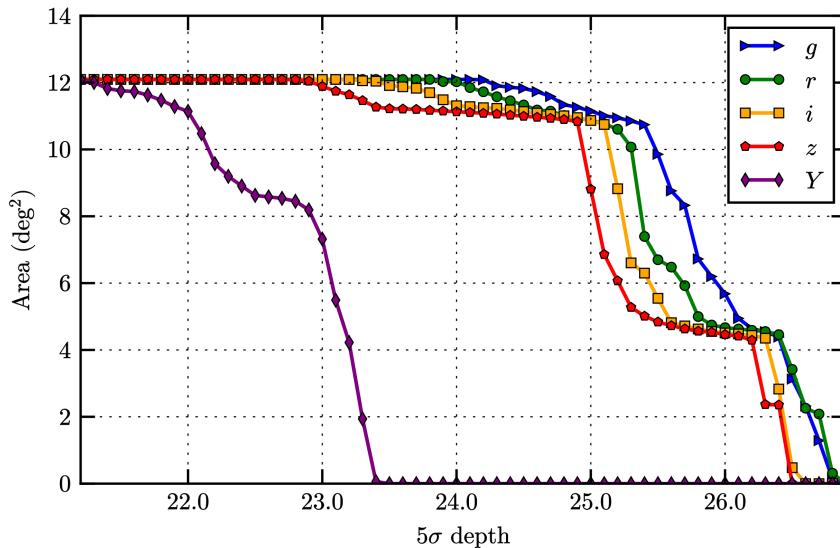


Figure 2.7 The area available to a given 5σ depth as a function of that depth, for the catalogue produced in this thesis. The plotted area contains only DES data overlapping VIDEO. Area-depth functions are shown for all five DES filters.

will be revisited in Section 2.1.2.4, where the VIDEO depths will be introduced and discussed.

Weight flags

To identify sources with questionable photometry, the DESDM catalogues contain a `FLAGS_WEIGHT` column that indicates when `SExtractor` extracted photometry on an object with low weights (as determined from the $r + i + z$ co-add weight maps). `SExtractor` assigns a `FLAGS_WEIGHT` value of 1 to objects which overlap pixels in the weight map below a threshold value `WEIGHT_THRESH`, and a value of 2 to objects that are adjacent to pixels below the `WEIGHT_THRESH` value³. All other, unflagged sources are allocated a `FLAGS_WEIGHT` value of 0. The DESDM pipeline used a `WEIGHT_THRESH` threshold of 0 (Morganson et al. 2018), which means in practice that objects with non-zero weight flags overlap or touch regions that were masked out by DESDM. As discussed in Section 2.1.1.3,

³More specifically, when the isophotal footprint assigned to an object by `SExtractor` overlaps or is adjacent to pixels which fall below this threshold in the detection image weight map.

these masked out regions belong to bad pixels or artefacts, for instance from faulty CCDs, chip gaps, saturated parts of bright stars, cosmic rays, and satellite or meteor trails (see Drlica-Wagner et al. 2018 and Morganson et al. 2018 for the specific details of the masking algorithm). In this thesis, the weight flags are later used to filter out artefacts and objects with unreliable photometry, during the computation of DES+VIDEO photometric redshifts and the assembly of a high-redshift galaxy sample (described in Sections 3.5.1 and 4.2.3.1 respectively).

2.1.2 The Vista Extragalactic Deep Observations Survey (VIDEO)

2.1.2.1 Overview

The VISTA Extragalactic Deep Observations survey (VIDEO; Jarvis et al. 2013) is a near-infrared survey conducted with the Visible and Infrared Survey Telescope for Astronomy (VISTA). Its scientific aims are to study the evolution of galaxies, clusters, and structures as a function of their environment over the cosmic epoch $0 < z < 4$, and to trace AGN and the most massive galaxies out into the era of reionisation. To this end, VIDEO covers 12 deg^2 in five near-infrared filters ($ZYJHK_s$). Imaging over these wavelengths is crucial to both science goals. Firstly, near-IR data is required for the proposed $z < 4$ galaxy evolution studies, because above $z > 1$ the bulk of the luminosity output required for stellar mass estimates is emitted in these wavebands. Secondly, when approaching the epoch of reionisation at $z \sim 6$, the light output of galaxies is almost entirely redshifted into the near-infrared. The VIDEO science goals also require deep imaging over a large area. Previous near-infrared studies have been either too shallow for high-redshift studies (e.g. VIKING, Edge et al. 2013; UKIDSS-LAS, Lawrence et al. 2007; VHS, McMahon et al. 2013), or covered too small areas for statistical studies of large-scale environments and for finding uncommon objects such as massive galaxies and AGN (e.g. CANDELS, see Table 1.1; UltraVISTA, see Table 1.2). The VIDEO survey bridges this gap, utilizing the state-of-the-art rapid imaging technology on the VISTA telescope that has made it possible to efficiently probe large cosmological volumes. Its combination of area and depth makes VIDEO an ideal candidate to combine with DES optical imaging, creating a dataset excellently suited to the photometric redshift research and high-redshift galaxy search in this thesis.

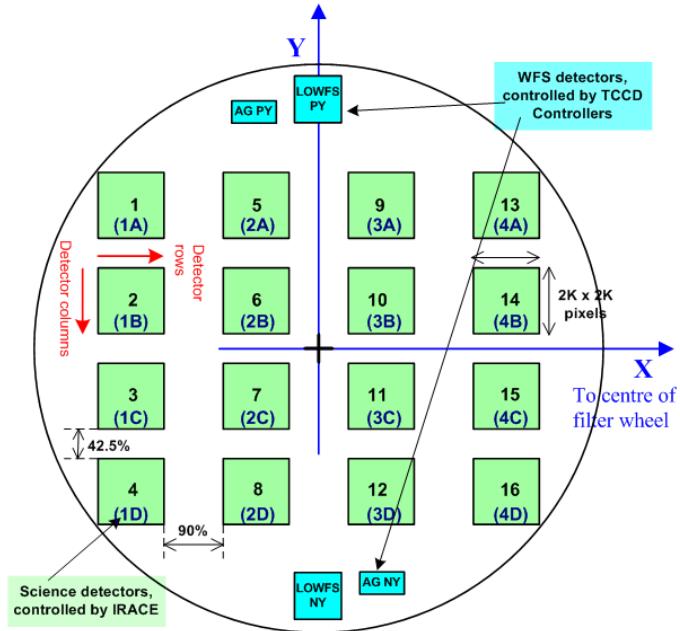


Figure 2.8 Focal plane of the VISTA Infrared Camera (VIRCAM), made up of 16 HgCdTe detectors. The detectors are spaced out with gaps of 0.9 and 0.425 of a detector in the x- and y-direction respectively. The diameter of the focal plane spans 1.65 deg, and the total effective area covered by all 16 detectors (excluding gaps) is 0.6 deg². *Image credit:* VISTA (www.vista.ac.uk)

2.1.2.2 Data collection

VIDEO observations have been conducted with the VISTA Infrared Camera (VIRCAM; Dalton et al. 2006) on the 4m VISTA telescope (Emerson et al. 2004; Sutherland et al. 2015) at the La Silla Paranal observatory. VIRCAM is the world’s largest near-IR camera, made up of 16 2048 × 2048 HgCdTe detector arrays totalling 67 megapixel, with a mean pixel scale of 0.339 arcsec pix⁻¹. The detectors are spaced out over a 1.65 deg diameter focal plane, with gaps of 0.9 and 0.425 of a detector along the x- and y-axes respectively, as illustrated by Figure 2.8. Excluding these gaps, the 16 detectors cover an effective 0.6 deg² field-of-view. The five ZYJHK_s broadband VIRCAM filters used for VIDEO are plotted in Figure 2.2, with detailed wavelength coverage specified in Table 2.1.

Observations commenced in October 2009 and finished in September 2018. Throughout these nine years, imaging was collected using an observing strategy detailed in Jarvis et al. (2013), which will be now be summarised briefly. The

VIDEO protocol was designed to deal with well-known challenges for near-infrared astronomy, concerning the brightness and variability of the sky background in this wavelength regime (particularly in the H - and K_s -bands). To overcome these issues, individual exposures were split up into small time chunks. Each pointing was observed NDIT times, with a detector integration time DIT (between 10 and 50 seconds depending on the filter), resulting in a total exposure time of $NDIT \times DIT$. The telescope was then ‘jittered’ randomly by a small offset of $\lesssim 20$ arcsec, and the $NDIT \times DIT$ sequence was repeated at the offset position. This jittering minimises the effect of both sky fluctuations and detector artefacts such as bad pixels. The totality of $NDIT \times DIT$ (jittered) exposures in a given position is called a ‘paw-print’ in VIDEO jargon. Due to the aforementioned gaps between VIRCAM detectors, these paw-prints do not fill a contiguous area. To cover a full tile, the telescope was moved in a sequence of six offsets in the x- and y-direction, with paw-prints taken in each of these positions. When added together, they produce a VIDEO tile with a total area of $\sim 1 \times 1.5 \text{ deg}^2$.

VIDEO FIELDS				
VIDEO name	Field	R.A. (J2000)	DEC (J2000)	Available filters
es1-north	ELAIS-S1	00:37:49	-43:29:53	ZYJHK _s
es1-south	ELAIS-S1	00:37:49	-44:29:53	ZYJHK _s
xmm1	XMM-LSS	02:17:42	-04:51:00	ZYHK _s
xmm2	XMM-LSS	02:22:00	-04:48:00	YJHK _s
xmm3	XMM-LSS	02:26:18	-04:44:00	ZYJHK _s
cdfs1	CDF-S	03:30:08	-27:34:30	ZYJHK _s
cdfs2	CDF-S	03:30:08	-28:38:00	JHK _s
cdfs3	CDF-S	03:34:44	-27:59:00	JK _s

Table 2.4 Central coordinates for all eight $\sim 1 \times 1.5 \text{ deg}^2$ VIDEO tiles. The positioning of these tiles, as well as the overlap with DES data, is shown in Figures 2.4, 2.5 and 2.6 for the ELAIS-S1, XMM-LSS and CDF-S respectively. The filters listed under ‘available filters’ are all the VIDEO wavebands that were observed in the DR5 data release used in this thesis. The number of filters differs across tiles, because the survey had not yet been completed in this release.

The VIDEO survey targeted three fields, which mostly overlap with the DES deep fields introduced in Section 2.1.1.2 and Table 2.2. The ELAIS-S1 field contains two VIDEO tiles extending over 3 deg^2 , and the XMM-LSS and CDF-S fields each encompass three VIDEO tiles totalling 4.5 deg^2 for each field. These eight tiles make up the total $\sim 12 \text{ deg}^2$ VIDEO footprint. The central coordinates of the VIDEO tiles are given in Table 2.4, and their placement is mapped out in Figures

2.4, 2.5 and 2.6, which also illustrate the intersection with the DES data. As mentioned previously in Section 2.1.1.5, the total overlap area comprises 12.1 deg^2 , with 10.8 deg^2 covered by deep field DES observations with limiting magnitudes $z_{5\sigma} \geq 24.9$.

This thesis uses imaging from the fifth VIDEO data release (DR5), which is the most recent publicly available dataset at the Edinburgh VISTA Science Archive⁴. This dataset includes observations taken up to 25 March 2015, and is identified by the date stamp **2015-03-25** internally to the VIDEO collaboration. As the survey was not yet completed at this time, some tiles in this release do not contain observations in every filter. Table 2.4 lists the filters that are available in each tile. Some tiles and filters have also progressed further than others in terms of exposure times, resulting in some depth variation between tiles. The extent of this variation will be presented in Section 2.1.2.4. Regarding future data releases, it is useful to mention once more that the DES+VIDEO catalogue production pipeline created in this thesis is largely automated, so that it will be straightforward to include future VIDEO data.

2.1.2.3 Data reduction

Data reduction for the VIDEO survey was performed by the Cambridge Astronomical Survey Unit (CASU) and Jarvis et al. (2013). A summary of the procedure is provided below, and the reader is referred to Jarvis et al. (2013) for the full details. The initial data processing steps were conducted by CASU using a pipeline specially developed for VIRCAM data, detailed in Irwin et al. (2004) and updated on the CASU website⁵. The CASU system firstly processed each NDIT \times DIT raw data frame, removing instrument signatures, performing photometric calibration, applying sky background subtraction⁶, and producing confidence (i.e. weight) maps. CASU then created paw-prints by stacking all jittered iterations of the processed raw data frames, applying outlier rejection to artefacts such as bad pixels, cosmic rays, and fast-moving transients. Stacked weight maps for each paw-print were created at the same time. Afterwards, Jarvis et al. (2013) combined these paw-prints into co-added tiles and co-added weight maps, using

⁴<http://horus.roe.ac.uk/vsa/>

⁵<http://casu.ast.cam.ac.uk/surveys-projects/vista/technical/data-processing>

⁶The background correction for VIDEO differs from the default CASU pipeline. As explained in Jarvis et al. (2013), the VIDEO depths require a special fine-tuned sky subtraction procedure.

SWarp to stack images via a weighted mean algorithm. Paw-prints with a seeing worse than $\text{FWHM} < 0.9 \text{ arcsec}$ were rejected before co-addition. During the co-addition process, all images were resampled to a pixel scale of $0.2 \text{ arcsec pix}^{-1}$. The resulting VIDEO images and weight maps make up the imaging from which the near-infrared photometry for this thesis is extracted.

Catalogues

Jarvis et al. (2013) also produced VIDEO catalogues from the co-add images, assembled via several iterations of **SExtractor**. In each run, Jarvis et al. (2013) used dual-image mode to extract photometry in a given band, with detection performed in a different band for each iteration (see Section 2.2.2.2 for an illustrative application of double-image mode). As an example, they created four catalogues for the Z -band, each of which was based on detections in one of the Y , J , H or K_s filters. They then merged these **SExtractor** iterations into a final catalogue for each tile, retaining only the longest wavelength detection for sources detected in multiple bands. Lastly, Jarvis et al. (2013) corrected their photometric errors in order to account for correlated noise. More detail on all these catalogues can be found in Jarvis et al. (2013).

It is emphasised that this thesis does not make use of the Jarvis et al. (2013) photometry for the DES+VIDEO catalogue. Instead, VIDEO fluxes in this thesis are extracted straight from the VIDEO co-adds via a forced photometry method that will be introduced in Section 2.2. The Jarvis et al. (2013) catalogues⁷ are briefly used afterwards to verify the forced photometry results. In this way, they do play a role in assessing the validity of the DES+VIDEO catalogue.

2.1.2.4 Data quality

It is useful to examine the quality of the VIDEO imaging. Just like the DES quality assessment in Section 2.1.1.5, this analysis will once again focus on the seeing, depths, and weight flags.

⁷Throughout this thesis, these datasets will be referred to as the Jarvis et al. (2013) catalogues, although the version used in this thesis is an updated version from the DR5 data release based on data up to 2015.

Seeing

The VIDEO observing protocol required that observations be carried out in good seeing – for each filter it was required that $\text{FWHM} < 0.8 \text{ arcsec}$. However, because observing conditions could change over an observation period, exposures with seeing up to $\text{FWHM} < 0.9 \text{ arcsec}$ were also included in the co-add tiles. Over the total survey, the mean seeing is $(0.84, 0.81, 0.79, 0.78 \text{ and } 0.78) \text{ arcsec}$ for the $Z, Y, J, H,$ and K_s filters respectively⁸. The smallest seeing of 0.72 arcsec was achieved in the cdfs2 H -band, and the largest value of 0.85 arcsec in the cdfs1 Z -band.

Depths

The VIDEO observing strategy detailed in Section 2.1.2.2 produced reasonably uniform coverage inside each tile. Unlike the DES imaging, the VIDEO tiles do not contain any chip gaps. Nevertheless, some sections in the tiles are slightly deeper than average. These regions are made up of an above average number of exposures, as a result of detector overlap between paw-print shifts. At the same time, some other areas contain below average depths. This is particularly the case in regions with poor detector performance, as well as in two $\sim 0.1 \text{ deg}$ strips at the edges which have only been sampled once (by a single paw-print). The depth variation that results from all these effects is captured by the co-add weight maps. An example weight map for the K_s -band in the xmm3 tile is shown in Figure 2.9.

In spite of the small divergent regions described above and in the caption to Figure 2.9, the overall coverage within a VIDEO tile is fairly uniform. However, even though the tiles themselves are much more uniform than the DES imaging, the VIDEO survey still contains some depth variation between tiles. This is due to the fact that the survey was not yet completed for the data release used in this thesis, so not all tiles had yet been allocated the same exposure time. To explore the variation of VIDEO depths in the DES+VIDEO catalogue, the author computed the 5σ limiting magnitudes in a 1.95 arcsec aperture for each tile. The procedure for calculating these values is as follows. Firstly, the 5σ limiting fluxes

⁸These values were derived by the author based on VIDEO collaboration measurements of the average seeing in each VIDEO tile. The mean was then obtained by averaging these tiles.

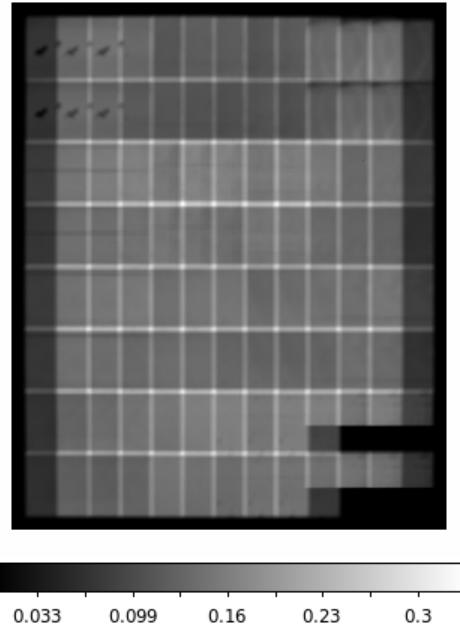


Figure 2.9 Weight map for the K_s -band in the xmm3 tile. Weight values are displayed on a linear scale indicated by the bar. The two black regions in the bottom right corner are due to time-varying quantum efficiency in detector 16. Because flat fielding in that detector is extremely challenging, data from those regions has been assigned zero weight. In the top right a small region has lower weights due to dead pixels in detector 1.

$F_{5\sigma}$ are derived. Because $F_{5\sigma}$ is simply defined as the flux for which the signal-to-noise ratio (S/N) equals 5, the following equation holds:

$$\text{S/N} \equiv \frac{F_{5\sigma}}{F_{\text{error}}} = 5, \quad (2.7)$$

where F_{error} is the VIDEO flux error in a 1.95 arcsec aperture. It follows that F_{error} can therefore be used to obtain the corresponding limiting flux $F_{5\sigma}$. Section 2.2.4.3 will describe the way those uncertainties are computed for the DES+VIDEO catalogue in this thesis. The method involves an approximation whereby flux errors in a given aperture are assigned a constant value throughout each VIDEO tile. Because this approximation entails that F_{error} is constant, for each tile the limiting fluxes can be computed via a single instance of Equation 2.7. The limiting fluxes are then converted into magnitudes to obtain the final depths in each tile:

$$m_{5\sigma} = m_{\text{zp}} - 2.5 \log F_{5\sigma} = m_{\text{zp}} - 2.5 \log (5F_{\text{error}}), \quad (2.8)$$

where $m_{\text{zp}} = -48.6$ is the AB zero-point. Table 2.5 shows the range in VIDEO depths derived in this way, listing the magnitude in the shallowest and deepest tiles for each filter.

VIDEO 5σ LIMITING MAGNITUDES	
Filter	5σ Depth (mag AB)
Z	24.8 - 25.3
Y	24.6 - 24.9
J	23.9 - 24.6
H	23.3 - 24.1
K_s	22.9 - 23.8

Table 2.5 Minimum and maximum VIDEO 5σ depths in a 1.95 arcsec aperture in the DES+VIDEO catalogue. These limiting magnitudes were estimated from the catalogue photometric errors as described in the text. Depths were approximated to be constant across tiles, and the range of depths observed originates from variation between tiles. Depth variation occurs because in the data release used in this thesis the full survey depth has not been achieved in all tiles.

Weight flags

Weight flags for the VIDEO portion of the DES+VIDEO catalogue are produced by **SExtractor** during the extraction of forced photometry (see Section 2.2.2.2). As with the DESDM catalogues, non-zero weight flags identify objects that contain or touch pixels with zero weight in the co-add weight maps. For the VIDEO data, these are pixels at the very edge of the imaging and in the regions covered by detector 16 (see the bottom right of Figure 2.9). In the same way as for DES, this thesis uses the VIDEO weight flags to filter out artefacts and objects with unreliable photometry, during the computation of photometric redshift and the selection of high-redshift galaxies (see Sections 3.5.1 and 4.2.3.1 respectively).

2.1.3 Spectroscopic redshifts

To complement the broadband imaging and catalogues, the DES collaboration has compiled a master spectroscopic catalogue of publicly available galaxy spec-

tra that overlap the DES footprint, published in the DES Science Portal database⁹ (Fausti Neto et al. 2018). This thesis uses the final **Y1A1** version (from 25 January 2016), with DESDM table name **brportal.e_10022146_295159**.

The master catalogue includes 759 535 galaxy spectra from 31 public surveys in the literature, and contains good quality flags for all objects. However, a quick investigation of the spectra revealed that some galaxies from the EBOSS survey¹⁰ have unreliable redshifts. These objects have been assigned spectroscopic redshifts of $z_{\text{spec}} > 3.0$, while the DES imaging clearly indicates that they are bright ($z_{\text{AB}} \lesssim 20$) resolved galaxies, implying they cannot be at such high redshifts. For this reason, it was decided to remove all 335 EBOSS objects with $z_{\text{spec}} > 3.0$ from the master catalogue, so that 759 200 galaxies remain. These objects are later matched to the DES+VIDEO sources by position (within a 1 arcsec radius), via a process described in Section 3.5.1. This matching strategy also removes any spectra with bad **FLAGS_WEIGHT** values in any DES+VIDEO filter. At the end, these steps produced a final sample of 35 596 galaxy spectra with DES+VIDEO photometry. The 21 spectroscopic surveys from which these spectra have been sourced are listed in Table 2.6, together with the fractional contribution of each survey to the total number of DES+VIDEO spectroscopic redshifts.

SOURCES OF SPECTROSCOPIC REDSHIFTS			
Survey	% of total redshifts	Survey	% of total redshifts
3DHST	18.9 %	SNLS_FORS	1.3 %
DES_AAOMEGA	16.5 %	SPARCS	0.8 %
VVDS	12.8 %	VUDS	0.3 %
ACES	11.9 %	2DF	0.3 %
GAMA	9.6 %	SNLS_AAO	0.3 %
EBOSS_DES_ELG	9.2 %	STALIN	0.2 %
NOAO_OZDES	8.5 %	CDB	0.2 %
UDS	3.5 %	6DF	< 0.1 %
PANSTARRS	2.4 %	MOSFIRE	< 0.1 %
ATLAS	1.6 %	LCRS	< 0.1 %
SDSS_OZDES	1.5 %		

Table 2.6 Publicly available spectroscopic redshift surveys that make up the spectra in the DES+VIDEO catalogue. Spectra have been sourced from the final **Y1A1** catalogue described in the text. The names in the ‘Survey’ column correspond to the **SOURCE** tag in this table.

⁹<http://cdcvn.fnal.gov/redmine/projects/des/wiki/SciencePortalProducts>

¹⁰The source name for this survey is **DES_EBOSS_ELG** in the master catalogue.

Figure 2.10 shows the redshift distribution of the DES+VIDEO spectra, for both the entire dataset and the ELAIS-S1, XMM-LSS, and CDF-S fields individually. With regard to the high-redshift galaxy search later in this thesis, it is important to note that the vast majority of spectra are concentrated at $z_{\text{spec}} < 1.0$, with only 11 at $z_{\text{spec}} > 4.0$ and none at $z_{\text{spec}} > 5.0$. The absence of available spectra at high redshifts impacts the photometric redshift strategy in Chapter 3.

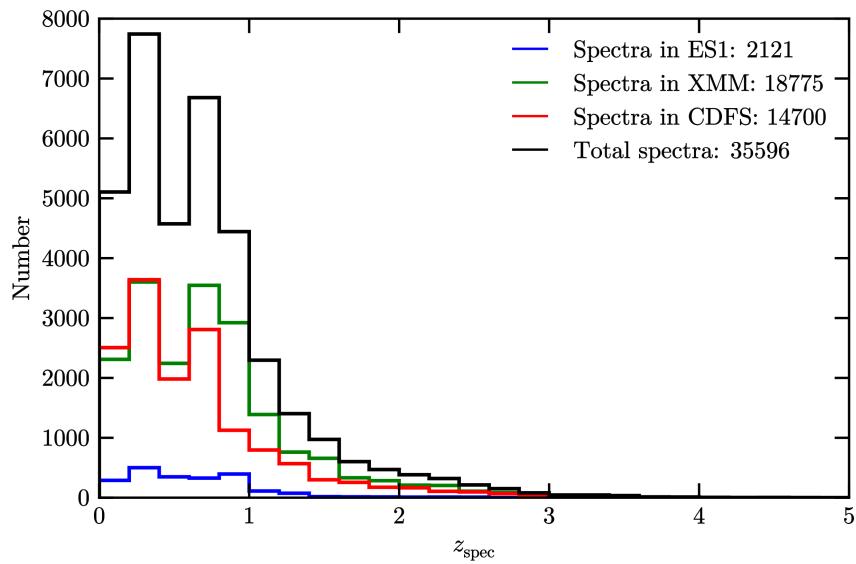


Figure 2.10 Spectroscopic redshift distribution of all spectra in the DES+VIDEO catalogue. The red, blue, and green histograms show the contributions from the ES1, XMM, and CDFS fields respectively, and the black histogram shows the combined distribution of all spectra. All histograms are plotted in redshift bins of $\Delta z_{\text{spec}} = 0.2$.

2.2 Forced photometry and catalogue production

2.2.1 Motivation

Now the DES and VIDEO datasets have been introduced, the upcoming part of this chapter will turn to the next step: merging these surveys into a single combined catalogue. The most straightforward way to achieve this merger would simply be to match the existing DES and Jarvis et al. (2013) VIDEO catalogues by position. However, the high quality of the DES data makes it possible to do better. By using a so-called *forced photometry* approach, this thesis aims to extract

more information from the VIDEO imaging than is possible with VIDEO alone. The concept behind forced photometry is to use imaging from a deeper or higher resolution survey to improve the photometry and positions that can be measured from shallower or lower resolution data (Banerji et al. 2015; Lang et al. 2016). In the case of DES and VIDEO, the two surveys have fairly similar resolutions, so the potential for improvement mainly revolves around the higher depths in the DES imaging. As demonstrated in Table 2.5, the VIDEO Z-band data reaches depths of $24.8 < Z_{AB} < 25.3$. While this is similar to the $z_{AB} < 24.9$ DES depths over the majority of the DES+VIDEO footprint, the 4.3 deg^2 deepest DES regions extend to $z < 26.2$, which is considerably deeper. Additionally, the combined $r + i + z$ is deeper than the z -band alone, as it enjoys the combined exposure times from the r , i and z imaging. Therefore, it is possible to use the deeper DES imaging to obtain secure detections for objects that would be too faint to pass the detection threshold in the VIDEO imaging alone. By extracting forced photometry from VIDEO based on these secure DES detections, the VIDEO measurements can be pushed to fainter magnitudes, increasing the number of sources with measured VIDEO fluxes. The concept is illustrated in Figure 2.11, which shows the increase in VIDEO measurements that can be achieved with forced photometry based on DES. Furthermore, in addition to the fact that it allows VIDEO fluxes to be obtained for fainter sources, forced photometry has a second benefit for this thesis. As mentioned in Section 2.1.1.3, the optical component of the DES+VIDEO catalogue consists of the DESDM catalogues. In order to maximise the amount of information contained within the combined catalogue, it is important to obtain VIDEO photometry for as many of these DESDM sources as possible. Forced photometry helps to achieve this objective, because it can measure VIDEO fluxes for all sources detected in DES.

To extract the forced photometry and to merge it with DES data, the author has created an automated pipeline. The rest of this section will describe this pipeline and the catalogue assembly process in general. The first part will present the implementation of forced photometry, the second part will report on the validation of this method, and the third part will describe the derivation corrected photometric uncertainties for the catalogue. Then, after these parts have identified and verified the necessary steps, the fourth section will summarise the automated pipeline to produce the desired DES+VIDEO catalogue.

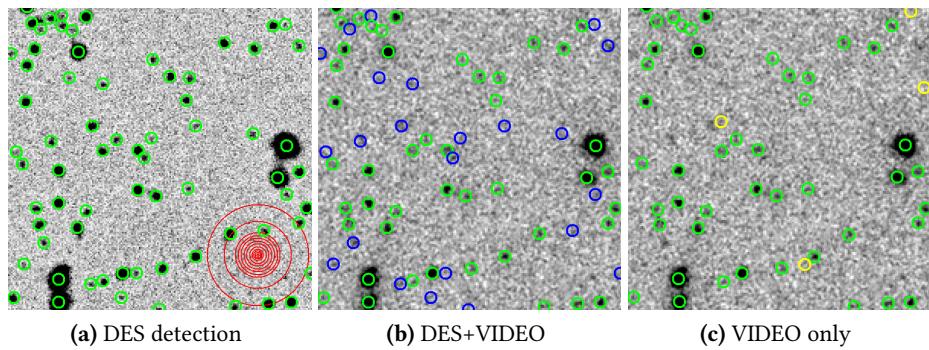


Figure 2.11 An illustration of the effectiveness of forced photometry in a section of the CDF-S where the DES data is significantly deeper than VIDEO. (a) DES $r + i + z$ detection image with 1.95 arcsec apertures centred on DES $r + i + z$ detections. All twelve aperture sizes in the DES+VIDEO catalogue (see Table 2.7) are superimposed to scale in red. (b) VIDEO Z-band image with DES detections from (a). The blue circles show sources which are detected in DES but not included in the VIDEO Jarvis et al. (2013) catalogue. (c) VIDEO Z-band image with detections from the Jarvis et al. (2013) VIDEO catalogue, based on the longest wavelength detection in the VIDEO Y-, J-, H- or K_s -band. Sources which are included in the Jarvis et al. (2013) catalogue but not detected in DES are shown in yellow. The abundance of blue circles compared to yellow circles illustrates that by using the deeper DES data, photometry can be obtained for VIDEO sources that are too faint to be detected in the VIDEO imaging.

2.2.2 Method

2.2.2.1 Image resampling and preparation

Before forced photometry can be extracted, the imaging must be suitably prepared. In the DES+VIDEO combination pipeline, this is achieved in the following way. Because the pipeline eventually measures forced photometry per DES co-add tile — for every combination of intersecting DES and VIDEO tiles — every VIDEO tile that overlaps a given DES tile is resampled to match the position, pixel scale, and size of that DES tile. This is achieved via **SWarp** v.2.38.0 (Bertin et al. 2002; Bertin 2010), using a **LANCOS3** kernel to rescale the VIDEO data from the VIDEO pixel scale of $0.2 \text{ arcsec pix}^{-1}$ to the DES scale of $0.263 \text{ arcsec pix}^{-1}$. Resampling is performed for every available VIDEO filter, and the **SWarp** configuration file used for this process is shown in Appendix A.1. The result is a series of VIDEO tiles covering the same $0.73 \times 0.73 \text{ deg}^2$ regions as the corresponding DES tiles. If a VIDEO tile does not fully overlap a given DES tile, **SWarp** sets the pixel values in the image and weight map to 0 in the areas where no VIDEO data is present.

For the DES imaging, preparation requires consideration of the overlap pattern between the DES and VIDEO tiles. There are two special cases, which are both addressed by the pipeline. The first case applies if a DES tile is not fully contained within a given VIDEO tile, i.e. if it lies on an edge. For those tiles, the pipeline trims the DES $r + i + z$ detection image to the size of the overlapping VIDEO image. The second case concerns DES tiles where there is data from multiple VIDEO tiles. In those tiles the pipeline duplicates the detection image for each VIDEO tile and cuts the duplicates down to the overlapping regions. This trimming is performed based on the resampled VIDEO K_s -band weight map, which was chosen because it is observed in every tile. After overlap has been accounted for, the pipeline then finishes the data preparation by assigning a value of **Nan** to all DES detection image pixels for which the corresponding VIDEO weight map pixels are zero. These zero weights occur a) for regions at the very edge of the VIDEO tile where there are no observations, b) for regions in the resampled VIDEO tiles where there is no overlapping DES data (see the previous paragraph), and c) for regions covered by the faulty detector 16 (see Section 2.1.2.4). This approach ensures that all detected sources lie within DES regions

that overlap existing and reliable VIDEO data, since the forced photometry code does not allow detected sources to contain **NaN** values.

2.2.2.2 Source extraction

As explained previously in Section 2.2.1, one of the advantages of forced photometry is that it can allow VIDEO fluxes to be measured for a maximum number of sources in the DESDM catalogues. To make full use of this strength, the DES+VIDEO catalogue pipeline must therefore replicate the DESDM source detection process as closely as possible, so that the measured VIDEO objects are a close match to the DESDM objects. Unfortunately, the photometry in the DESDM pipeline was obtained with a special of **SExtractor** that the author does not have access to, so it is not possible to copy that source detection method outright. In spite of this, the author has obtained a very close approximation to the DESDM results (as will be shown in Section 2.2.3) via a description of the DESDM configuration file, obtained from internal communication with the DES collaboration. Together with the relevant DES, VIDEO, and **SExtractor** documentation in the literature, this internal communication helped to inform the source extraction method for the DES+VIDEO pipeline. A summary of the adopted approach is presented below.

Forced photometry is extracted in every available VIDEO filter, by running **SExtractor** v.2.19.5 (Bertin & Arnouts 1996) in dual image mode for each overlapping DES+VIDEO tile combination. This process uses the DES $r + i + z$ co-adds (prepared as described in the previous section) as detection images, and the corresponding resampled VIDEO tiles as measurement images. Appendix A.2 shows the **SExtractor** configuration file used in this process. With this setup, **SExtractor** measures photometry through a series of steps. A summary of this process, including the configuration choices made for this thesis, is presented below. More information can be found in the **SExtractor** documentation¹¹.

Firstly, **SExtractor** creates a background map for both the DES detection and VIDEO measurement images. To do this, it estimates the background in a grid of $16 \times 16 = 256$ pixel boxes (the value of 256 used in this thesis has been imposed via the **BACK_SIZE** parameter). Within these boxes, pixels brighter than 3σ above the median are iteratively clipped. To account for boxes where the

¹¹<http://www.astromatic.net/pubsvn/software/sExtractor/trunk/doc/sExtractor.pdf>

background is overestimated due to bright stars, median filtering is applied to the grid, using a filter size of 3 boxes (as specified by the **BACK_FILTERSIZE** parameter). The final background maps are obtained by smoothing the grid of boxes via a bicubic-spline interpolation. The resulting maps are then subtracted from the corresponding detection and measurement images. In parallel with the background maps, **SExtractor** produces a map of the RMS background noise, which is used to scale the input weight maps to the appropriate absolute level.

After subtracting the background, **SExtractor** convolves the detection image with a 3×3 pixel convolution kernel. This improves the detectability of faint extended sources¹² by increasing the contrast with the background noise. The kernel file **sex.conv** used in this thesis is identical to the one used for the DESDM catalogues, and contains the following elements:

$$\mathbf{conv} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (2.9)$$

SExtractor then identifies source detections from the convolved background-subtracted $r + i + z$ detection images. For this thesis, the algorithm considers an object detected if it consists of six or more adjacent pixels (specified via **DETECT_MINAREA** = 6) with a signal-to-noise ratio of 1.6σ (**DETECT_THRESH** = 1.6σ) above the local background¹³. For blended detections consisting of overlapping sources, **SExtractor** attempts to separate the detections into subcomponents using a multi-thresholding algorithm (Bertin & Arnouts 1996). For this thesis, a subcomponent of a composite object is considered separate if its total integrated value is greater than 0.001 of the composite (**DEBLEND_MINCONT**=0.001), and the maximum number of subcomponents is 32 (**DEBLEND_NTHRESH**=32).

For every detected object, **SExtractor** extracts photometry from the background-subtracted VIDEO measurement images. The photometry consists of two main types of measurement:

- **Aperture photometry** is measured by counting the flux of pixels inside twelve fixed-size circular apertures, with a range of diameters between

¹²<http://www.astromatic.net/pubsvn/software/sextorator/trunk/doc/sextorator.pdf>

¹³To be more precise, the required threshold for detection is adjusted for each pixel according to the weight map, such that each pixel must be above a threshold $t_i = 1.5 \times \sqrt{\sigma_i^2}$, where i denotes an individual pixel and σ_i^2 is the variance in that pixel as determined from the weight image.

0.49 arcsec and 17.53 arcsec. A list of all the diameters is provided in Table 2.7. To illustrate their sizes, the apertures are superimposed on part of a DES detection image in Figure 2.11a. The diameters have been chosen to match the ones used by DESDM, so that all DES+VIDEO fluxes are measured in the same size apertures. **SExtractor** outputs the fluxes and magnitudes as **FLUX_APER** and **MAG_APER** respectively.

- **Auto photometry** employs a flexible elliptical aperture scaled to encompass (most of) the flux of a given source. The method implemented in **SExtractor** follows an approach pioneered by Kron (1980) and adapted by Bertin & Arnouts (1996). It computes the shape of the desired elliptical aperture (i.e. the ratio between the semi-major and semi-minor axes) from the second order moments of the flux distribution in the detection image. The size of the ellipse is determined by a dimensionless quantity that **SExtractor** refers to as the Kron radius R_{Kron} , which is a product of the first moment of the flux distribution and a scale factor k . For this thesis, the value of k and the minimum permitted Kron radius $R_{\text{Kron,min}}$ are set to values of 2.5 and 3.5 respectively (via the **PHOT_AUTOPARAMS** parameter). **SExtractor** uses R_{Kron} to obtain the semi-major axis a and semi-minor axis b of the final scaled elliptical aperture in units of pixels:

$$a = R_{\text{Kron}} \times \mathbf{A_IMAGE}, \quad (2.10)$$

$$b = R_{\text{Kron}} \times \mathbf{B_IMAGE}, \quad (2.11)$$

where **A_IMAGE** and **B_IMAGE** are the aforementioned shape parameters computed from the second order moments, supplied in units of pixels. If the calculated value of R_{Kron} for a given source is less than $R_{\text{Kron,min}} = 3.5$, instead of a scaled aperture **SExtractor** uses an aperture of fixed area instead (still with an elliptical shape determined from the imaging). Its area corresponds to that of a circular aperture with a radius of 7.41 pixels (**PHOT_AUTOAPERS** = 7.41)¹⁴. **SExtractor** saves the resulting auto fluxes and magnitudes as **FLUX_AUTO** and **MAG_AUTO** in the output catalogues, to-

¹⁴This corresponds to a diameter of 1.97 arcsec, roughly equal to the seeing.

gether with **A_IMAGE**, **B_IMAGE**, R_{Kron} (listed as **KRON_RADIUS**), and various other shape parameters.

APERTURE DIAMETERS			
	Aperture size (arcsec)	Aperture size (pix)	
0.49	1.85	4.87	18.52
0.97	3.70	5.84	22.22
1.46	5.55	6.81	25.93
1.95	7.41	7.79	29.63
2.92	11.11	11.68	44.44
3.90	14.81	17.53	66.67

Table 2.7 Aperture diameters for the aperture photometry in the DES+VIDEO catalogue. The aperture sizes were chosen to match those of the DESDM catalogues, so that the DES+VIDEO photometry is measured in identical apertures.

2.2.3 Source extraction method validation

It is important to confirm that the configuration in the DES+VIDEO pipeline indeed replicates the DESDM configuration as closely as possible. To verify this, the author used the DES+VIDEO setup for a single **SExtractor** run in the DES0227-0458 tile, specifying the combined $r + i + z$ image as the detection image and the DES z -band image as the measurement image. The current section compares the result to the DESDM catalogue in the same tile.

Within the single tile that was studied, the DESDM catalogue contains 126 624 detected sources, whereas the DES+VIDEO method in this thesis – let us refer to this as the ‘Schooneveld configuration’ – finds 126 980 sources. When merging these catalogues by position¹⁵ within a 1 arcsec radius (permitting no duplicate matches), the total number of matches is 123 993. The numbers imply that the DESDM catalogues contain 2631 objects ($\approx 2\%$) not detected by the Schooneveld configuration. Conversely, the Schooneveld setup produces 2987 sources ($\approx 2\%$) that are not present in the DESDM catalogues. Figure 2.12 illustrates the difference in detections between the two methods as a function of magnitude. When the author inspected the images for the divergent detections, it became clear that these objects are clustered around bright stars. They are

¹⁵Based on the **ALPHAPEAK_J2000** and **DELTAPEAK_J2000** output parameters.

also mostly undetected or extremely faint in the z -band. This suggests that the difference is caused by small differences in the way the two source extraction methods handle the detection threshold in edge cases, which may have to do with the specifics of the background estimation. Possible explanations are the fact that the DESDM configuration used a different version of **SExtractor**, or a perhaps slightly different weights image. Nonetheless, compared to the ~ 127000 sources in the test tile, the number of ~ 3000 divergent detections is very small, and it is overall reassuring that the two methods retrieve very similar sets of objects.

The DESDM and Schooneveld configurations were also compared in terms of the position of the brightest pixel for each object (given by the **ALPHAPEAK_J2000** and **DELTAPEAK_J2000** parameters). The agreement was found to be excellent: 98.5% of all detections have identical positions and 99.7% differ by at most 1 pixel. This demonstrates that the forced photometry is working as desired, since it confirms that the detections in the Schooneveld configuration correspond to the same sources in the DESDM catalogues. Together, this result and the result from the previous paragraph show that the Schooneveld and DESDM configurations retrieve largely the same sources, almost entirely in exactly the same positions. Therefore, it is concluded that the Schooneveld configuration can indeed achieve the aim of obtaining VIDEO photometry for a maximal number of DES sources.

Additionally, Figure 2.13 illustrates that the Schooneveld configuration also retrieves similar photometry to the DESDM configuration. It is suspected that any discrepancies are largely due to the fact that the DESDM pipeline possibly uses a slightly different background estimation procedure. Because **SExtractor** obtains its photometry from the background-subtracted images, magnitudes are sensitive to the specifics of the subtraction process. This suspicion is corroborated by the observation that sources with a magnitude difference of $z_{AB,\text{Schooneveld}} - z_{AB,\text{DESDM}} > 0.01$ preferentially reside in regions where the **SExtractor BACKGROUND** parameter diverges between the two catalogues.

All in all, the good degree of agreement in all three of the above tests implies that the source extraction method in this thesis is adequate for the purpose of obtaining forced photometry for the DES+VIDEO catalogue.

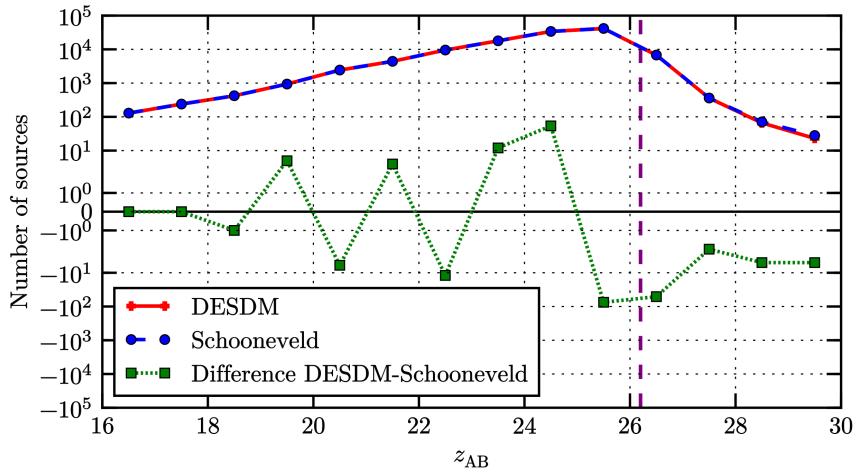


Figure 2.12 z -band magnitude in a 1.95 arcsec diameter aperture versus number of detections in bins of 1 mag, for detections in the DES0227-0458 tile only. Blue pluses show data from the DESDM catalogues. Green squares show the Schooneveld configuration results, from a single **SExtractor** run with the DES+VIDEO configuration used in this thesis, as described in the text. The difference between these data points is shown in red. The vertical purple dotted line shows the mean 5σ depth in the DES0227-0458 tile.

2.2.4 Correction of the VIDEO photometric uncertainties

Before the VIDEO forced photometry can be incorporated into the DES+VIDEO catalogue, its uncertainties must be adjusted to the correct values. The photometric noise properties estimated by **SExtractor** assume that images are made up of uncorrelated pixels, so that the noise follows a Poisson distribution. As a result, the uncertainties listed in the **SExtractor** output catalogues are simply based on the pixel-to-pixel rms. However, this assumption of uncorrelated noise is not valid for the VIDEO images, as they have been resampled several times. The VIDEO survey data reduction process (see Section 2.1.2.3) initially rescaled the images from the VIRCAM pixel scale of $0.339 \text{ arcsec pix}^{-1}$ to the VIDEO co-add scale of $0.2 \text{ arcsec pix}^{-1}$, and the resampling for this thesis (described in Section 2.2.2.1) has further rescaled the imaging to the $0.263 \text{ arcsec pix}^{-1}$ DES pixel scale. In both instances, this process has introduced correlations between pixels, which means that the **SExtractor** errors are expected to give an inaccurate representation of the true VIDEO uncertainties (see e.g. Labb   et al. 2003 and

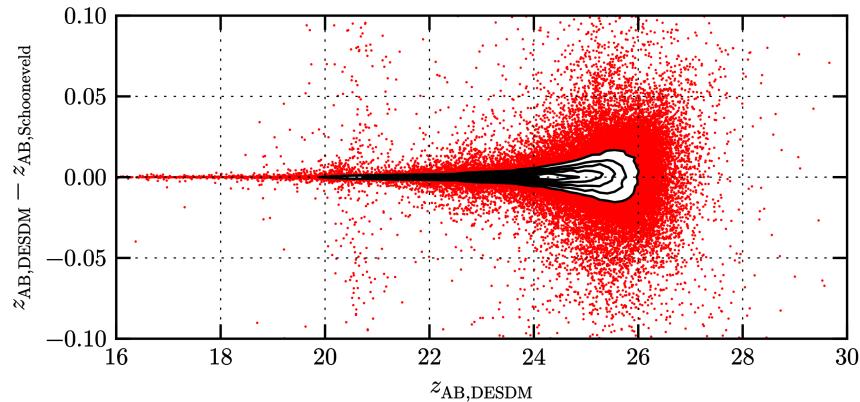


Figure 2.13 Difference in z -band magnitude in a 1.95 arcsec diameter aperture between the DESDM configuration and the Schooneveld configuration, plotted as a function of DESDM magnitude. Both configurations contain data from the DES0227-0458 tile only. The Schooneveld configuration consists of a single **SExtractor** run with the DES+VIDEO configuration used in this thesis, as described in the text. Each plotted point corresponds to an object common to both outputs.

Gawiser et al. 2006 for a demonstration that correlated fluxes lead to incorrect **SExtractor** errors).

To obtain a more accurate measure of the VIDEO errors, the author has computed the correct uncertainties via a technique developed by Labb   et al. (2003) and Gawiser et al. (2006), which is commonly applied in the literature (e.g. Quadri et al. 2007; Whitaker et al. 2011; Muzzin et al. 2013; Jarvis et al. 2013). Conceptually, this approach involves measuring the flux inside a large number of fixed-size apertures placed randomly in empty regions of the image. The distribution of fluxes within these apertures can be approximated by a Gaussian curve. The true photometric error, which corresponds to the background noise, can then be computed from the standard deviation of this Gaussian.

Before the discussion will move on to applying this method to the data, it is important to note that correlated noise is not an issue for the DES uncertainties. Because the DES co-add pixel scale of $0.263 \text{ arcsec pix}^{-1}$ is identical to that of the DECam CCDs, the DESDM data reduction process described in Section 2.1.1.3 did not need to apply any rescaling to resample the DECam exposures into the DES co-adds. This means that the co-addition introduced only minimal flux correlations, exclusively from stacking with **SWarp**. As a result, the original

SExtractor errors in the DESDM catalogues indeed closely capture the true photometric uncertainties, and no corrections are required.

2.2.4.1 Aperture flux errors

The above method for correcting the VIDEO uncertainties has been implemented via a custom code developed by the author. In this algorithm, errors are calculated for each VIDEO tile and filter individually, following a process that will now be described below. Firstly, the error-calibration pipeline creates a segmentation image for each VIDEO co-add tile¹⁶ via **SExtractor**, using essentially¹⁷ the same configuration as described in Section 2.2.2.2 and Appendix A.2. In such segmentation maps, **SExtractor** assigns an integer value to pixels designated as part of a detected object, and a value of zero to all other pixels. An example region of such a segmentation map is shown in Figure 2.14. To identify empty regions, the pipeline firstly randomly chooses 500 000 positions on the segmentation map. For a given aperture size the code then selects a random subset of 10 000 positions with empty apertures (i.e. not encircling any non-zero pixels that overlap a detected object), which it uses to extract the required background fluxes from the VIDEO co-add tiles. For each tile and filter, the random selection of 10 000 empty apertures from the 500 000 random positions and the subsequent photometry procedure is executed seven times – once for each of the first seven aperture sizes in Table 2.7 (i.e. with diameters between 0.49 arcsec and 4.87 arcsec). The larger apertures are omitted, because these are more likely to encircle some low surface brightness sources just below the detection limit and are thus less reliably truly empty. Fluxes from random apertures on the very edge of a given tile (where the pixel flux values are zero) are then removed by rejecting any fluxes that equal exactly zero. The result is a series of just under¹⁸ 10 000 fluxes in random empty apertures for every VIDEO tile, VIDEO filter, and aperture size.

For each tile, filter, and aperture size, the pipeline fits a Gaussian curve to the distribution of fluxes. Figure 2.15a illustrates this process, showing the results

¹⁶For clarity, these are the original VIDEO data products presented in Section 2.1.2.3, before the resampling and cutting up by the DES+VIDEO pipeline described in Section 2.2.2.2.

¹⁷The only thing that needed to be changed in the input files are the values specified for the aperture diameter, since these must be supplied in pixels. Because the VIDEO tiles have a pixel scale of $0.2 \text{ arcsec pix}^{-1}$, the aperture sizes in these **SExtractor** runs are scaled by a factor of

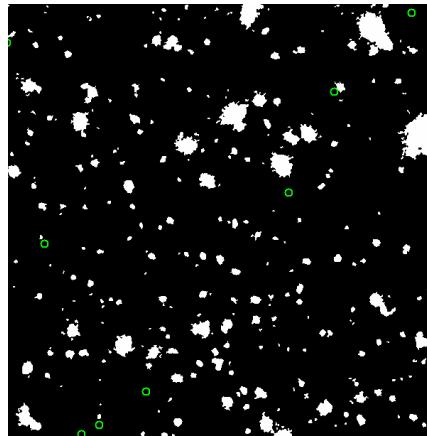


Figure 2.14 Segmentation image for the VIDEO K_s -band in the xmm3 tile, with 1.95 arcsec apertures randomly placed in empty regions. The white areas consist of pixels that **SExtractor** has assigned to a detection. For the purpose of error estimations, apertures are designated as empty if they do not encircle any of these pixels.

for the 0.97, 1.95, and 2.92 arcsec apertures in the xmm3 K_s -band. The standard deviation σ of the fitted Gaussians corresponds to the desired quantity, i.e. the background error in a given aperture. It is reassuring that upon visual inspection of the pipeline fits, all the measured flux distributions were found to be fitted well by a Gaussian, as expected if the method in the DES+VIDEO pipeline indeed measures the background noise in empty regions (Labbé et al. 2003; Gawiser et al. 2006).

Up to this point, the error-estimation algorithm can successfully compute the background error for each tile and filter in the seven smallest aperture sizes. To smooth out the effects of random aperture selection for these seven sizes, as well as to obtain uncertainties for the five largest aperture sizes, the background error σ is expressed as a function of aperture radius, from which maximally reliable errors for all twelve aperture sizes can be interpolated and extrapolated. Following Gawiser et al. (2006) and Whitaker et al. (2011), this dependence of σ on the aperture radius r is parameterised by a power law:

0.2/0.263 = 0.760 compared to the values listed in Table 2.7 and Appendix A.2. The true size of the apertures (i.e. the size in arcsec) remains the same.

¹⁸The numbers ended up as slightly less than 10 000 because of the rejected fluxes from apertures that happened to lie on the very edge of a tile.

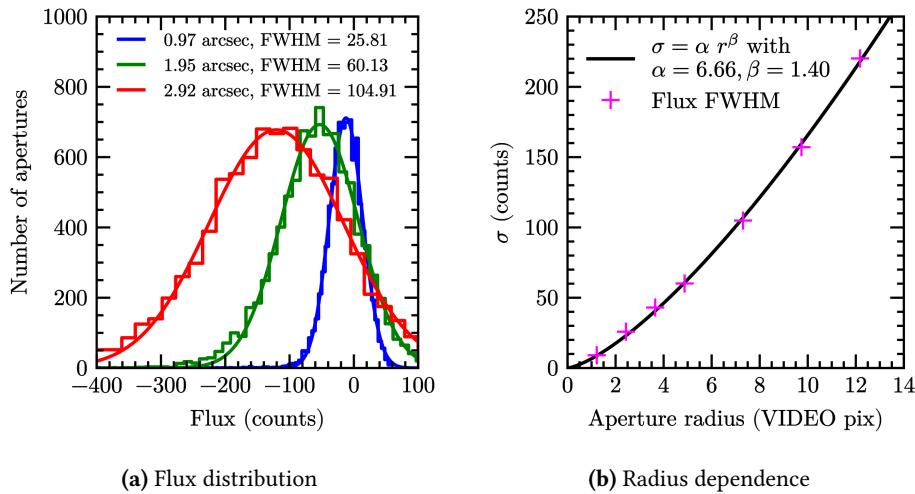


Figure 2.15 An illustration of the two-part fitting algorithm to estimate VIDEO photometric errors, showing results for the K_s -band in the xmm3 tile. **(a)** Histograms of flux distributions for three of the seven measured apertures (i.e. 0.97, 1.95, and 2.92 arcsec; in blue, green, and red respectively), binned in intervals adjusted proportionally to the standard deviation. Each distribution has been fitted to a Gaussian, which is overplotted. For each aperture size, the measured background error σ is equal to the Gaussian FWHM listed in the legend. **(b)** The background error σ for all seven 0.49–4.87 arcsec apertures as a function of aperture radius. The values of σ (shown in pink pluses) have been computed via the Gaussian fitting procedure illustrated in (a). Radii are given in the pixel scale of the VIDEO tiles (i.e. 0.2 arcsec pix $^{-1}$), and the apertures from (a) correspond to the second, fourth, and fifth data points from the left. The values of σ have been fitted to a power law $\sigma = \alpha r^\beta$ to obtain values for α and β . The power law for the best-fit values of α and β is shown in black. This power law is then used to calculate the final errors for all twelve DES+VIDEO aperture sizes.

$$\sigma = \alpha r^\beta, \quad (2.12)$$

where $\beta = 1$ in the case of no correlations between pixels, and $\beta = 2$ for perfect correlations. The pipeline fits this power law to the values of σ in all seven measured apertures to obtain α and β . As an example, the resulting curve for the xmm3 K_s -band is shown in Figure 2.15b. Here, as well as in all other fields and filters, the output value of β was crucially found to lie between 1 and 2, confirming the presence of correlations between pixels.

As a final step, the pipeline then computes the desired DES+VIDEO catalogue flux errors (**FLUXERR_APER**) for all twelve apertures in Table 2.7, by inserting the best-fit parameters α and β into Equation 2.12. At the same time these flux errors are also converted to magnitude errors (**MAGERR_APER**). The resulting measurements of **FLUXERR_APER** and **MAGERR_APER** in every tile and filter are later used by the DES+VIDEO catalogue creation pipeline to replace the original errors computed by **SExtractor**.

Assumptions

It is important to point out that the above process for computing the corrected VIDEO errors is based on a number of assumptions. Firstly, the error measurements assume that the uncertainties are constant within a given VIDEO tile. In practice, this is not completely true, because some parts of the tiles contain a higher or lower number of exposures than average (as described in Section 2.1.2.4). This variation causes slight differences in the true errors, as a higher density of exposures leads to lower true background errors (and vice versa). Nevertheless, since the VIDEO coverage is reasonably uniform within each tile (as demonstrated in Figure 2.9), the assumption of constant errors is reasonably accurate. It is noted that Jarvis et al. (2013) have also made the same assumption.

Secondly, it is assumed that the flux uncertainties can be directly equated with the background error. In reality, the true uncertainties also contain a contribution from the Poisson counting of electrons in the detectors (Quadri et al. 2007). However, this effect is only significant for the very brightest sources. In the Jarvis et al. (2013) catalogues, which do include Poisson counting, its contribution to the error only reaches $>1\%$ at $Z_{AB} < 20.8$ in the Z -band and at $K_s < 17.4$ in the K_s -band. The DES+VIDEO catalogue is primarily aimed at computing pho-

tometric redshifts for galaxies and finding high-redshift galaxies. Since the vast majority of these objects are significantly fainter than the threshold at which Poisson counting becomes important, the omission of the Poisson contribution is justified for the purposes of this thesis.

2.2.4.2 Auto flux errors

In contrast to aperture photometry, which is measured in circular apertures of a fixed size, **SExtractor** measures auto photometry in elliptical apertures which are scaled to the size of each individual detection (as explained previously in Section 2.2.2.2). By using Equation 2.12, the DES+VIDEO pipeline can also estimate accurate errors for those varying apertures. In order to do so, the error estimation code defines an ‘effective radius’ R_{eff} , which is the radius that corresponds to a circular aperture of the same area A_{circ} as the ellipse:

$$\pi R_{\text{eff}} = A_{\text{circ}} \equiv A_{\text{ell}} = \pi ab. \quad (2.13)$$

Here, A_{ell} is the area of an elliptical aperture with semi-major and semi-minor axes a and b . The expressions that **SExtractor** uses to calculate a and b have been introduced previously in Equations 2.10 and 2.11. Together with Equation 2.13, these yield:

$$R_{\text{eff}} = \sqrt{R_{\text{Kron}} \times \mathbf{A_IMAGE} \times R_{\text{Kron}} \times \mathbf{B_IMAGE}}. \quad (2.14)$$

If the Kron radius computed by **SExtractor** falls below the specified minimum value introduced in Section 2.2.2.2, the effective radius R_{eff} instead corresponds to that of a 1.97 arcsec diameter circular aperture. For each object, the error calibration pipeline computes the auto uncertainties by inserting the value of R_{eff} into Equation 2.12, using the best-fit parameters α and β computed from the flux FWHM in circular apertures as described in Section 2.2.4.1. The resulting auto errors and corresponding magnitude errors later replace the **FLUXERR_AUTO** and **MAGERR_AUTO** uncertainties computed by **SExtractor**. Similarly to the aperture flux procedure, the auto flux calibration assumes that the background is constant within a tile, and that the contribution from Poisson counting is insignificant.

DES+VIDEO ERRORS COMPARED TO JARVIS ET AL. (2013)						
	Error (2 arcsec aperture) (counts)		Error (2 arcsec aperture) (counts)		Error (2 arcsec aperture) (counts)	
Tile	Z-band		Y-band		J-band	
	Schooneveld	Jarvis	Schooneveld	Jarvis	Schooneveld	Jarvis
cdfs1	24.4	24.8	31.2	31.3	31.6	32.2
cdfs2	-	-	-	-	44.7	46.5
cdfs3	-	-	-	-	48.7	51.1
es1-north	15.3	14.5	22.7	21.2	56.4	60.6
es1-south	17.3	16.9	27.7	27.1	57.0	62.2
xmm1	14.0	13.0	22.6	21.6	-	-
xmm2	-	-	22.9	22.2	30.6	31.3
xmm3	15.0	13.6	22.6	21.6	30.3	30.6
H-band						
Tile	Schooneveld	Jarvis	Schooneveld	Jarvis	K_s -band	
cdfs1	49.7	51.1	66.0	68.8		
cdfs2	101.9	109.9	73.9	76.3		
cdfs3	-	-	140.2	151.6		
es1-north	53.6	54.0	63.6	66.0		
es1-south	47.2	47.7	64.2	67.8		
xmm1	64.2	67.4	70.1	72.1		
xmm2	45.8	47.0	67.4	70.4		
xmm3	46.8	48.0	63.4	66.7		

Table 2.8 A comparison between the VIDEO errors computed via the procedure in this thesis (labelled ‘Schooneveld’) and the errors from the Jarvis et al. (2013) catalogues. The Jarvis et al. (2013) errors in this table include the random background error only (i.e. they do not contain the contribution from Poisson counting of electrons in the VIRCAM detectors). Dashes indicate filters which were not observed in a given tile. Since Jarvis et al. (2013) photometry used a 2 arcsec diameter aperture, the Schooneveld errors have been scaled from the DES radius of 0.97 arcsec to the VIDEO radius of 1 arcsec via Equation 2.12 – i.e. $\sigma_{\text{2arcsec}} = \sigma_{\text{1.95arcsec}} (1/0.97)^\beta$. The values of β for each filter and tile have been derived from the best-fit parameters of Equation 2.12, which have been obtained as described in Section 2.2.4.1. The agreement between all pairs of errors is generally good, with all values agreeing within 10% and a majority within 5%.

2.2.4.3 Verification of the corrected errors

To verify that the error estimation algorithm in this thesis functions as intended, the author has compared its output to the errors in the Jarvis et al. (2013) catalogues (which were similarly empirically determined from the standard deviation of flux in randomly placed apertures). Table 2.8 presents the results of this comparison. Reassuringly, the agreement is generally good. In all tiles and filters, the two error values agree within < 10%, with most differing by less than 5%.

The corrected errors were also compared to the original (uncorrected) **SExtractor** errors. The analysis revealed a considerable difference, with **SExtractor** underestimating the uncertainties by up to a factor of 2 in some filters. This observation quantitatively underlines the extent to which flux correlations introduced by image resampling can cause problematic inaccuracies in error estimations that are simply based on pixel-to-pixel rms. It therefore confirms that if such resampling has occurred, other measurement techniques, such as the empty apertures method pioneered by Labb   et al. (2003) and Gawiser et al. (2006) and implemented in this thesis, must be used to identify the errors accurately.

2.2.5 Combined DES+VIDEO catalogue production pipeline

Now all the necessary sub-procedures for compiling the DES+VIDEO catalogue have been introduced, this section presents the algorithm that combines them into a single data processing pipeline. It consists of the following steps:

1. Firstly, the pipeline prepares the images via the procedure presented in 2.2.2.1, resampling the VIDEO tiles to the size and pixel scale of the DES tiles and trimming the DES tiles to the overlap of each VIDEO tile.
2. The pipeline then measures forced photometry for the VIDEO data with **SExtractor**, using the method described in Section 2.2.2.2. To recap, for each combination of overlapping DES and VIDEO tiles, objects are detected based on the DES $r + i + z$ co-adds and their photometry is measured from the resampled VIDEO tiles.
3. Then, in each DES tile, the pipeline creates and optimises a combined *grizY* catalogue via two steps:

- (a) The raw DESDM catalogues for the individual g , r , i , z and Y filters are merged into one. Because they are all based on exactly the same detection image, the tables for each filter contain the exactly same ordered list of sources, so they can simply be concatenated.
 - (b) The photometry in this combined table is corrected for position-dependent variations in galactic reddening and photometric calibration, via the SLR corrections described in Section 2.1.1.4. The SLR-corrected values are appended to the table, adding new columns for the auto photometry and 1.95 arcsec diameter aperture photometry.
4. For each VIDEO tile overlapping a given DES tile, the pipeline combines the VIDEO **SExtractor** output from step 2 with the DES $grizY$ catalogues from step 3. To perform this merging, the VIDEO output for every available $ZYJHK_s$ filter is matched to the DES tables by position within a 1 arcsec radius. If a VIDEO filter was not observed in a given tile, the corresponding fluxes, magnitudes, and errors are set to -99. The result is a series of $grizYZYJHK_s$ DES+VIDEO catalogues for every combination of DES+VIDEO tiles.
 5. The pipeline then corrects the VIDEO aperture and auto photometry errors to account for correlated flux, using the error-calibration code described in Section 2.2.4.
 6. All observed fluxes and flux errors are converted from counts to units of $\text{erg s}^{-1} \text{Hz}^{-1}$. This is achieved by multiplying all fluxes and flux errors by a factor of $10^{0.4(-48.6-m_{\text{zp}})}$, where $m_{\text{zp}} = 30.0$ is the photometric zero-point of the original imaging, and -48.6 is the zero-point of the AB magnitude system.
 7. Limiting magnitudes (10σ , 7.5σ , 5σ , 3σ , and 1σ) for each $grizY$ DES filter are added to the catalogues. The 10σ depths have been provided by the DES collaboration (see Section 2.1.1.5), and the other four depths are converted from the 10σ depths via the method described in Section 2.1.1.5.
 8. The last step concatenates all the catalogues from the previous step. Firstly, in DES tiles that contain multiple overlapping VIDEO tiles, the step 7 catalogues for each VIDEO tile are stacked into a single table, such that there

is one complete catalogues for each DES tile. Then, the final DES+VIDEO catalogue is produced by stacking the (stacked) catalogues from all DES tiles.

The above pipeline was run on the DES and VIDEO datasets, producing a catalogue with 2 443 576 sources. Two important points must be made regarding duplicates and aperture corrections; these will be addressed in the following two sections.

2.2.5.1 Duplicates

Due to the way the pipeline algorithm stacks data, the output DES+VIDEO catalogue contains some duplicate objects. These arise from tile overlap, both between VIDEO tiles and between DES tiles. In the former case, objects within a particular DES tile are matched to near-IR detections from multiple VIDEO tiles. In the latter case, objects are multiply detected in DES because they lie on the edges of the DES tiles, which overlap by ~ 1 arcmin (as described in Section 2.1.1.3). It was decided to retain all such duplicates in the main version of the DES+VIDEO catalogue. Duplicate detections are useful for the high-redshift search in this thesis, because some high-redshift galaxies are likely to have higher quality photometry in one of the duplicate instances. A maximally inclusive range of detections also benefits the catalogue as a resource for the community, since any astronomer who wishes to make use of the dataset can apply their own duplicate removal according to the specific aims of their research project.

Because some parts of this thesis do require a dataset without duplicates (such as the verification tests that will be described in Section 2.3), the author has also created a non-duplicate version of the DES+VIDEO catalogue. This was achieved by eliminating objects in two stages. The first step removed duplicates from overlapping VIDEO tiles, retaining the objects with the highest VIDEO K_s -band depth in a 1.95 arcsec aperture. This selection resulted in the rejection of 190 599 catalogue entries. The second stage then dealt with duplicates originating from the 1 arcmin overlap between DES tiles. To remove these, the DES+VIDEO catalogue was matched to a list of objects specially assembled by the author from the DESDM catalogues. During this assembly, duplicate detections were removed by retaining only the entries closest to the center of their respect-

ive tiles, which is essentially equivalent to trimming each tile by 0.5 arcmin. Care was taken to retain objects on the sides where the VIDEO fields extend beyond the DES footprint (because these sides do not contain duplicates from DES tile overlap). Altogether, the second duplicate rejection stage eliminated 100 402 DES+VIDEO objects, resulting in a non-duplicate DES+VIDEO catalogue of 2 152 575 objects.

2.2.5.2 Note on aperture corrections

Photometric datasets sometimes include aperture corrections, which are applied to the photometry in order to account for flux that falls outside the measurement apertures (Labbé et al. 2003; Quadri et al. 2007; Walcher et al. 2011). For an aperture of a given size, the correction can be obtained from the growth-curve of bright stars, by calculating what fraction of the stellar PSF falls outside the aperture. Consequently, such computations require accurate estimates of the PSF across each image. This would make it challenging to estimate accurate aperture corrections for the DES photometry, because the inhomogeneity of the co-add imaging (see Section 2.1.1.5) means that the PSF can vary significantly within a tile. Fortunately, aperture corrections are not strictly necessary for the work in this thesis. As will be explained in detail in Section 3.2.3, the photometric redshift code chosen for this thesis can apply offsets to photometry based on a training sample of galaxies with known spectra, so that colours are adjusted empirically according to the best-fit templates. This essentially automatically incorporates aperture corrections into the colours¹⁹.

For the reasons described above, it was decided not to apply aperture corrections to the DES+VIDEO photometry. A quick estimate based on results from Labb   et al. (2003) suggests that the flux missed as a result of this would amount to roughly ~ 0.25 mag AB for a 1.95 arcsec diameter aperture, given the DES+VIDEO seeing of FWHM $\sim 0.8\text{--}1.0$ arcsec. A thorough computation of DES+VIDEO aperture corrections is left as a suggestion for future work.

¹⁹It must be noted that this feature only corrects the colours internally to the photo-z code, and that it leaves the fluxes and magnitudes in the DES+VIDEO catalogue uncorrected.

2.3 Catalogue verification and results

Several verification tests have been performed to verify the accuracy of the DES+VIDEO catalogue, all of which shall be presented in this section. These assessments compare its astrometry and photometry to that of the Jarvis et al. (2013) VIDEO catalogues. The investigation also explores the number counts in both datasets, in order to measure the effectiveness of the forced photometry strategy in this thesis. At the end, further number counts are presented to provide an overview of the final DES+VIDEO catalogue.

2.3.1 Astrometry verification

The first verification test concerns whether the astrometry between the DES and VIDEO surveys is consistent. For this comparison, the non-duplicate DES+VIDEO sources were matched to the Jarvis et al. (2013) objects by position within a 1 arcsec radius. The position offsets between the two datasets were then quantified by defining two metrics for each object in the matched catalogue²⁰:

$$\Delta RA = RA_{\text{DES}} - RA_{\text{VIDEO}} = \text{ALPHAPEAK_J2000_G} - \text{ALPHAPEAK}, \quad (2.15)$$

$$\Delta DEC = DEC_{\text{DES}} - DEC_{\text{VIDEO}} = \text{DELTAPEAK_J2000_G} - \text{DELTAPEAK}. \quad (2.16)$$

Figure 2.16a plots the resulting values of ΔRA and ΔDEC for the 199 252 brightest objects, selected via a cutoff (SLR-corrected) magnitude of $z_{\text{DES,AB}} < 22.0$. The mean offsets are $\overline{\Delta RA} = -0.012$ arcsec and $\overline{\Delta DEC} = 0.027$ arcsec, and the median offsets are $\Delta RA_{50} = -0.026$ arcsec and $\Delta DEC_{50} = 0.016$ arcsec. The quantities ΔRA and ΔDEC can also be used to quantify the total separation Δ for each source:

$$\Delta = \sqrt{\Delta RA^2 + \Delta DEC^2}. \quad (2.17)$$

A histogram of the distribution of Δ is shown in Figure 2.16b, for the same 199 252 brightest sources as before. The distribution of Δ is strongly peaked, with the highest number of sources showing a separation in the 0.08–0.10 arcsec

²⁰Because the DES positions are based on the combined $r + i + z$ images, the position for the DES g -band is identical to the DES position in all $grizY$ filters.

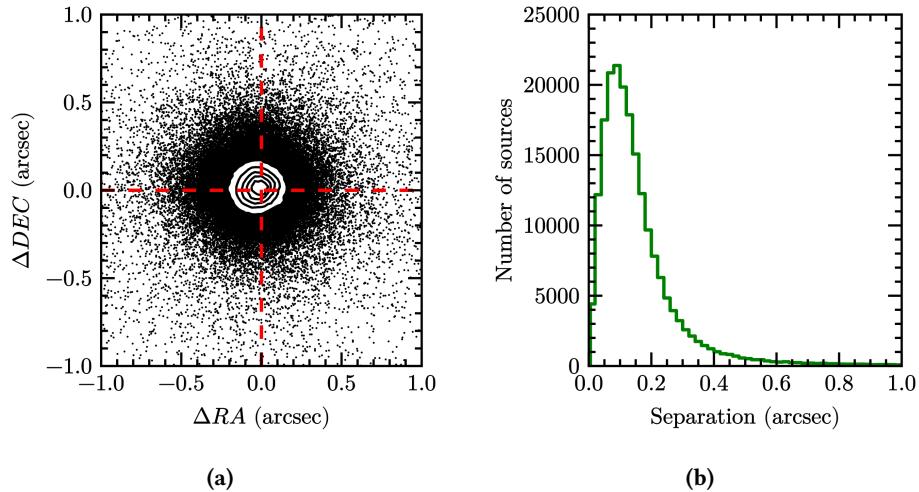


Figure 2.16 Comparison between DES and VIDEO astrometry for the 199 252 brightest non-duplicate objects, selected by imposing an SLR-corrected magnitude of $z_{\text{DES,AB}} < 22.0$. **(a)** Distribution of the difference ΔRA and ΔDEC between the two surveys as defined in Equations 2.15 and 2.16. The dashed red lines correspond to an offset of 0. **(b)** Histogram of the total separation $\Delta = \sqrt{\Delta RA^2 + \Delta DEC^2}$. Both (a) and (b) indicate excellent astrometric agreement.

bin. The mean and median values of Δ equal $\bar{\Delta} = 0.24$ arcsec and $\Delta_{50} = 0.12$ arcsec respectively. It is important to note that the mean and median for every quantity listed above is smaller than the 0.263 arcsec size of a single DES pixel. Together, all the results in this section therefore confirm that the astrometry of the DES and VIDEO surveys is indeed consistent.

2.3.2 Photometry verification

The next validation test compares the photometry between the DES+VIDEO and Jarvis et al. (2013) VIDEO catalogues. For this test, the non-duplicate DES+VIDEO detections were again matched to the Jarvis et al. (2013) objects within a 1 arcsec radius. The magnitude offset between the two surveys was then computed for all matched detections, using magnitudes measured in a ≈ 2 arcsec diameter aperture (note: the two catalogues differ slightly in the aperture sizes that have been used – 1.95 arcsec for DES+VIDEO and 2 arcsec for Jarvis et al. 2013). Figure 2.17 shows the resulting magnitude differences for each VIDEO filter as a function of magnitude. With mean offsets of < 0.022 mag AB in all filters, the two

catalogues are generally in agreement, especially when considering that the Jarvis et al. (2013) apertures are slightly larger. Despite this broad agreement, one may nevertheless note a slight bias towards bigger mean offsets (i.e. brighter Jarvis et al. 2013 magnitudes) for redder filters, ranging from 0.007 for the Z -band to 0.022 in the K_s -band. There are several possible explanations for this trend. Firstly, the difference in aperture sizes may play a role. Even though the mean VIDEO PSF actually has a larger FWHM at shorter wavelengths (see Section 2.1.2.4), it is possible that it has more extended wings at longer wavelengths, so that the (smaller) DES+VIDEO aperture catches less of the total flux in the redder filters. Secondly, the DES+VIDEO apertures for a given object have been placed in the same position for all bands (based on the $r + i + z$ detection image), whereas the positions in the Jarvis et al. (2013) catalogues have been derived from the morphology of the object in the reddest detected band (see Section 2.1.2.3). For extended sources, this morphology can differ between filters if some regions of a galaxy contain bluer emission than others. Therefore, it is possible that for these sorts of objects the DES+VIDEO apertures are less optimally positioned, so that the DES+VIDEO measurements miss some red flux that is included in the Jarvis et al. (2013) catalogues. Nevertheless, the fact that even in the K_s -band the mean offsets are no larger than 0.022 mag AB demonstrates that the general agreement is good. This warrants a good degree of confidence in the accuracy of the DES+VIDEO photometry.

2.3.3 Forced photometry verification and results

To assess the merits of the DES+VIDEO forced photometry strategy, the current section investigates whether the DES+VIDEO method indeed retrieves measurements for a larger number of sources than Jarvis et al. (2013). For this analysis, the DES+VIDEO source counts are compared to the number that would have been obtained via simple catalogue matching between the DESDM and Jarvis et al. (2013) catalogues. In order to estimate the latter number, the author assembled a position-matched dataset via the same catalogue assembly procedure as before (applying step 1, 3, 4, 6, and 8 from the algorithm presented in Section 2.2.5), this time using the Jarvis et al. (2013) catalogues instead of DES+VIDEO

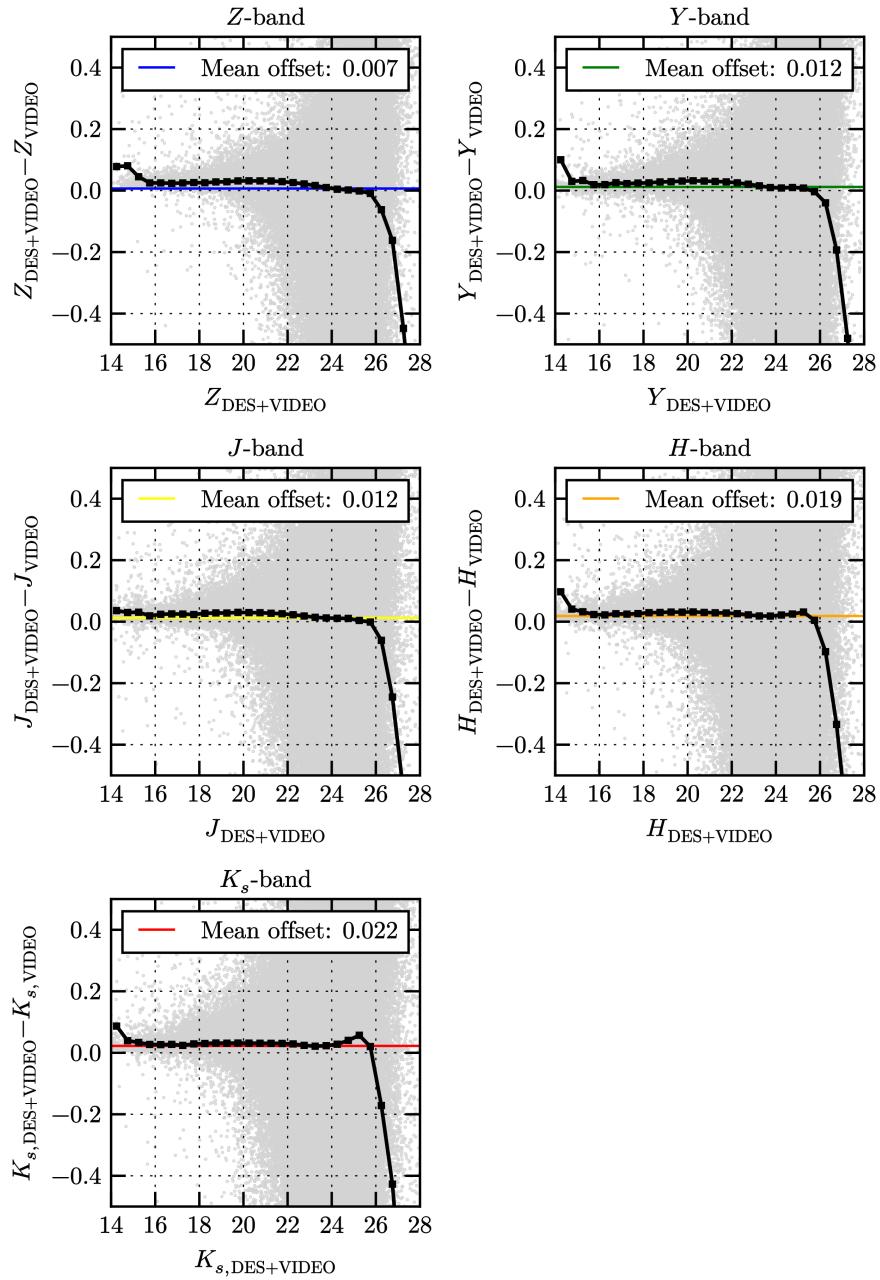


Figure 2.17 Comparison between VIDEO magnitudes from DES+VIDEO ($m_{\text{DES+VIDEO}}$) and Jarvis et al. (2013) (m_{VIDEO}), for all objects detected in both datasets. The grey dots show the magnitude difference $m_{\text{DES+VIDEO}} - m_{\text{VIDEO}}$ of each source as a function of DES+VIDEO magnitude. The black squares indicate the mean difference in bins of 0.5 mag. The coloured horizontal lines show the total mean value averaged over all grey data points, which is also listed in the legend. It must be noted that DES+VIDEO values have been measured in a 1.95 arcsec diameter aperture, as opposed to 2 arcsec for Jarvis et al. (2013).

forced photometry in step 4. Including duplicates²¹, position matching retrieved 1 862 154 objects, which is 24% less than the 2 443 576 sources obtained before via forced photometry. The unique (i.e. non-duplicate) total equals 1 676 054 with the Jarvis et al. (2013) catalogues, a 22% decrease compared to the 2 152 575 unique sources in the forced photometry dataset. These two figures indicate that DES+VIDEO forced photometry indeed significantly increases the amount of information extracted from the VIDEO survey. Even more gains were made in those regions where DES imaging is significantly deeper than VIDEO. For example, in the cdfs1 tile, forced photometry retrieves fluxes for 393 616 objects (including duplicates), compared to 257 812 from matching to Jarvis et al. (2013) – an improvement of $\approx 35\%$. Without duplicates, these numbers are 376 079 vs 246 755, implying an increase of $\approx 34\%$.

Further analysis of the number counts also shows that DES+VIDEO forced photometry extracts data for fainter sources than is possible with VIDEO alone. Figure 2.18a illustrates this finding by plotting the number of DES+VIDEO vs matched Jarvis et al. (2013) objects as a function of K_s -band magnitude, per deg² in bins of 0.5 mag. Like in the previous section, these magnitudes are measured in a ≈ 2 arcsec diameter aperture (1.95 arcsec for DES+VIDEO and 2 arcsec for Jarvis et al. 2013). Importantly, the DES+VIDEO number counts agree well with the Jarvis et al. (2013) results at the bright end, but come out considerably higher at the faint end as the Jarvis et al. (2013) distribution starts to roll over. This is exactly the behaviour that one would expect if DES+VIDEO forced photometry indeed extracts data from objects that are too faint to be detected reliably in VIDEO alone. To quantify this increase in faint source counts, the maximum of the K_s distribution was determined for both catalogues (grouping magnitudes in smaller bins of 0.01 mag for increased resolution). Without forced photometry, this peak²² lies at $K_{s,AB} = 23.635$, whereas the DES+VIDEO peak occurs at $K_{s,AB} = 24.065$. This indicates that the DES+VIDEO method pushes the VIDEO measurements ≈ 0.4 mag fainter (possibly even a bit more given the fact that the DES+VIDEO apertures are slightly smaller than those from Jarvis et al. 2013). The above analysis was repeated for the cdfs1 tile studied in the previous

²¹The new matched catalogue contains duplicates, originating from the same tile overlap pattern that exists in the DES+VIDEO catalogue. A non-duplicate version was also produced.

²²All the maxima in this paragraph are given without duplicates, the results including duplicates are shown in Figure 2.18 for completeness.

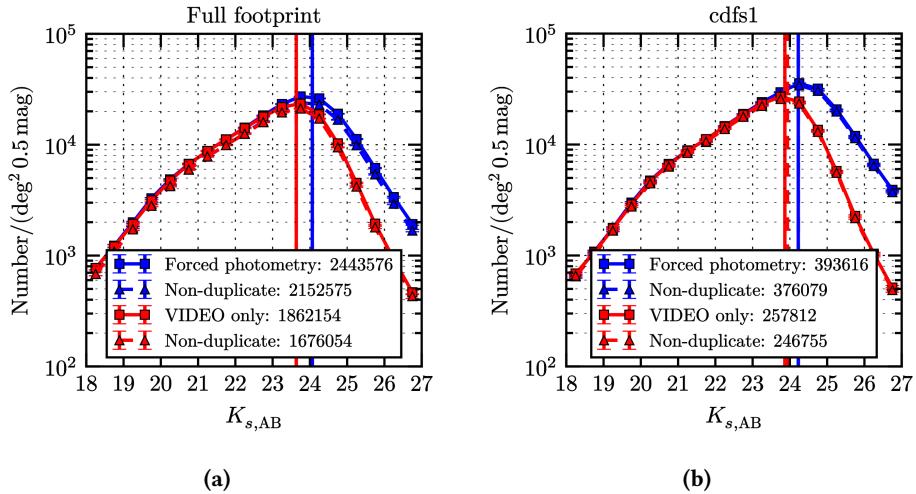


Figure 2.18 Number counts per deg^2 as a function of K_s -band magnitude, in bins of 0.5 mag. DES+VIDEO forced photometry (measured in 1.95 arcsec apertures) outcomes are shown in blue. The outcomes from position matching to the Jarvis et al. (2013) VIDEO catalogues (measured in 2 arcsec apertures) are shown in red. Solid lines with squares denote results including duplicates, and the dashed lines and triangles denote results after duplicates have been removed. All data points contain error bars corresponding to Poisson counting uncertainties, but these are smaller than the symbol size in almost all cases. Maxima for all distributions are computed in smaller 0.01 mag bins and denoted by vertical lines. **(a)** Results for the full DES+VIDEO footprint. The DES+VIDEO maxima both lie at $K_{s,\text{AB}} = 24.065$ (with and without duplicates), and the Jarvis et al. (2013) maxima are at $K_{s,\text{AB}} = 23.635$ (idem). **(b)** Results for the cdfs1 tile only. The DES+VIDEO maxima lie at $K_{s,\text{AB}} = 24.225$ (with and without duplicates), and the Jarvis et al. (2013) maxima are at $K_{s,\text{AB}} = 23.865$ with duplicates and at $K_{s,\text{AB}} = 23.935$ without duplicates.

paragraph. These results are plotted in Figure 2.18b. Here, the Jarvis et al. (2013) maximum occurs at $K_{s,\text{AB}} = 23.935$, while the DES+VIDEO distribution peaks at $K_{s,\text{AB}} = 24.225$, i.e. ≈ 0.3 mag fainter. This is slightly less than the ≈ 0.4 mag increase for the full footprint, most likely due to the fact that the K_s data in this tile is slightly deeper than average (see Table 2.8). Because Jarvis et al. (2013) based their source detections on the reddest filter in which a given object is detected (which is often the K_s -band), in this tile their detection strategy performs slightly above average. Therefore, the DES+VIDEO forced photometry provides a comparatively smaller increase in faint K_s -observed sources. One may recall from the previous paragraph that forced photometry does, however, perform above average in the cdfs1 tile in terms of overall source counts (when compared to the full footprint). This can be attributed to the fact that the Z -band imaging is shallower than average in this tile, which means that the corresponding Jarvis et al. (2013) catalogue contains fewer ‘blue’ objects that are undetectable in $YJHK_s$ but would nevertheless show up in deeper Z -band imaging. At the same time, the DES z -band is much deeper than VIDEO in this region, so the DES+VIDEO pipeline does manage to retrieve these blue objects.

The fact that forced photometry can be used to obtain photometry for more and for fainter objects has been shown several times before (e.g. Banerji et al. 2015; Lang et al. 2016; Nyland et al. 2017). In fact, the very similar study by Banerji et al. (2015) demonstrated this principle for the DES wide-field (detection) and VHS (measurement) data. Exploiting the fact that their DES data is considerably deeper than VHS (~ 4 mag AB when comparing i - and J -band magnitudes), they found that forced photometry increases the number of objects with measured VHS fluxes by a factor of 4.5 and pushes the VHS 80% completeness limit 1.5 mag deeper. Comparatively, the 1.22 times increase in numbers and 0.4 mag increase in depth found by this thesis are not quite as enormous, since the DES deep field and VIDEO surveys are more evenly matched in terms of limiting magnitudes (< 2 mag AB difference between i and J). However, the DES+VIDEO results nevertheless add to the existing body of evidence in favour of using forced photometry to combine surveys. On top of that, they also demonstrate that doing so can lead to considerable gains even if one survey is only moderately deeper than the other.

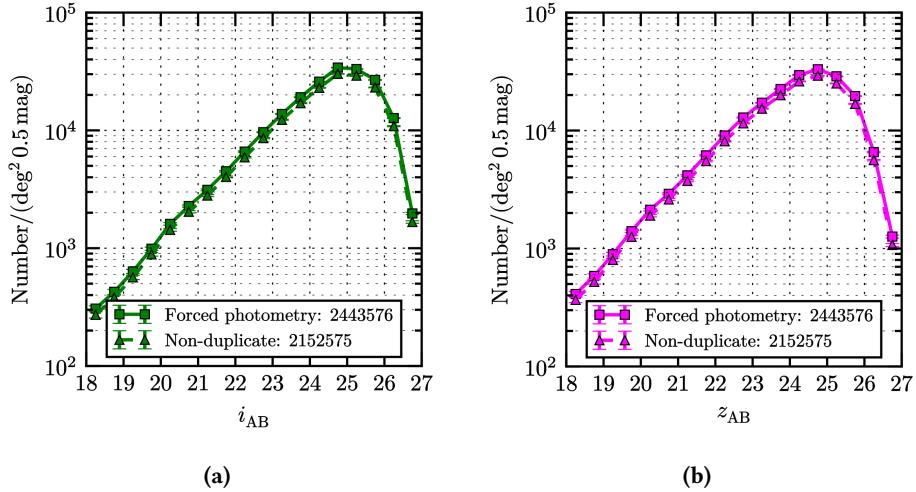


Figure 2.19 Number counts for the DES+VIDEO catalogue per deg^2 as a function of 1.95 arcsec magnitudes in bins of 0.5 mag. The solid lines with squares show results including duplicates, and the dashed lines and triangles show results after duplicates have been removed. Data points for each bin include error bars corresponding to uncertainties from Poisson counting, but these are smaller than the symbol size in almost all cases. **(a)** Results for the DES i -band. **(b)** Results for the DES z -band.

2.3.4 Catalogue results

The validation tests above confirm that the astrometry and photometry of the DES+VIDEO catalogue are indeed accurate, and that the forced photometry is functioning as intended. This establishes the catalogue as a suitable general resource for the scientific community, and demonstrates that it can provide a solid backbone for the photometric redshifts and high-redshift galaxy search later in this thesis.

For the sake of providing a comprehensive overview of the full catalogue, Figure 2.19 shows the DES+VIDEO number counts in the i -band and z -band as a function of magnitude. The distributions behave as expected from the area-depth distributions in Figure 2.7. The z -band distribution peaks roughly $z_{\text{AB}} \approx 24.75$, which approximately corresponds to the $z_{5\sigma} \geq 24.9$ magnitude limit for the largest 10.8 deg^2 section of the total footprint. The number counts then decline slightly with increasing magnitude, falling sharply after $z_{\text{AB}} \sim 25.75$. This turning point is reasonably close to the $z_{5\sigma} \geq 26.2$ limiting magnitude of the deepest 4.3 deg^2 . Similarly, the maximum of the i -band distribution lies around $i_{\text{AB}} \approx 25.0$, in agree-

ment with the limiting $i_{5\sigma} \geq 25.1$ limiting magnitudes in the largest 10.7 deg^2 i -band area. The second turnover at $i_{\text{AB}} \approx 26.0$ again lines up approximately with the $i_\sigma \geq 26.3$ depths in the deepest 4.4 deg^2 i -band regions of the catalogue.

Chapter 3

Photometric redshifts

3.1 Motivation

A key result of Chapter 2 has been the production of a DES+VIDEO source catalogue that covers 10.8 deg^2 to depths of $z_{5\sigma} \geq 25.0$ and a smaller sub-area of 4.3 deg^2 to $z_{5\sigma} \geq 26.2$. At the time of writing, this dataset contains the largest footprint to comparable depths over optical and near-IR wavelengths. This makes the DES+VIDEO catalogue a highly attractive resource to search for the brightest ($m_{AB} < 25.0$) galaxies at $z \gtrsim 5$. As explained in Section 1.3, the study of galaxy evolution in the early universe would benefit greatly from the discovery of more of these rare sources, and finding extremely bright high-redshift galaxies is therefore one of the core aims of this thesis. Due to their rarity, no more than a few dozen of these objects are expected within the DES+VIDEO footprint (as will be argued later in Section 4.3.1.1). This number is evidently small when compared to the catalogue total of 2 152 575 unique sources, which indicates that the search will require an efficient and clean selection method. Section 1.3.1 described several effective methods that have been developed in the literature, of which Lyman-break colour selection and photometric redshifts are the most appropriate for the current dataset. Even though both strategies have historically been highly successful, this thesis will opt for a photometric redshift approach. The main reason behind this choice is that for the purpose of selecting candidates, colour-colour methods only use the information from the three bands used in the colour cuts, whereas photometric redshifts incorporate data from all available wavelengths. Given that the DES+VIDEO catalogue covers the full

optical+near-IR range in 10 filters of which 9 are deep enough for the photo-zs in this thesis, a photometric redshift approach makes better use of all the available information. Furthermore, as described in Section 1.2.2, some photometric redshift methods offer sophisticated probabilistic data handling options, providing another edge over simple colour-cut methods. Features such as the inclusion of a redshift-luminosity prior (see Section 3.2.2) or probability distribution (see Section 3.2.1.1), can be used to remove objects consistent with a low-redshift solution, in order to increase the purity of a high-redshift sample.

In addition to the fact that photometric redshifts offer a more sophisticated treatment of the available flux information, they also deliver redshifts for all objects, instead of only selecting a particular high-redshift sample. This is an important asset for the DES+VIDEO catalogue. Given its utility as a general resource for the scientific community (as proposed in Sections 1.1.5 and 1.2.4), equipping the DES+VIDEO catalogue with photometric redshifts for all 2 443 576 objects will be of general interest. Moreover, because 35 596 sources have spectroscopic redshifts, their photo-zs can also be used for a case-study of photometric redshift performance. As explained in Sections 1.2.3 and 1.2.4, because photometric redshifts play a crucial role in modern astronomy, such performance tests are highly important, and therefore form another aim of this thesis.

The task of calculating photometric redshifts for the DES+VIDEO catalogue is the topic of the current chapter. As part of that process, several photo-z options and features are tested in terms of their performance. At the end, the most suitable configuration is selected to calculate the final DES+VIDEO redshifts. As a concluding remark, it is helpful to emphasise the priorities that have informed some of the choices involved here. Despite the secondary goal of providing photometric redshifts for general use, it must be stressed that the final DES+VIDEO redshift catalogue is primarily created with the selection of $z \gtrsim 5$ objects in mind. Thus, the tweaking of the photometric redshift code is aimed towards producing optimal results for high-redshift galaxies, possibly at the cost of fine accuracy for the bulk of the objects at $z \lesssim 1.5$. This is a conscious choice, and anyone who wishes to use the photometric redshifts for low-redshift research must bear this caveat in mind.

3.2 Description and implementation of the photometric redshift algorithm

3.2.1 Choice of code

With a strong case established in favour of calculating photometric redshifts for the DES+VIDEO dataset, the next step involves selecting a particular code from the large suite of publically available codes described in Section 1.2.2. To recap briefly, these codes can be grouped into two categories: template fitting methods and empirical methods. Empirical methods rely on spectroscopic redshifts to calibrate an observed colour-redshift relationship empirically, and therefore require a large and representative set of spectroscopic redshifts to function. There is evidence that empirical codes outperform template fitting methods when such a set is available (Abdalla et al. 2011; Salvato et al. 2019), and that template codes are more suited to regimes where this is not the case, such as at high redshift (Salvato et al. 2019). As demonstrated in Section 2.1.3, the spectroscopic coverage in the DES+VIDEO catalogue is sparse at high redshifts, with only 11 objects at $z_{\text{spec}} > 4.0$. Due to these extremely small numbers, empirical codes will not be reliable in this redshift range. Since optimal performance at $z \geq 5$ is the main priority of the photometric redshift catalogue, a template fitting code is deemed to be the most suitable option.

As described in Section 1.2.2.3, studies in the literature comparing different codes have failed to produce a clear consensus on a single best template code (Hildebrandt et al. 2010; Abdalla et al. 2011; Dahlen et al. 2013). The choice of code for this thesis was therefore informed primarily by what extra features the different codes provide. It was decided to use **LePHARE** (Ilbert et al. 2006, 2009), predominantly on the basis that it offers a way to calibrate magnitude offsets within photometric bands (a detailed description of this feature and how it benefits the DES+VIDEO catalogue will be presented in Section 3.2.3). The fully probabilistic Bayesian approach is also advantageous, especially because it permits the option to include a luminosity prior (see Section 3.2.2), and can make use of a redshift probability distribution function $P(z)$ (see Section 3.2.1.1). Furthermore, **LePHARE** is a well established code in the literature, and the fact that it has been used in many successful studies demonstrates its quality. For instance, the highly accurate multiwavelength photometric redshift catalogue in

the COSMOS field (Ilbert et al. 2009; previously introduced in Section 1.1.4.2) was produced with **LePHARE**. Jarvis et al. (2013) also noted good performance when applying **LePHARE** to an optical+near-infrared dataset very similar to the DES+VIDEO catalogue. Another significant case in favour comes from a comparative analysis of 19 photometric redshift codes by Hildebrandt et al. (2010), who placed **LePHARE** in the top 5 in terms of performance, showing the lowest scatter and outlier fraction. The latter in particular is important for reducing low-redshift contamination in the search for high-redshift galaxies.

Having settled on **LePHARE** as the choice of code, the remainder of this section will present the way that this piece of software computes photometric redshifts, along with several of its key features and options.

3.2.1.1 χ^2 minimisation

LePHARE operates using a standard χ^2 minimisation algorithm, which has long been a favoured procedure for template fitting codes (see Section 1.2.2.2 and e.g. Puschell et al. 1982, Bolzonella et al. 2000, Brammer et al. 2008). In very rough terms, these χ^2 methods work by comparing the observed colours to those of templates at various redshifts, to find which redshift produces the best fit to the data.

The following paragraphs present in detail the χ^2 minimisation algorithm as implemented in **LePHARE**. Firstly, the user specifies a set of predicted template SEDs. Each template SED T is redshifted between¹ $z_{\min} = 0.0$ and $z_{\max} = 9.0$ in steps of $\delta z = 0.04$. The code then convolves the SEDs at each redshift with the filter transmission curves (supplied by the user) to produce a predicted template flux $F_{T,z,i}$ for each template T , redshift z , and filter i . A redshift-dependent amount of Lyman- α absorption by the intergalactic medium² is accounted for following Madau (1995). If specified, dust extinction and emission lines are also applied to the predicted template magnitudes (see Sections 3.3.3 and 3.3.4 for a further explanation of these features). Using the observed fluxes $F_{\text{obs},i}$ and errors $\sigma_{\text{obs},i}$ in each filter i for a particular observed source, one can define the χ^2 statistic:

¹The values of z_{\max} and δz can be altered by the user through the **Z_STEP** parameter in the configuration file; the values listed here were used in this thesis.

²This absorption corresponds to the Lyman- α forest previously described in Section 1.3.1.2.

$$\chi^2(z, T, A) = \sum_i^{N_f} \left[\frac{F_{\text{obs},i} - AF_{T,z,i}}{\sigma_i} \right]^2, \quad (3.1)$$

where A is a normalisation factor, and N_f is the number of observed filters. For a given object, this χ^2 is a measure of the discrepancy between the observed fluxes and the predicted fluxes from a template SED T at a redshift of z with normalisation A . Lower values of χ^2 indicate a better fit to the data. Using a slightly involved procedure detailed below, **LePHARE** then varies the free parameters T , z , and A , and estimates the best-fit photometric redshift from the combination of parameters that minimises the χ^2 value.

When varying the free parameter z , the code saves the minimal χ^2 value for each increment of z (marginalising over T and A), and converts these values into a maximum likelihood function $F(z)$:

$$F(z) = \exp[-\chi^2_{\min}(z)/2]. \quad (3.2)$$

This maximum likelihood function $F(z)$ is directly proportional to the redshift probability function $P(z)$, which can simply be obtained from $F(z)$ by normalising. In other words, $P(z) \propto F(z)$ and $P(z) = F(z) / \int F(z) dz$. Summarising the previous line of reasoning yields:

$$P(z) \propto \exp[-\chi^2_{\min}(z)/2]. \quad (3.3)$$

At this stage it is important to mention that if a luminosity prior is used, the right hand side of Equations 3.2 and 3.3 will be weighted by the prior probability. This weighting process will be further introduced and mathematically justified in Section 3.2.2.

LePHARE then uses the maximum likelihood function $F(z)$ — or in the case where a prior is used, the prior-weighted version³ — to compute several types of output redshifts:

³Considering the properly normalised Bayesian reasoning in Section 3.2.2, it is technically more consistent to frame the concepts in this section in terms of prior probability $P(z)$, rather than the unnormalised $F(z)$. However, the use of ‘maximum likelihood’ was retained for the sake of maximal overlap with the terms used in the **LePHARE** documentation. Fortunately, as the overall normalisation factor does not affect the redshift output results, this issue is merely technical.

Z_BEST is the best redshift estimate, based on a parabolic interpolation of the maximum likelihood $F(z)$ between steps δz (following Bevington 1969).

Z_BEST_68_LOW and **Z_BEST_68_HIGH** are the lower and upper limits on the 68% confidence interval for **Z_BEST**. Using the notation z_{lim} to represent either **Z_BEST_68_LOW** or **Z_BEST_68_HIGH**, these quantities are found by requiring that $\Delta\chi^2 = |\chi^2(z_{\text{lim}}) - \chi^2(z_{\text{best}})| = 1.0$.

Z_ML is the median of the maximum likelihood distribution $F(z)$. The associated 68% confidence limits **Z_ML_68_LOW** and **Z_ML_68_HIGH** are again derived by requiring that $\Delta\chi^2 = |\chi^2(z_{\text{lim}}) - \chi^2(z_{\text{ML}})| = 1.0$.

Z_SEC provides the best redshift estimate for a potential secondary solution in cases where the $F(z)$ distribution has several maxima (for an example of such a source see Figure 4.4). The probability threshold above which such a secondary solution will be calculated can be specified with the **MIN_THRES** parameter, which was set to 0.1 in this thesis.

This thesis uses the values of **Z_BEST** as the final photometric redshifts. The uncertainties σ_z on these redshifts can be calculated using the 68% confidence limits: $\sigma_z = (Z_{\text{BEST_68_HIGH}} - Z_{\text{BEST_68_LOW}})/2$.

3.2.1.2 Procedure for missing data

It is common for photometric catalogues to have varying waveband coverage, as some fields in the surveys may not (yet) be imaged in every available waveband. Section 2.1.2.3 described how the VIDEO part of the DES+VIDEO catalogue is an example of such a dataset. To handle this common issue, **LePHARE** can recognise objects with missing data if the user has set the input fluxes *and* errors in the unobserved filters to -99.0. The code then excludes these filters from the χ^2 minimisation in Equation 3.1. In this way, unobserved data points can be distinguished from entries where an object is simply undetected in some wavebands, since non-detections have negative fluxes but positive flux errors. This solid treatment of the distinction between missing data and non-detections is very important when searching for $z \gtrsim 5$ galaxies. As explained in Section 1.3.1.2, those sources are often undetected in the bluest filters due to their strong Lyman breaks.

3.2.1.3 Reduced χ^2

Because the number of available filters N_f can vary from object to object, sources with more available filters will have higher average values of χ^2 , since their χ^2 sum in Equation 3.1 includes more terms. To allow for direct goodness-of-fit comparisons between sources with different values of N_f , it is common to capture the best fit in terms of a so-called reduced χ^2 (written as χ_v^2). For each source, this quantity is defined as follows:

$$\chi_v^2 \equiv \frac{\chi^2}{v}. \quad (3.4)$$

Here, χ^2 is the best-fit value for the source in question as given by Equation 3.1. v encapsulates the number of degrees of freedom, which is the number of input data points minus the number of free parameters in the fits. For the photometric redshift fitting in this thesis, there are N_f data points and three free parameters – namely z , T , and A . Therefore, $v = N_f - 3$, which results in the following formula for χ_v^2 :

$$\chi_v^2 = \frac{\chi^2}{N_f - 3}. \quad (3.5)$$

3.2.2 $N(z)$ prior

As a template fitting method, the χ^2 minimisation algorithm in **LePHARE** inherently suffers from colour-redshift degeneracies, an issue previously described in Section 1.2.2.2. A common cause of error is confusion between the Lyman break and the 4000 Å break, which can lead to catastrophic ($|z_{\text{phot}} - z_{\text{spec}}| > 1$) errors in the photometric redshifts (Ilbert et al. 2006). While near-IR photometry can help to break these degeneracies (Benítez 2000; Ilbert et al. 2006), it is important to investigate further methods to address the problem.

In an attempt to break colour-redshift degeneracies, **LePHARE** contains the option to include a so-called $N(z)$ prior in the fitting procedure. The underlying idea behind this prior is to adjust the photometric redshifts according to pre-existing information from a general galaxy redshift distribution $N(z)$ that has been previously measured. To understand this, let us reconsider the probability function $P(z)$ for a given source found from χ^2 minimisation in Equation 3.3. For each redshift z , it is possible to weight this $P(z)$ by the likelihood that the

source is at that particular redshift *given its magnitude*, because this likelihood can be estimated from the known distribution of galaxies $N(z)$ with that magnitude (for example, bright galaxies are known to be uncommon, and therefore unlikely, at high redshifts). Bayesian probability theory provides a way of formalising these ideas mathematically. In this framework, the weighting likelihood described above is called a prior probability (or simply prior). This factor will generally⁴ break colour-redshift degeneracies, by indicating that one of the possible redshifts is more likely given known data.

The use of priors in photometric redshift fitting was first introduced by Benítez (2000), who provided a detailed description of the statistical foundation for this approach within Bayesian probability theory. To grasp the core reasoning behind his approach, start by considering a single source with observed colours C and observed magnitude m_0 . For this source, $P(z, T|C, m_0)$ represents the probability that it lies at redshift z and has a spectral type T . Under the assumption that the set of possible spectral types is complete, summation over all spectral types generates the probability that the source is at redshift z , given the observed C and m_0 :

$$P(z|C, m_0) = \sum_T P(z, T|C, m_0). \quad (3.6)$$

The quantity $P(z|C, m_0)$ is the desired end product, i.e. the prior-weighted probability that the code will use to compute the final photometric redshifts. To deconstruct $P(z, T|C, m_0)$, Bayes' theorem can be applied:

$$P(z, T|C, m_0) = \frac{P(z, T|m_0)P(C|z, T)}{P(C)}, \quad (3.7)$$

which when inserted into Equation 3.6 trivially yields:

$$P(z|C, m_0) = \sum_T \frac{P(z, T|m_0)P(C|z, T)}{P(C)}. \quad (3.8)$$

Here $P(C)$ is a normalisation factor, and $P(C|z, T)$ is the probability of the observed colours C for a given redshift and spectral type. The latter quantity corresponds to the original, unweighted $P(z)$ from Equation 3.3. The quantity

⁴Strictly speaking, it will naturally only break the degeneracy if the information contained within the prior indicates that one of the degenerate redshifts has a higher probability.

$P(z, T|m_0)$ in Equation 3.8 is referred to as the prior probability and expresses the probability that the observed source has a certain redshift and spectral type given its observed magnitude.

To implement the Bayesian framework into **LePHARE**, Ilbert et al. (2006) devised a functional form for the prior by parameterising $P(z, T|m_0)$ as a function of z , T , and several free parameters. Following a computational procedure developed by Benítez (2000), they calibrated the free parameters by fitting that equation to the observed redshift distribution for each spectral type $N_T(z)$ in a deep and representative I -selected spectroscopic sample from the VVDS (Le Fèvre et al. 2005). In the end, the resulting $N(z)$ prior supplied within **LePHARE** captures the distribution of $B - I$ colour that they found in this dataset. Further details of the functional form for the prior they derived can be found in Ilbert et al. (2006).

Several studies have demonstrated that the use of priors enhances photometric redshifts (Benítez 2000; Ilbert et al. 2006; Tanaka 2015). Improvements have been observed particularly strongly in the outlier fraction, which is the fraction of objects with significantly poor⁵ photo-zs, as well as in the scatter. For instance, when applying a prior, Tanaka (2015) found a reduction in outliers from 45% to 25% and a lowering of the scatter by a factor of 0.45. Benítez (2000) also found that using a prior reduces outliers, specifically spurious redshift estimations from photometric noise. Minimising the number of outliers is very important for the high-redshift search later in this thesis. Because of the comparative abundance of low-redshift objects compared to true high-redshift candidates in the DES+VIDEO catalogue, even a small increased probability of catastrophic outliers can send a considerable number of contaminating low-z objects into $z \gtrsim 5$ redshift bins. A prior is therefore likely to be useful for this thesis.

3.2.3 Adaptive offsets

Photometric surveys are known to contain small systematic uncertainties in the absolute calibration of photometric zero-points, expected to be of the order of ~ 0.05 mag (Brodwin et al. 2006; Ilbert et al. 2006). Several studies have attempted to quantify the size of these offsets in various surveys (Brodwin et al. 2006; Ilbert et al. 2006, 2009; Dahlen et al. 2013). To do this, they applied template fit-

⁵The exact definition of what is considered a bad redshift estimate varies between publications.

ting to samples of objects with known spectroscopic redshifts, and measured the residual difference between the observed magnitude and the magnitude of the best-fitting template for each source (fixing the redshift to the measured spectroscopic value). All found non-negligible discrepancies in most filters, generally of the order of several 0.01 mag — comparable to the expected value of the zero-point calibration offsets (although these residuals can also trace other factors that will be touched upon shortly). These magnitude residuals can lead to inaccuracies in photometric redshifts (Ilbert et al. 2006, 2009). In their comprehensive multiwavelength study of the COSMOS field, Ilbert et al. (2009) found that this effect was of the same order as the photo-z precision. It is therefore important to account properly for any magnitude residuals.

For this reason, **LePHARE** contains the option to calculate so-called adaptive offsets, which provide magnitude corrections that can be applied to the photometry. The code calibrates these offsets on a training sample with measured spectroscopic redshifts, using a subroutine described in Ilbert et al. (2006) and summarised below. First of all, for each galaxy g with spectroscopic redshift z_{spec} in the training sample, **LePHARE** uses Equation 3.1 with a fixed value of $z = z_{\text{spec}}$ to find the statistic $\chi^2(T, A)$. Minimising $\chi^2(T, A)$ yields a best-fit template T and corresponding scale factor $A_{T,g}$. Then, the algorithm computes the predicted fluxes $F_{T,g,i}$ for this template in a given filter i , which it uses to compute the total combined flux discrepancy ψ_i^2 for all galaxies:

$$\psi_i^2(s_i) = \sum_g^{N_{\text{gal}}} \left[\frac{A_{T,g} F_{T,g,i} - F_{\text{obs},i} + s_i}{\sigma_{\text{obs},i}} \right]^2. \quad (3.9)$$

Here, N_{gal} is the number of galaxies in the spectroscopic subsample, $F_{\text{obs},i}$ and $\sigma_{\text{obs},i}$ are the observed flux and flux uncertainty within the filter i , and s_i is a free parameter which corresponds to the adaptive (flux) offset for that filter. s_i is found by determining the value that minimises ψ_i^2 . Because the best-fit template may change after incorporating these flux offsets, **LePHARE** then adds the corrections s_i to the observed fluxes and repeats the entire process described above, redoing the template fitting and recalculating all the values of s_i . This iterative process is repeated until s_i converges in every filter. After converting to magnitudes, the code outputs these final values as the adaptive offset corrections that can be applied to the observed photometry.

In the event of randomly distributed flux errors, the values of the adaptive offsets are expected to be (close to) zero, but in the case of systematic offsets in the photometry they will have a non-zero value. Typical values in the literature are generally of the order of several 0.01 mag, but rarely exceed 0.1 (Ilbert et al. 2006, 2009; Dahlen et al. 2013). It is important to note that while such non-zero offsets mostly tend to reflect errors in the calibration of survey zero-points, other factors can also contribute. These include band-to-band variation in the seeing, and uncertainties in the predicted template colours, for instance through inaccurate filter transmission curves, incomplete template sets, or an incorrect extinction curve. Regardless of the exact origin, the adaptive offsets capture real systematics that ought to be accounted for. Several studies have demonstrated the effectiveness of including offset corrections, indicating that they can reduce the photo-z bias (Ilbert et al. 2009) as well as the scatter and outlier fraction (Dahlen et al. 2013). It is emphasised that correcting for systematic photometric offsets is particularly important for catalogues that combine data from several surveys (such the DES+VIDEO catalogue presented in this thesis), due to the fact that different surveys can use different zero-point calibration processes. Furthermore, because of the difficulties described in Section 2.2.5.2 regarding DES+VIDEO aperture corrections, adaptive offsets are expected to provide an additional considerable benefit for this thesis; since they account for any systematic flux variation within a filter, the offsets essentially provide a rough empirical aperture correction of sorts.

3.3 Templates

The SEDs with which **LePHARE** performs its template fitting must be specified by the user. This section presents a brief description of the templates that are used in thesis, and of the way that the code applies dust extinction and emission lines to the template magnitudes.

3.3.1 Galaxy templates

The default installation of **LePHARE** is equipped with a wide range of common galaxy template sets. For this thesis, two sets were chosen from this selection. The discussion below will introduce both sets of SEDs, and present the reasons why those were considered suitable.

3.3.1.1 **AVEROI_NEW** templates

The collection of templates referred to as **AVEROI_NEW** in the **LePHARE** documentation is a set of 62 empirical galaxy SEDs, assembled by Arnouts et al. (2007). The templates are based on the original four observed CWW spectra (Coleman et al. 1980) with the addition of two starburst spectra from Kinney et al. (1996). The original CWW set consists of optical spectra for elliptical (Ell), spiral (Sbc, Scd) and irregular (Irr) spectral types, which Arnouts et al. (2007) extrapolated into ultraviolet and infra-red wavelengths using the GISSEL synthetic models (Bruzual & Charlot 2003). They then linearly interpolated these six spectra to create a final set of 62 SEDs, which they optimised using a sample of spectroscopic redshifts (following Ilbert et al. 2006). A representative subset of the **AVEROI_NEW** templates is plotted in Figure 3.1a.

The CWW spectra that form the basis of the **AVEROI_NEW** SEDs are a common choice of template among photo-z studies in the literature (e.g. Fernández-Soto et al. 1999; Arnouts et al. 1999), especially when augmented by starburst templates (Sawicki et al. 1997; Brodwin et al. 2006; Ilbert et al. 2006). For the CWW+starburst sets in particular, this choice enjoys strong observational support, as the combination of these templates has been shown to span the range of galaxy properties observed in the Hubble Deep Field over a wide redshift range of $0 < z < 6$ (Benítez 2000; Fernández-Soto et al. 2002). This completeness well into the high-redshift regime suggests that the **AVEROI_NEW** templates make a highly appropriate choice for this thesis.

3.3.1.2 **COSMOS** templates

The **COSMOS** template set consists of 31 theoretical SEDs assembled by Ilbert et al. (2009) for their photo-z study of the COSMOS field. Nine of these templates were initially created by Polletta et al. (2007), who used the GRASIL code (Silva et al. 1998) to generate templates that fitted the observed UV to mid-IR broadband colours of spectroscopically confirmed objects in the VVDS survey (Le Fèvre et al. 2005). Three of these spectra represent elliptical types and six are spiral types (S0, Sa, Sb, Sc, Sd, Sdm). Ilbert et al. (2009) linearly interpolated these nine SEDs to produce 19 templates. Because this collection does not adequately sample the blue end of galaxy colour space, Ilbert et al. (2009) generated 12 additional templates using Bruzual & Charlot (2003) models with starburst ages from 0.03

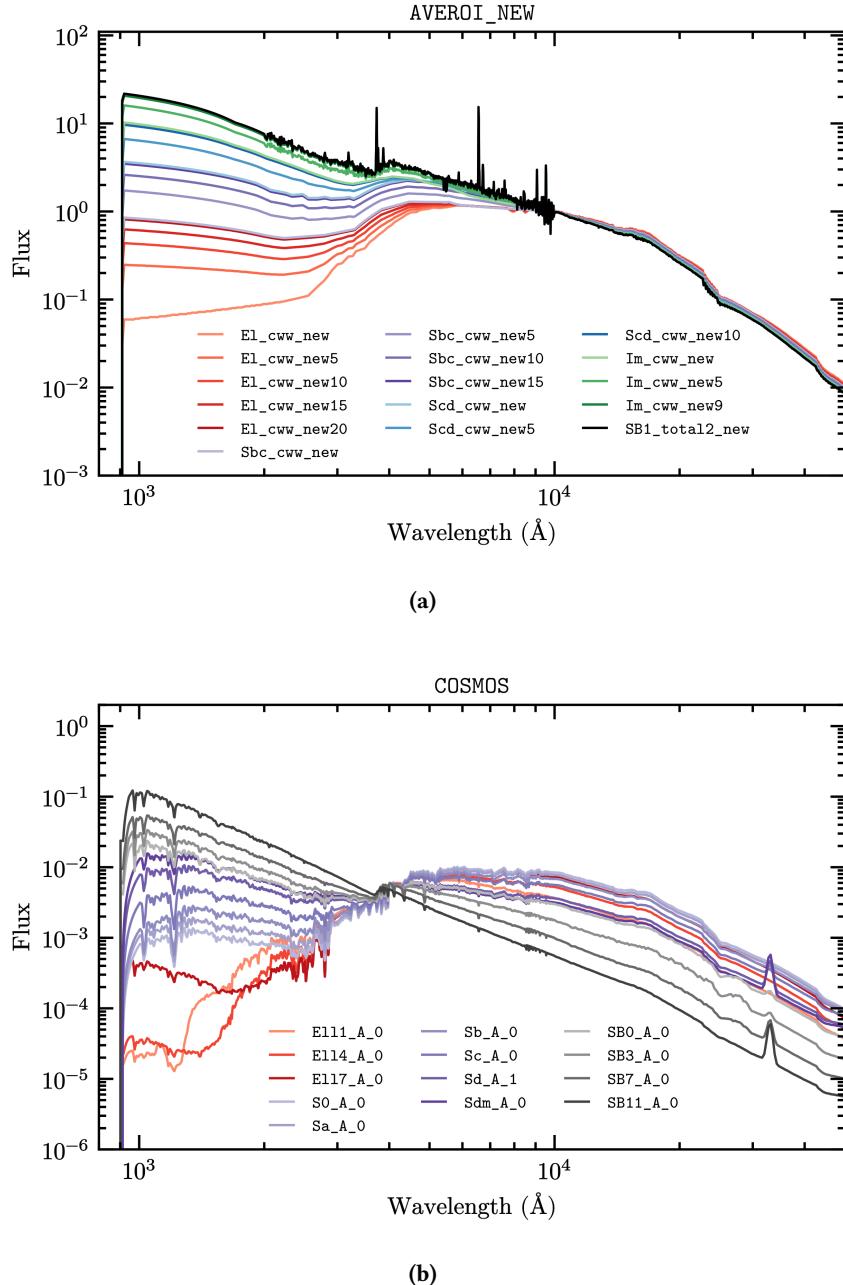


Figure 3.1 Spectra for the two sets of templates used to calculate photometric redshifts for the DES+VIDEO catalogue. The plots display a subset of templates covering a representative fraction of the full template set for (a) the **AVEROI_NEW** templates and (b) the **COSMOS** templates. The names of these templates have been copied directly from the **LePHARE** installation. Broadly following the Hubble classification scheme for galaxy morphologies, spectra beginning in **E1** or **E11** correspond to elliptical galaxies, **Im** to irregular galaxies, **SB** to starburst galaxies, **SO_A_0** to a lenticular galaxy, and the other names starting in **S** to various forms of spirals.

to 3 Gyr, and extended these beyond $3 \mu\text{m}$ using the Sdm Polletta et al. (2007) template. Together, all 31 SEDs form the **COSMOS** template set. Figure 3.1b shows a subset of these spectra.

Because the **COSMOS** SEDs have been selected to fit the UV-mid-IR colours of an observed sample, Ilbert et al. (2009) claim that they provide a better joining of the UV and mid-IR than template sets based on the CWW spectra (such as the **AVEROI_NEW** templates), which have been extrapolated from the optical theoretically. For this reason, it was deemed of interest to this thesis to test the **COSMOS** templates as well. Moreover, the **COSMOS** SEDs have also produced noteworthy observational success. As described above, they were used by Ilbert et al. (2009) in the production of their COSMOS 30-band photometric redshift catalogue. Section 1.1.4.2 previously mentioned that this catalogue is extremely accurate, with a dispersion of only $\sigma_{\Delta z}/(1+z_{\text{spec}}) = 0.007$ (for the brightest $i_{\text{AB}}^+ < 22.5$). Ilbert et al. (2009) also demonstrated a 3–5 times improvement in accuracy compared to the previous 16-band COSMOS catalogue (Mobasher et al. 2007) and CFHTLS-Deep results (Ilbert et al. 2006), which were both based on modified versions of the CWW templates. Even though the exceptional accuracy is almost certainly for a large part thanks to the remarkable 30-band coverage, it also corroborates the quality of the **COSMOS** templates, and suggests that these may be an suitable choice for this thesis. The indicated improvement over CWW+starburst template sets is also interesting, and warrants a comparative investigation with the **AVEROI_NEW** performance using the DES+VIDEO dataset.

3.3.2 Stellar and QSO templates

In addition to galaxy SEDs, **LePHARE** also determines the best-fit solutions for stellar and quasar (QSO) templates. This feature can be useful, because a comparison of the best-fit χ^2 for each SED type can identify whether the observed colours of a given source are most consistent with a galaxy, a star, or a quasar. Because stellar sources will have nonsensical photo-zs, they can contaminate photometric redshift catalogues, so it is important to have an effective method for differentiating between stars and galaxies. This task will be addressed in Section 3.7.1, which will describe the process of finding a suitable star-galaxy separation method based on the SED fitting with **LePHARE**. Rejecting stars and low-redshift QSOs is also be important to the high-redshift galaxy search, so this thesis makes extensive use of the template fitting feature in Chapter 4 as well.

To make sure that **LePHARE** can fit stars accurately, it is important to include a good selection of stellar SEDs. For this reason, it was decided to use the broadest set of stars contained in the default **LePHARE** installation. This collection consists of a total of 254 templates, composed of 131 empirical star SEDs from Pickles (1998), 4 observed white dwarf spectra from Bohlin et al. (1995), 18 theoretical models of late-type stars, and a collection of 99 low mass stars and brown dwarfs from Chabrier et al. (2000). Having a large selection of SEDs to cover the colours of low mass stars as comprehensively as possible is especially important for this thesis, since low mass stars and brown dwarfs are known to be major contaminants of $z \gtrsim 5$ candidates (McLure et al. 2009; Willott et al. 2013; Bowler et al. 2015; Finkelstein 2016).

The QSO templates used in this thesis are based on models by Salvato et al. (2009), and are taken from a collection labelled **QSO_MARA** in the default **LePHARE** installation. These models contain AGN-to-total luminosity ratios of $0.1 < L_{\text{AGN}}/L_{\text{tot}} < 0.9$, which cover a range of relative contributions to the combined spectrum from AGN or host galaxy emission.

3.3.3 Extinction

All the templates presented so far are intrinsic SEDs, which do not include the effects of reddening by dust. Instead, **LePHARE** applies the necessary corrections for dust extinction itself, during the stage where it computes the predicted template magnitudes for Equation 3.1 (see Section 3.2.1.1). The basic mechanism of dust extinction and its implementation in the code will be briefly described below.

Dust grains cause reddening of galaxy spectra by absorbing high energy UV and optical starlight and re-radiating it thermally at IR and submillimeter wavelengths. At a given wavelength λ this process causes a shift A_λ in the observed magnitude compared to the intrinsic (i.e. non-reddened) magnitude:

$$m_{\text{obs}}(\lambda) = m_{\text{int}}(\lambda) + A_\lambda, \quad (3.10)$$

where m_{obs} and m_{int} are the observed and intrinsic magnitudes respectively, and the shift A_λ is what is referred to as the extinction (or attenuation). The wavelength dependence of A_λ is customarily captured by the extinction curve $k(\lambda)$, which is defined as:

$$k(\lambda) = \frac{A_\lambda}{E(B-V)}. \quad (3.11)$$

Here, $E(B - V)$ is a reddening parameter which is equal to the discrepancy between observed and intrinsic $B - V$ colour:

$$E(B - V) = (B - V)_{\text{obs}} - (B - V)_{\text{int}}. \quad (3.12)$$

The exact shape of the extinction curve $k(\lambda)$ depends on the content of the interstellar medium, which differs among galaxies. A functional form has been found for a small number of local objects, notably the Milky Way (Seaton 1979; Allen 1976), Large Magellanic Cloud (LMC; Fitzpatrick 1986), Small Magellanic Cloud (SMC; Prevot et al. 1984; see Figure 3.2), and starburst galaxies (Calzetti et al. 2000; see Figure 3.2). **LePHARE** allows the user to specify an extinction law and includes the extinction curves from the above studies by default. When applying extinction, the user must also provide a range of values for $E(B - V)$. With these user-specified choices for $k(\lambda)$ and $E(B - V)$, A_λ can be inferred from Equation 3.11 and applied to the template SEDs via Equation 3.10, to compute predicted galaxy magnitudes for varying amounts of dust attenuation.

For the purpose of this thesis, it was decided to follow the recommended dust extinction procedure for each template set as specified in the **LePHARE** documentation⁶. Following these instructions, the **AVEROI_NEW** templates of star-forming galaxies are attenuated via the SMC extinction curve. For the **COSMOS** templates, the SMC law is applied to the redder star-forming galaxies, and the Calzetti law is used for the bluest starburst galaxies. The specific details of the application of dust extinction will be addressed in Section 3.4.

3.3.4 Emission lines

To ensure that predicted template magnitudes are maximally accurate, it is important to consider the effect of emission lines. Ilbert et al. (2009) argue that strong emission lines can cause notable changes in colour even in broadband filters (up to ~ 0.15 mag), and that for accurate photometric redshifts, emission lines must therefore be taken into account.

⁶More specifically, in the **README** files of the default installation.

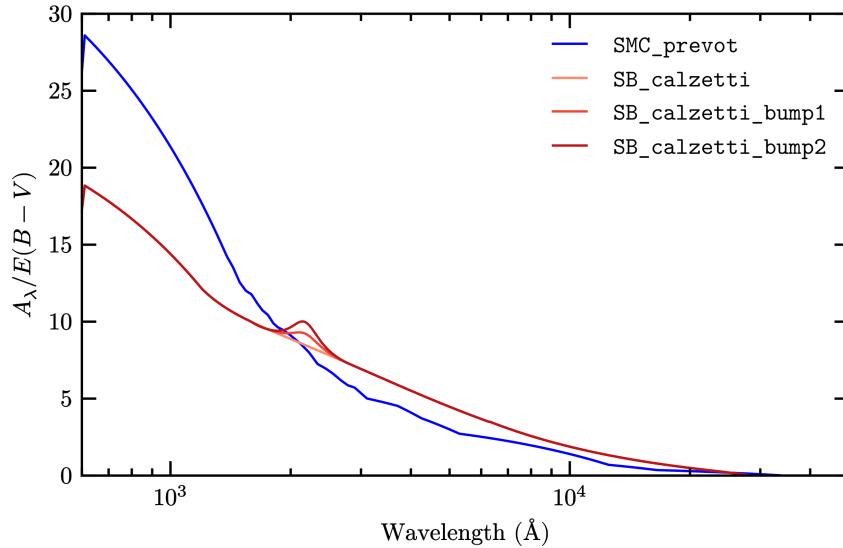


Figure 3.2 Dust extinction curves $k(\lambda) = A_\lambda/E(B - V)$ for the SMC (Prevot et al. 1984) and Calzetti et al. (2000) attenuation laws as included in the default **LePHARE** installation. There are three available options for the Calzetti et al. (2000) law, which have differing strengths of the 2175 Å bump (see e.g. Stecher 1965; Mathis 1994; Xiang et al. 2011 for more information on this feature).

There is an option within **LePHARE** to add Ly α , H α , H β , [O_{II}], and [O_{III}] emission lines to the galaxy templates. This feature has been implemented via a process described in Ilbert et al. (2009), which infers the emission line strengths from template UV fluxes. Via a relation from Kennicutt (1998), the UV luminosity is firstly used to estimate the global star formation rate (SFR):

$$\text{SFR}(\text{M}_\odot \text{yr}^{-1}) = 1.4 \times 10^{-28} L_{\nu, \text{UV}} (\text{erg s}^{-1} \text{Hz}^{-1}), \quad (3.13)$$

which can be related to the [O_{II}] emission line luminosity via:

$$\text{SFR}(\text{M}_\odot \text{yr}^{-1}) = (1.4 \pm 0.4) \times 10^{-41} L_{[\text{O}_{\text{II}}]} (\text{erg s}^{-1}). \quad (3.14)$$

Putting Equation 3.13 and 3.14 together, and converting luminosity to flux or magnitude, Ilbert et al. (2009) obtain:

$$\log \{F_{[\text{O}_{\text{II}}]}(10^{-17} \text{erg s}^{-1} \text{cm}^2)\} = -0.1 \times M_{\text{UV}} + 10.65 - \frac{DM(z)}{2.5}, \quad (3.15)$$

where $F_{[\text{O}_{\text{II}}]}$ is the $[\text{O}_{\text{II}}]$ line flux, $DM(z)$ is the distance modulus, and M_{UV} is the UV absolute magnitude measured at 2300 Å from the (dust-corrected) template SEDs. Because **LePHARE** can directly measure the latter two quantities, it is able to measure the value of $F_{[\text{O}_{\text{II}}]}$ for each (redshifted) template. It then estimates the other emission lines from their intrinsic, unextincted flux ratios with $[\text{O}_{\text{II}}]$: $[\text{O}_{\text{III}}/\text{O}_{\text{II}}] = 0.36$, $[\text{H}\beta/\text{O}_{\text{II}}] = 0.61$, $[\text{H}_\alpha/\text{O}_{\text{II}}] = 1.77$, and $[\text{H}_\alpha/\text{O}_{\text{II}}] = 2$ (McCall et al. 1985; Moustakas et al. 2006; Mouhcine et al. 2005; Kennicutt 1998). To obtain the final line strengths, these values are corrected with the appropriate attenuation. At the end, the code adds all the emission lines to the desired predicted magnitudes for the template fitting in 3.1. It only applies emission lines to templates with a dust free colour bluer than $(NUV - r)_{\text{ABS}} \leq 4$.

3.4 Photometric redshift configurations

Now the proposed photometric redshift code and template SEDs have been presented, the next step is to introduce the input details for the code. The previous few sections have mentioned several options and parameters in **LePHARE** that must be specified by the user. The values of these, and various other required inputs, can be provided in configuration files. Throughout this thesis, a particular configuration is referred to as a ‘setup’. For the DES+VIDEO dataset, a number of setups with different parameters have been tested and compared in terms of their photometric redshift performance, which is assessed by comparing the photo-zs for spectroscopically confirmed galaxies to their measured spectroscopic redshifts. The current section will introduce all these different test configurations. Afterwards, Section 3.5 will then describe the process of computing the photo-zs and various performance assessment measures, which will be analysed in detail in Section 3.6.

The photo-z tests have been conducted on the spectroscopic DES+VIDEO subsample only (see Section 2.1.3). The objective is to find the optimal configuration (judged primarily by high-redshift performance) for both the **AVEROI_NEW** and **COSMOS** SEDs, previously introduced in Section 3.3.1. The end of this chapter will describe how these setups are used to compute photometric redshifts for the full DES+VIDEO catalogue. There are several reasons for studying two sets of templates in detail. Firstly, computing photo-zs with several template sets makes the DES+VIDEO catalogue more useful as a general resource, as future

users will be able to choose which template set is best suited to their particular research. Secondly, the discussion in Section 3.3.1.2 offered some indications that the **COSMOS** templates might provide more accurate photo-zs than the **AVEROI_NEW** SEDs, which suggests that it is interesting to use the DES+VIDEO data to compare their performance in detail. Thirdly, as one of the aims of this thesis is to perform a precise investigation of how the setup choice impacts photo-zs, the use of different templates provides a better template-independent means of assessing the effect of each configuration parameter.

3.4.1 The **default** configurations

A first attempt at a suitable setup for either template set was constructed by the author, based on a configuration created by Jarvis et al. (2013) for a very similar dataset, and on findings by previous studies. These two setups are subsequently referred to as ‘**default**’, and the corresponding configuration file for the **AVEROI_NEW** templates is shown in full in Appendix A.3.

3.4.1.1 Templates

The first **default** input file uses the **AVEROI_NEW** templates. To recap, Section 3.3.1.1 mentioned that these SEDs were chosen based on the consideration that similar CWW+starburst template combinations have been successful in many photo-z studies, and that they have been shown to represent galaxies all the way out to $z \sim 6$. Following recommendations in the **LePHARE** documentation, the SMC extinction curve (see Figure 3.2) is applied to all Scd, irregular (Im in Figure 3.1a) and starburst (SB in Figure 3.1a) spectral types, with $E(B-V)$ values of 0.0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.7, 1.0. This range is larger than the recommended values listed in the **README** file of the **LePHARE** installation, which instead suggests a maximum permitted value of $E(B-V) = 0.3$. The higher **default** maximum was chosen in order to capture very red dusty galaxies, which can contaminate samples of high-redshift candidates (however, the original recommendation of $E(B-V) \leq 0.3$ will also be tested via an alternative configuration presented shortly in Section 3.4.2). Finally, the **AVEROI_NEW default** configuration adds emission lines to the templates, on the basis that Ilbert et al. (2009) claimed that this improves the photometric redshifts (see Section 3.3.4).

The second **default** setup uses the **COSMOS** SEDs. Earlier on, Section 3.3.1.2 suggested that these templates may be a suitable choice based on previous observational success. In the **COSMOS default** configuration, dust extinction is applied with values of $E(B - V) = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5$. The Calzetti extinction law (see Figure 3.2) is applied to the SB3 galaxy type and all bluer types (the reader is referred to Figure 3.1b for a clarification of the **COSMOS** spectral types), and the SMC law (see Figure 3.2) is applied to all types redder than SB3 (including SB3). No extinction is applied to galaxies redder than Sb. The Calzetti law is applied three times, once for each of the three strengths of the 2175 Å bump in Figure 3.2. All the above decisions directly follow the recommendations in the **LePHARE** installation, which are also covered in Ilbert et al. (2009). For the same reasons listed for the **AVEROI_NEW** templates, emission lines are enabled for the **COSMOS default** configuration.

3.4.1.2 Photometric redshift code options for the **default** configurations

Sections 3.2.2 and 3.2.3 introduced two options in **LePHARE** designed to aid in the redshift estimations. The **default** setups each use both these features, discussed below and implemented as described.

$N(z)$ prior

The **default** configurations both enable the option to apply a luminosity prior $N(z)$, described in Section 3.2.2. The discussion there described how the use of a prior is likely to be helpful for this thesis, as it has been shown to reduce outliers. As mentioned, the $N(z)$ prior in **LePHARE** refers to the $B - I$ colour of objects in the VVDS. While the DES observations naturally do not include B -band data, Sánchez et al. (2014) have demonstrated that DES $g - i$ colour is a reasonable substitute, so the **default** configurations use these two filters as input for the prior.

Adaptive offsets

Section 3.2.3 presented several reasons why the inclusion of adaptive offsets, which calibrate zero-point offsets via spectroscopic redshifts, might be beneficial for the DES+VIDEO catalogue. In particular, the ability to incorporate aperture

corrections, and the capacity to compensate for uncertainties in the photometric zero-point calibration are considered especially important to this thesis. With these arguments in mind, it was decided to activate this option in the **default** configurations.

Inconveniently, a quick first attempt at calculating suitable adaptive offsets for each DES+VIDEO tile revealed that the offsets determined by **LePHARE** vary considerably from tile to tile. This is not unexpected — differences in imaging quality and cosmic variance can cause such an effect in several ways. For starters, the seeing fluctuates across tiles, which leads to different compensation for aperture corrections. Secondly, due to the varying depth and cosmic variance, the galaxy populations observed in each tile are different as well. Because of the effect that template incompleteness can have on the adaptive offset values, differences in spectral types can lead to disparity in the offsets.

The reasoning above implies that the adaptive offsets would be most accurate when computed on a DES tile-to-tile basis. Unfortunately, some DES tiles contain an insufficient number of spectroscopic redshifts to make the calculated values statistically sound (e.g. there exist 19 DES tiles with <100 spectra). However, each of the eight VIDEO tiles does include a sizeable number (>1000) of spectra. It was therefore decided to compute the adaptive offsets individually for each VIDEO tile. This ensures that the offsets are based on VIDEO imaging with constant limiting magnitudes. There are still depth differences among each of the DES co-adds covering a given VIDEO tile, but this variation is generally smaller than the DES variation over the full footprint (see Figures 2.4, 2.5, and 2.6). Therefore, the proposed strategy method provides a good balance, allowing the adaptive offsets to capture most of the depth variation, while also ensuring that they are based on a statistically sound number of spectra.

Because the uncertainties in the adaptive offsets computed by **LePHARE** are not better than 0.01 mag (Ilbert et al. 2009), a value of 0.01 is added in quadrature to the apparent magnitude errors in the input photometry⁷.

⁷This is taken care of by specifying a value of 0.01 mag for the **ERR_SCALE** parameter, which leads **LePHARE** to increase the errors automatically.

3.4.1.3 Input photometry for the **default** configurations

Naturally, in order to be able to compute photometric redshifts, **LePHARE** requires the user to supply input photometry. For the **default** setups, these input files consist of fluxes in all the DES+VIDEO *grizZYJHK_s*-bands (note that this excludes the DES Y-band, which is too shallow compared to the other bands to be useful for the photo-zs in this thesis), measured in a fixed 1.95 arcsec diameter aperture. The choice for fixed-size aperture fluxes over auto photometry was mirrored on similar studies in the literature (e.g. Muzzin et al. 2009; Jarvis et al. 2013; Bowler et al. 2012, 2014). The relatively small 1.95 arcsec size was chosen because high-redshift candidates are expected to be faint, and measuring fluxes over too large an aperture compromises the signal-to-noise for faint sources (Salvato et al. 2019). The aperture diameter is therefore designed to be not much larger than the expected spatial extent of $z \gtrsim 5$ galaxies. In previous ground-based surveys with similar seeing, galaxy sizes around $z \sim 6$ have generally been dominated by the seeing (Willott et al. 2013; Bowler et al. 2014, 2015). Given that the DES and VIDEO FWHM is around ~ 1 arcsec, 1.95 arcsec apertures are expected to enclose the majority of galaxy flux, while also being small enough to minimise photometric noise. A comparable 2 arcsec diameter has also been used by previous studies with similar datasets (Jarvis et al. 2013; Bowler et al. 2012), and has been shown to deliver good results at low redshifts as well (Muzzin et al. 2009; Jarvis et al. 2013).

3.4.2 Alternative configurations

To investigate the effect of different input parameters on the photo-zs, the performance of the **default** setup for both template sets is compared to other possible configurations. Each of these rival setups is identical to the **default** version with the exception of one factor. This section presents all the alternative setups for this investigation, alongside a clarification of which parameters have been changed.

The configurations for the following four setups are template dependent, and are defined individually for each template set:

AVEROI_NEW less_ext This setup uses the original extinction recommendation found in the **README** file for the **AVEROI_NEW** templates in the **LePH-**

ARE installation. These values equal $E(B-V) = 0.0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30$. The aim of this setup is to investigate whether this lower extinction range is sufficient to produce accurate photo-zs, by testing whether it performs as well as (or better than) the wider range of up to $E(B-V) = 1.0$ used in the **default** setup.

AVEROI_NEW more_ext As emphasised throughout this chapter, a priority for the DES+VIDEO photo-zs is optimal performance at $z \gtrsim 5$. Very red dusty galaxies are a known contaminant at these redshifts (Finkelstein 2016), because their extremely red colours replicate those of true high-redshift sources in some filters. The **more_ext** setup tests whether maximally reducing contamination at high redshifts requires even higher extinction than the **default** values, in order to capture adequately the colours of any extremely red low-redshift interlopers. The permitted extinction range in the **more_ext** setup equals $E(B-V) = 0.0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.3, 1.6, 2.0$.

COSMOS ext, more_ext The **default** setup for the **COSMOS** templates uses extinction values of up to $E(B-V) = 0.5$. For the same reasons as for the **AVEROI_NEW more_ext** setup above, the **COSMOS ext** and **more_ext** configurations test whether a larger amount of reddening is required for optimal (high-redshift) results. The permitted extinction values are $E(B-V) = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 1.0$ for the **ext** setup, and $E(B-V) = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.3, 1.6, 2.0$ for the **more_ext** configuration.

*The remainder of the alternative setups are the same for both the **AVEROI_NEW** and **COSMOS** template sets (i.e. each configuration contains the same change to the respective **default** setups).*

no_emlines To evaluate whether the photometric redshift accuracy improves by adding emission lines to the templates (as claimed by Ilbert et al. 2009), emission lines are turned off in the **no_emlines** setups.

no_adapt The use of adaptive offsets is disabled in the **no_adapt** setups, in order to assess how the offsets affect the photo-z results.

no_prior Due to fact that known distributions of galaxies are heavily skewed towards low redshifts, including an $N(z)$ prior will inherently bias photo-

z estimates against high-redshift solutions (Benítez 2000). This could potentially remove some genuine high-redshift galaxies from any resulting $z \geq 5$ sample. It is therefore interesting to investigate whether it is possible to avoid using a prior, while still maintaining a manageable level of high-redshift contamination by low-redshift interlopers. Interestingly, several results in the literature indicate that the inclusion of near-infrared photometry may help to break the colour-redshift degeneracies that can lead to those contaminating catastrophic outliers (Benítez 2000; Ilbert et al. 2006). To investigate whether the DES+VIDEO catalogue does in fact contain enough photometric data to permit not using a prior, the **no_prior** configurations test the photo-z performance without the use of an $N(z)$ prior.

auto As mentioned previously in Section 3.4.1, it is common in the literature to use aperture magnitudes as input photometry, and the **default** configurations follow this custom. To test the validity of this assumption, the **auto** setups use an input catalogue based on auto fluxes instead.

des, des_auto While the default setups use data from all nine $grizZYJHK_s$ DES+VIDEO bands, the **des** and **des_auto** configurations only include flux measurements from the DES *griz* filters, using aperture and auto photometry respectively. By assessing the photo-z performance without the VIDEO data, these setups can probe the extent to which near-IR coverage improves the photometric redshifts.

3.5 Configuration testing

3.5.1 Photometric redshift computation method

The next stage in the investigation aims to quantify which is the most suitable configuration for the DES+VIDEO catalogue. To this end, photometric redshifts have been computed for the subset of galaxies with spectroscopic redshifts, using all of the 18 configurations introduced in the previous section (9 for each template set). The redshifts have been obtained via a procedure with several stages, which ensure that the resulting photo-zs are as reliable as possible. This algorithm consists of the following steps:

1. Firstly, the list of sources with spectroscopic redshifts (see Section 2.1.3) is combined with the DES+VIDEO dataset. To do this, the spectra are matched within a 1 arcsec radius to the DES+VIDEO tables from step 7 of the assembly pipeline in Section 2.2.5, i.e. the photometric catalogues for each combination of overlapping DES and VIDEO tiles. At this stage, the objective of the spectroscopic redshifts is to calibrate adaptive offsets. Therefore, in order to make sure that the input data for these offsets is maximally reliable, the algorithm removes spectroscopic objects with a bad **FLAGS_WEIGHT** value in any DES or VIDEO band. This criterion also applies to the DES Y-band, which is not used in any subsequent photo-z computations. The reason why it is nonetheless included here relates to the need for reliable input fluxes. As described in Section 2.1.1.5, the **FLAGS_WEIGHT** value in a given band is non-zero if the object overlaps a region with bad data or no data in that filter. In the case of the DES data, these bad regions include chip gaps. The images in all DES bands are severely more noisy in the vicinity of these gaps, but the DESDM catalogue uncertainties have not been suitably adjusted to account for this, which results in unreliable photometry around those regions. Because the weight maps in the five *grizY*-bands have slightly different widths for the chip gaps, using the **FLAGS_WEIGHT** parameter in all bands ensures the exclusion of as many objects as possible in the noisy border regions.
2. For each VIDEO tile, the algorithm then merges the spectroscopic catalogues from step 1 together for all the DES+VIDEO tiles that overlap that VIDEO tile. In the merging process, any duplicate spectra from overlapping DES tiles are removed, so that the combined catalogues include each spectrum only once per VIDEO tile. The duplicates are removed by rejecting any objects that do not exist within the non-duplicate version of the DESDM catalogues (assembled previously by the author as described in Section 2.2.5.1).
3. To calculate photometric redshifts, **LePHARE** is run individually on each of the VIDEO tiles, using all of the 18 setups introduced in Section 3.4 (9 for each template set). The input contains photometry and spectroscopic redshifts from the catalogues in step 2, and the output consists of the standard **LePHARE** output, which includes photometric redshifts for all sources, as

well as adaptive offsets (for the configurations where they were enabled). Running **LePHARE** in the above way automatically achieves the goal previously laid out in Section 3.4.1.2, of calibrating the adaptive offsets for each VIDEO tile individually.

4. For each configuration, the **LePHARE** results from all eight VIDEO tiles are then merged together. To achieve this, the algorithm stacks the eight output files from step 3. The objective of these merged datasets is to enable a statistical analysis of the photo-z performance for each setup (which will be described later in Section 3.5.2). Initially, the simple stacked catalogues contain duplicates from the overlapping regions between tiles. While these duplicate objects were needed in the earlier steps to make sure that the adaptive offsets are based on a maximal number of spectra, they introduce bias in the statistics. Duplicates are therefore removed; when a single DES object is covered by multiple VIDEO tiles, the algorithm retains the data from the highest ranking VIDEO tile. The author determined this ranking from Table 2.8, based on the depths and available bands. In order of descending priority, the order is: (es1-north, es1-south) for the ELAIS-S1, (xmm3, xmm2, xmm1) for the XMM-LSS and (cdfs1, cdfs2, cdfs3) for the CDF-S. The final product is a series of 18 photo-z catalogues (one for each configuration), where each spectrum is included only once.

After the above algorithm was applied to the DES+VIDEO dataset, the resulting 18 photo-z catalogues each contain a total of 35 596 sources with spectroscopic and photometric redshifts.

3.5.2 Accuracy metrics for assessing photo-z performance

The redshift catalogues from the previous section can be used to assess the photo-z performance of each configuration. To do so, this thesis uses eight accuracy metrics, largely following Sánchez et al. (2014). In this framework, the individual bias (i.e. the photometric-to-spectroscopic discrepancy) for each object i is defined as $\Delta z_i = z_{\text{phot},i} - z_{\text{spec},i}$, and N represents the total number of objects used in the estimation of the statistics. Using these notations, eight statistical measures have been defined to capture various aspects of the photo-z distribution:

- (i) **Bias.** Two metrics capture the average bias. The first of these is the mean bias $\bar{\Delta}z$, which is defined as follows:

$$\bar{\Delta}z = \frac{1}{N} \sum_i^N \Delta z_i. \quad (3.16)$$

Secondly, the median bias Δz_{50} corresponds to the median of the Δz distribution, defined as the (interpolated) middle value of an ordered list of Δz_i values. In other words, the median is the value of Δz that satisfies the following equation:

$$P_{50} \equiv P(\Delta z \leq \Delta z_{50}) = \frac{1}{2}. \quad (3.17)$$

- (ii) **Scatter.** The scatter is measured by two separate statistics, corresponding to the two bias metrics defined above. Firstly, there is the standard deviation σ of the mean bias $\bar{\Delta}z$, which is expressed as:

$$\sigma = \sqrt{\frac{1}{N} \sum_i^N (\Delta z_i - \bar{\Delta}z)^2}. \quad (3.18)$$

The second statistic measures the 68 percentile width σ_{68} of the Δz distribution about the median bias Δz_{50} ; in other words, $\Delta z_{50} \pm \sigma_{68}$ encloses 68% of the Δz distribution:

$$\sigma_{68} = \frac{1}{2}(\Delta z_{84} - \Delta z_{16}), \quad (3.19)$$

where Δz_{84} and Δz_{16} are defined analogously to Δz_{50} in Equation 3.17.

In the specific case where the values of Δz_i are normally distributed, it is the case that $\sigma = \sigma_{68}$. However, in general the bias distribution is not Gaussian, and σ_{68} captures the scatter in the core of the Δz distribution σ is more sensitive to the tails.

- (iii) **Outliers.** The number of objects with significantly bad photo-z estimates is quantified in the outlier fractions. The 2σ outlier fraction $f_{2\sigma}$, which measures the fraction of objects more than 2σ away from the mean, can

be written as:

$$f_{2\sigma} = \frac{1}{N} \sum_i^N W_{2,i}, \quad (3.20)$$

where

$$W_{2,i} = \begin{cases} 1, & \text{if } |\Delta z_i - \bar{\Delta z}| > 2\sigma \\ 0, & \text{if } |\Delta z_i - \bar{\Delta z}| \leq 2\sigma \end{cases}.$$

Similarly, the 3σ outlier fraction $f_{3\sigma}$ is expressed as follows:

$$f_{3\sigma} = \frac{1}{N} \sum_i^N W_{3,i}, \quad (3.21)$$

where

$$W_{3,i} = \begin{cases} 1, & \text{if } |\Delta z_i - \bar{\Delta z}| > 3\sigma \\ 0, & \text{if } |\Delta z_i - \bar{\Delta z}| \leq 3\sigma \end{cases}.$$

- (iv) **High-redshift performance.** The six numbered equations above, all based on Sánchez et al. (2014), capture the photo-z accuracy for the full redshift range (within a given configuration). To determine the performance at high redshifts specifically, the author has defined two additional metrics. The first quantity N_{int} captures the galaxy contamination by counting the number of low-redshift ($z_{\text{spec}} < 2.0$) interlopers in a $z_{\text{phot}} > 4.0$ sample. It is defined as follows:

$$N_{\text{int}} = \sum_i^N W_{\text{int},i}, \quad (3.22)$$

where

$$W_{\text{int},i} = \begin{cases} 1, & \text{if } z_{\text{phot}} > 4.0 \text{ and } z_{\text{spec}} < 2.0 \\ 0, & \text{otherwise} \end{cases}.$$

The second high-redshift metric N_{good} counts the number of actual high-redshift ($z_{\text{spec}} > 3.5$) objects that are retrieved at high redshift (with z_{phot} within 0.5 of z_{spec}):

$$N_{\text{good}} = \sum_i^N W_{\text{good},i}, \quad (3.23)$$

where

$$W_{\text{good},i} = \begin{cases} 1, & \text{if } z_{\text{spec}} > 3.5 \text{ and } |z_{\text{spec}} - z_{\text{phot}}| < 0.5 \\ 0, & \text{otherwise} \end{cases}.$$

The reader may have noticed that the meaning of the phrase ‘high-redshift’ for the two metrics N_{int} and N_{good} is slightly different from the general $z \geq 5$ definition used throughout this thesis. This is because the spectroscopic subsample does not contain any spectroscopic redshifts above $z_{\text{spec}} = 5.0$, and only 11 objects at $z_{\text{spec}} > 4.0$. To obtain the closest possible measurement of the $z \geq 5$ performance, it is therefore necessary to lower the definition of high redshifts for N_{good} , and to assume that this metric gives an indication of the true accuracy at $z \geq 5$. It is emphasised that placing the threshold at $z_{\text{spec}} > 3.5$ should avoid N_{good} being significantly biased by issues with colour-redshift degeneracies from confusion between the Lyman and 4000 Å breaks (since a Lyman break between 912 Å and 1216 Å enters the g-band between $z = 3.36$ and $z = 2.27$ respectively and breaks the degeneracy). Therefore, there ought to be no major issues inherent within the photometric redshift method that implies that the $z_{\text{spec}} > 3.5$ performance does not carry over to $z \geq 5$. Regarding N_{int} , the $z_{\text{phot}} = 4.0$ cutoff is likewise below $z = 5.0$, this time to ensure that N_{int} captures contamination over a slightly wider photo-z range. When this metric is used to find the best configuration for the $z \geq 5$ galaxy search, the $z_{\text{phot}} = 4.0$ threshold minimises the potential effect of low-redshift objects around $z_{\text{phot}} \sim 4.5$ scattering into a $z_{\text{phot}} \geq 5$ sample. Furthermore, the lower cutoff also ensures consistency with the value used for N_{good} .

3.5.3 Results

The statistics introduced above have been calculated for all the configurations from Section 3.4. Table 3.1 shows the results, grouped together by template set. For every one of the setups, the distribution of spectroscopic vs photometric redshift is also plotted in Figure 3.3. Together, these plots and metrics provide a comprehensive picture of the photo-z performance. Naturally, smaller values for the metrics generally imply better accuracy for all indicators except N_{good} , where a larger value is better (since it indicates that a larger number of $z_{\text{spec}} > 3.5$ objects

were correctly assigned a high redshift). When considering the performance, one must bear in mind that some metrics are interdependent. Most importantly, the 2σ and 3σ outlier fractions $f_{2\sigma}$ and $f_{3\sigma}$ depend on the value of σ . Because a higher scatter widens the bounds of $f_{2\sigma}$ and $f_{3\sigma}$, lower values for these outlier fractions only indicate a better performance if the scatter σ is the same or lower. Lastly, for the purpose of this thesis, only the absolute value of the mean and median bias is important, because it only matters how much these quantities differ from zero.

PHOTO-Z ACCURACY FOR THE **AVEROI_NEW** TEMPLATES

setup	$\bar{\Delta}z$	Δz_{50}	σ	σ_{68}	$f_{2\sigma}$	$f_{3\sigma}$	N_{int}	N_{good}
default	0.00259	0.01675	0.47682	0.09439	0.04335	0.02669	105	23
less_ext	0.02224	0.01706	0.51961	0.10407	0.04610	0.02910	120	23
more_ext	0.00220	0.01675	0.47683	0.09448	0.04340	0.02677	104	23
no_emlines	-0.00112	0.01127	0.47774	0.09221	0.04321	0.02677	109	23
no_prior	0.01475	0.01590	0.52508	0.09914	0.04332	0.02641	176	27
no_adapt	-0.00260	0.01488	0.50593	0.12308	0.04363	0.02801	119	22
auto	-0.01653	0.01007	0.46644	0.10813	0.04377	0.02641	72	14
des	-0.00438	0.01122	0.48223	0.17451	0.04976	0.03113	5	16
des_auto	-0.02364	0.00006	0.48641	0.17312	0.05001	0.03101	3	9

PHOTO-Z ACCURACY FOR THE **COSMOS** TEMPLATES

setup	$\bar{\Delta}z$	Δz_{50}	σ	σ_{68}	$f_{2\sigma}$	$f_{3\sigma}$	N_{int}	N_{good}
default	-0.01853	-0.01125	0.40009	0.07213	0.04304	0.02405	52	14
ext	-0.01990	-0.01131	0.39913	0.07264	0.04335	0.02444	51	14
more_ext	-0.02002	-0.01132	0.39943	0.07267	0.04338	0.02433	51	14
no_emlines	-0.00610	-0.01380	0.40535	0.11407	0.04267	0.02211	51	15
no_prior	-0.00530	-0.01230	0.43921	0.07499	0.04023	0.02118	106	26
no_adapt	0.01191	-0.00163	0.41498	0.11867	0.04267	0.02467	49	17
auto	-0.04709	-0.01920	0.41747	0.08238	0.04692	0.02436	40	9
des	-0.01641	-0.01694	0.53677	0.16559	0.05765	0.03616	5	9
des_auto	-0.05554	-0.02422	0.49295	0.15919	0.05588	0.03051	8	7

Table 3.1 Values of the photometric redshift accuracy metrics defined in Section 3.5.2, for all the configurations introduced in Section 3.4.

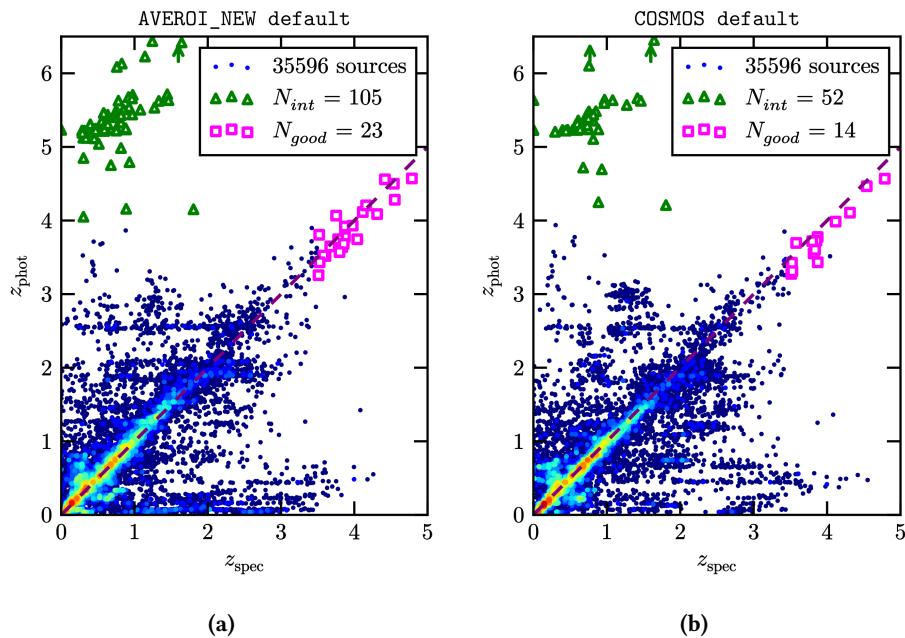
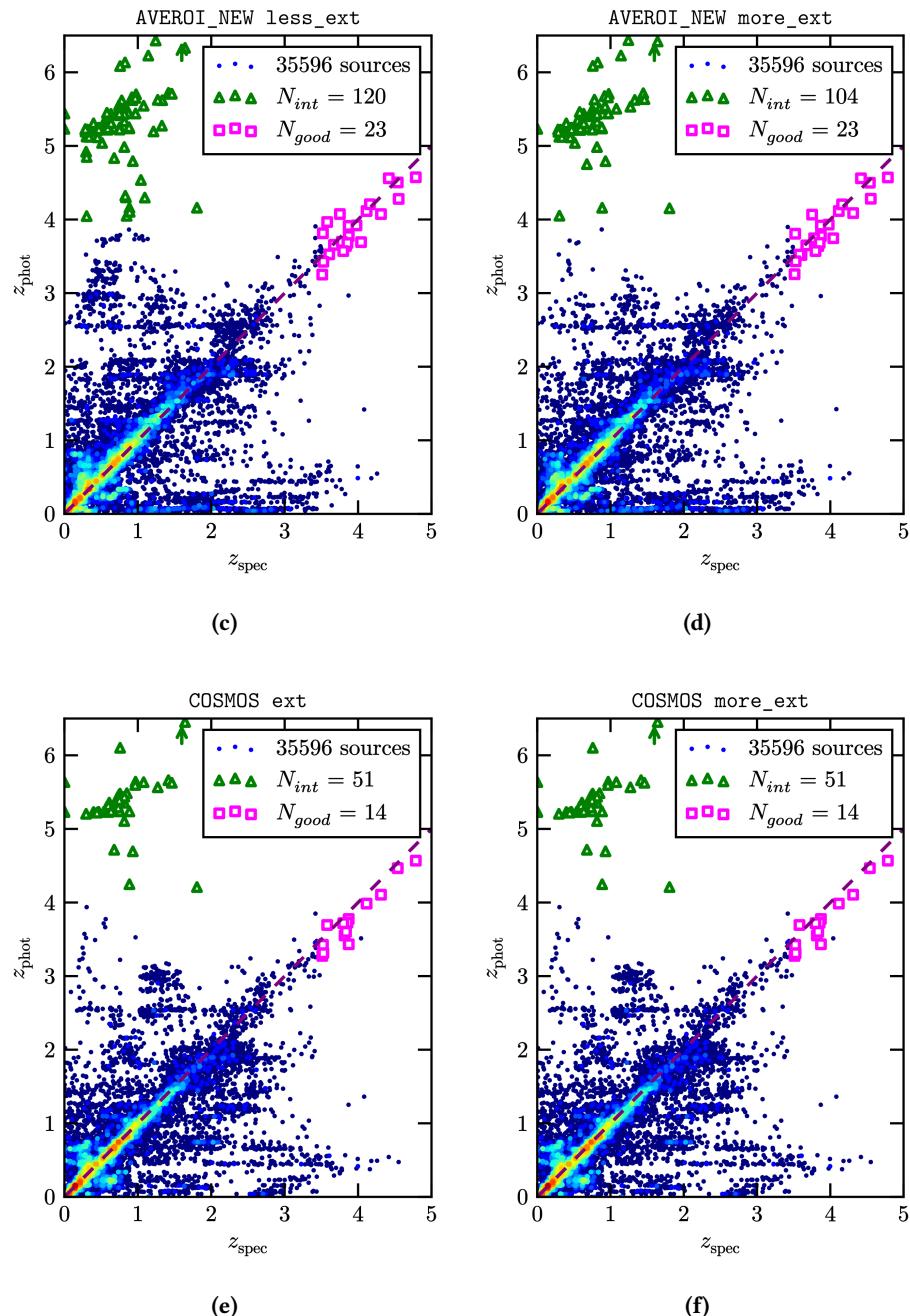


Figure 3.3 The distribution of spectroscopic vs photometric redshifts for all configurations listed in Table 3.1. The colour of each data point shows the number density in bins of 0.01, on a rainbow scale of blue (low) to red (high). The colours have been normalised over all figures, so that the same colour depicts the same number density across all plots. The pink squares represent good high-redshift fits, defined via N_{good} in Equation 3.23. The green triangles show low-redshift interlopers defined via N_{int} in Equation 3.22. The green arrows in some figures indicate interlopers for which the photometric redshift falls outside the plotted window (i.e. $z_{\text{phot}} > 6.5$).

Figure 3.3 *continued*

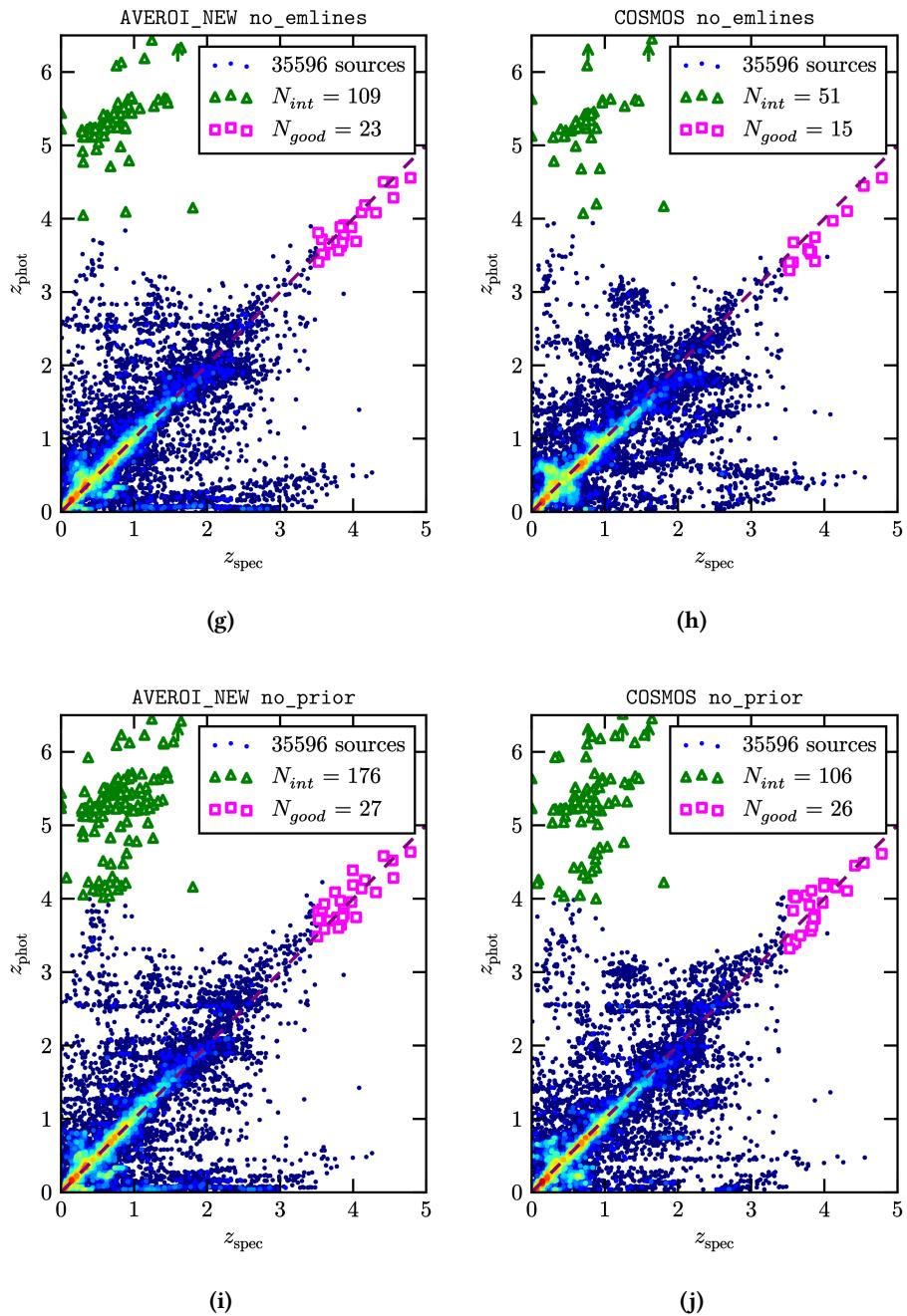
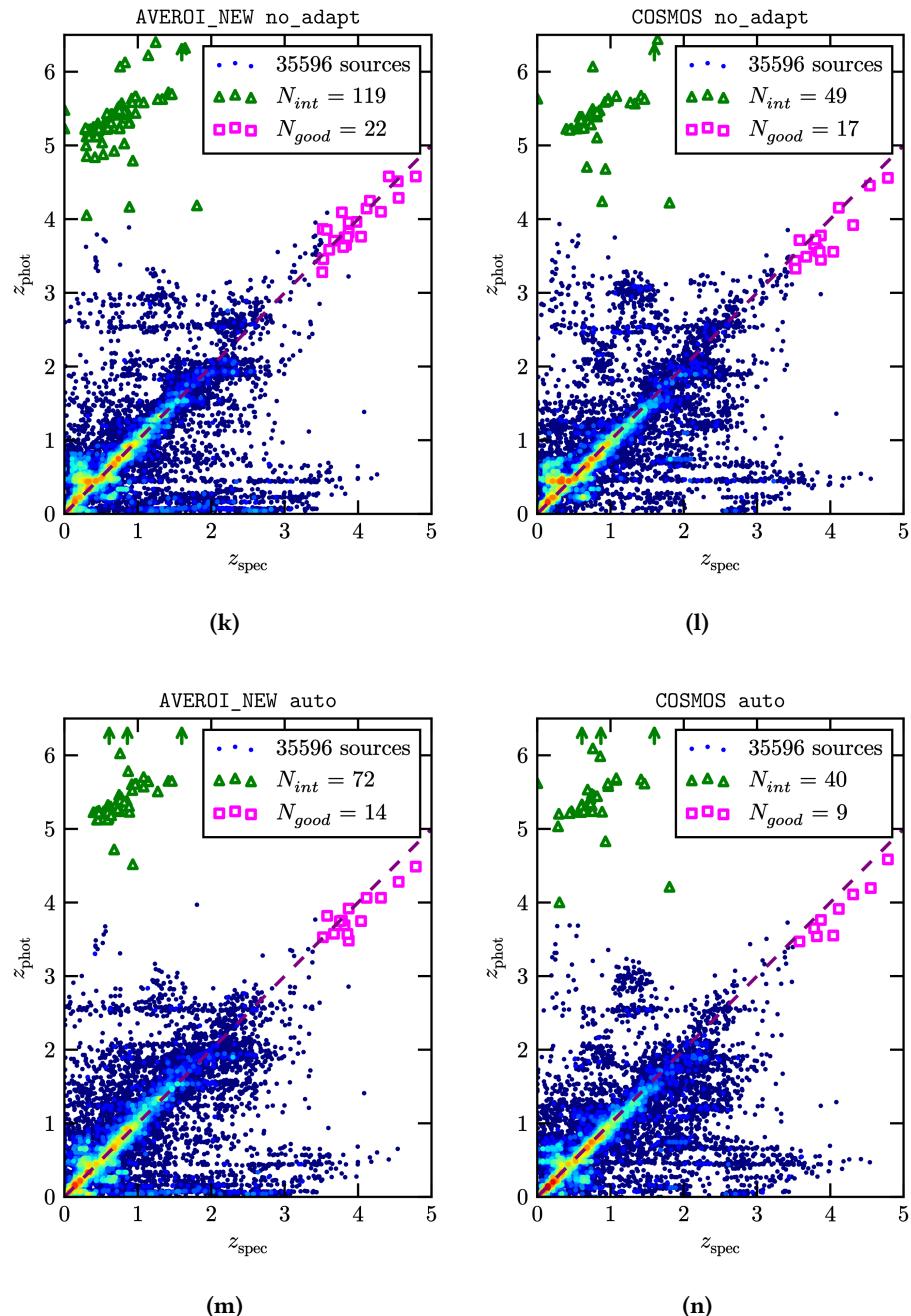
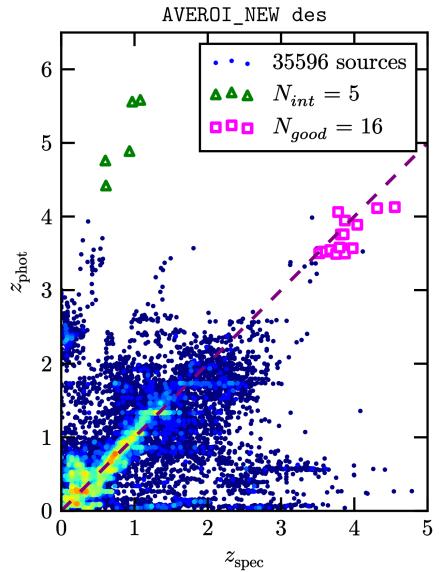
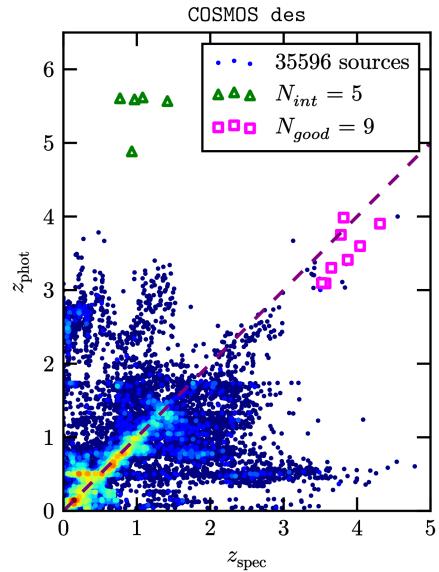


Figure 3.3 continued

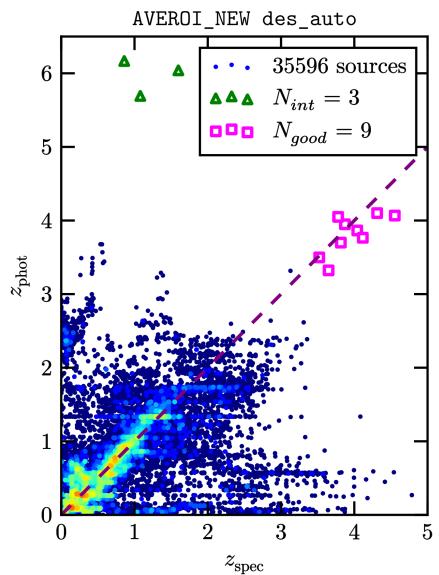
**Figure 3.3 continued**



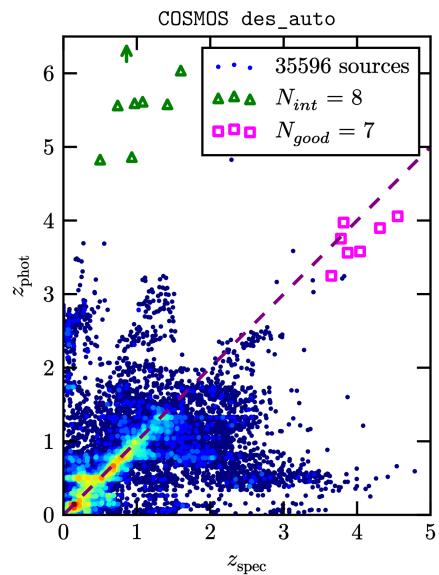
(o)



(p)



(q)



(r)

3.6 Discussion on configuration testing

The part of this chapter concerned with photometric redshift testing has now reached its final stage, where the preceding results can be used to explore the effect of all the different options that have been trialled. To this end, the current section will present a comparative evaluation of the photometric redshift performance for all the setups. In the analysis, configurations are compared by the values of their accuracy metrics in Table 3.1 and by their photo-z distributions in Figure 3.3.

The end goal is to find the most suitable way to estimate photo-zs for the DES+VIDEO catalogue, which, as emphasised throughout this thesis, needs to be specifically geared towards optimal results at high redshifts. At several points the discussion will also comment on the implication of the test results for general photo-z research. These analyses will disregard the high-redshift accuracy metrics, and looks only at those statistics that apply to the full testing sample – i.e. the bias, scatter, and outlier fractions. Those primarily reflect the accuracy for the low-redshift ($z_{\text{spec}} \lesssim 1.5$) majority of galaxies in the sample, which is also approximately the range that most other photo-z studies tend to focus on.

3.6.1 Templates

3.6.1.1 Template sets

In order to investigate which template set performs best, the **AVEROI_NEW** and **COSMOS** entries in Table 3.1 are compared directly for each setup⁸. A general observation is that the **COSMOS** templates perform better than the **AVEROI_NEW** templates for most statistical measures and most configurations. The DES+VIDEO tests thus confirm the claim by Ilbert et al. (2009) mentioned earlier in Section 3.3.1.2, that the **COSMOS** SEDs are generally more accurate than template sets based on the CWW spectra (such as the **AVEROI_NEW** SEDs).

In terms of metrics, a notable exception is that the **AVEROI_NEW** templates consistently show a larger value for N_{good} (except in the **des_auto** setup where the values are identical). Secondly, the **AVEROI_NEW** templates also perform bet-

⁸The **AVEROI_NEW less_ext** and **COSMOS ext** configurations are not directly comparable, because they use very different amounts of extinction. Therefore they are not considered in this template set comparison.

ter in terms of mean bias $\bar{\Delta}z$ in all but one case (the **no_prior** configuration being the exception). Regarding configurations, there are two instances where the **AVEROI_NEW** templates are conclusively more accurate than **COSMOS**, namely the **des** and **des_auto** setups. For the **des** setup, the mean and median bias $\bar{\Delta}z$ and Δz_{50} are 0.0120 and 0.0057 lower for the **AVEROI_NEW** templates, and the σ , $f_{2\sigma}$, and $f_{3\sigma}$ values are 10.2%, 13.7% and 13.9% lower. For the **des_auto** configuration the metrics are lower by 0.0319 and 0.0242 for $\bar{\Delta}z$ and Δz_{50} , and 13.2%, 10.5% and 16.4% for σ , $f_{2\sigma}$ and $f_{3\sigma}$. A possible explanation for why the **AVEROI_NEW** templates perform slightly better when using exclusively optical data might be found within Section 3.3.1.2. To recap, there it was stated that Ilbert et al. (2009) claim that the better accuracy of the **COSMOS** templates is because they produce a better joining of the UV to the infrared. If this is indeed the reason why these SEDs perform better for the full DES+VIDEO filterset, it would follow that this advantage would not apply when only optical data is used. At the same time, the **AVEROI_NEW** templates might be more reliable over visible wavelengths because that part of these spectra is based on the empirically measured CWW SEDs, whereas the **COSMOS** SEDs are theoretical. Furthermore, the CWW spectra are local, and the spectroscopic dataset used for the statistics is predominantly low-redshift.

Let us now return to the observation that the **COSMOS** templates generate superior statistics on the whole. Particularly interesting is the discovery that the **default COSMOS** setup produces better results than the **more_ext AVEROI_NEW** setup. Over the course of the following few sections, it will be demonstrated that those two are the best choice for their respective template sets. Upon comparing these configurations, the **COSMOS** metrics come out 0.0055 lower for Δz_{50} , and lower by 16.1% in σ , 23.7 % in σ_{68} , 0.8% in $f_{2\sigma}$, and 10.2% in $f_{3\sigma}$. The number of low-redshift interlopers N_{int} is also twice as small. The only instances where the **more_ext AVEROI_NEW** templates perform better is with regard to the mean bias $\bar{\Delta}z$ (lower by 0.016), and the number of successfully retrieved high-redshift objects N_{good} : the **COSMOS** setup correctly identifies only 14 high-redshift galaxies, whereas the **AVEROI_NEW** results find 23. In the end, the latter advantage is one of the major reasons why the high-redshift search later in this thesis makes use of the **AVEROI_NEW** templates. Section 4.2.2 will describe the reasons for this choice in detail.

The remainder of the setup comparisons will now determine the best configuration parameters for each template set. Such an analysis is useful both for finding the best way to apply each template set, and for a template-independent assessment of all the parameter options. The topic of template set choice will be revisited at the end in Section 3.6.3.3.

3.6.1.2 Extinction

The next point of inquiry involves the effects of dust extinction on the photo-zs. The observed trends with various amounts of extinction differ between the two template sets. To start with, the discussion will focus on the **AVEROI_NEW** templates. It is evident from Table 3.1 that the **default** setup with extinction values of up to $E(B-V)=1.0$ performs better than the **less_ext** setup with a maximum extinction of only $E(B-V)=0.3$ (recommended in the **LePHARE** installation). The superior performance is demonstrated across $\bar{\Delta}z$ (better by 0.0197), Δz_{50} (0.0003), σ (better by 9.0%), σ_{68} (10.3%), $f_{2\sigma}$ (6.3%), $f_{3\sigma}$ (9.0%) and N_{good} (16 fewer interlopers). N_{int} is the only statistic where both setups are equally accurate. Specific reasons for this improvement can be found in Figure 3.3a and Figure 3.3c. The latter plot demonstrates that the **less_ext** configuration contains a population of galaxies at $0 \leq z_{\text{spec}} \leq 1.5$ that are assigned high photo-zs of $z_{\text{phot}} \geq 2.5$. When inspecting the imaging and photometry for these sources, it becomes apparent that they are extremely red galaxies. It is likely that they are so red that without templates with adequate dust extinction the 4000 Å break at $z < 1.5$ is mistaken for a 1216 Å Lyman break at $z > 2.5$. However, when higher extinction values are included through the **default** setup, these galaxies are fitted well. The findings above thus suggest that the $E(B - V) \leq 0.3$ dust extinction values listed in the **LePHARE** installation do not span the full range of attenuation observed in real galaxies. In fact, this shortcoming was partly implicitly acknowledged by the Arnouts et al. (2007) study that the **AVEROI_NEW** templates are derived from, which includes $E(B - V)$ values of up to 0.6. Nonetheless, the DES+VIDEO results indicate that this is also not sufficient — out of a sample of 166 $z_{\text{phot}} \geq 2.5$ interloper galaxies that are poorly fitted by the **less_ext** extinction range, 16 are best fitted by templates as red as $E(B - V) = 1.0$. This number is of the same order as the number of true high-redshift objects $N_{\text{good}} = 24$. This thesis therefore emphasises that it is important for high-redshift studies to include templates

with large amounts of extinction, in order to accurately model very attenuated sources.

The **more_ext** configuration is aimed at testing whether extinction values above $E(B - V) = 1.0$ would further improve the photo-zs. Table 3.1 shows that this is not the case for the spectroscopic testing sample, as the difference between the **default** and **more_ext** metrics is negligible (the **default** results show only a marginal 0.0004 lower $\bar{\Delta}z$, and any improvements are at the subpercent level for the σ , σ_{68} , $f_{2\sigma}$ and $f_{3\sigma}$ metrics). In fact, only 10 out of 35 596 sources differed by more than 0.2 in photometric redshift across the two configurations. The two configurations could therefore be considered equivalent for the (low-redshift) population that dominates the spectroscopic subsample. Nevertheless, there is a potential benefit to including redder templates for the high-redshift search in the full catalogue, since extremely red galaxies are a common contaminant. As the statistical metrics indicate that there is no worsening performance when using the **more_ext** setup, this potential advantage was considered a sufficient reason to regard the extinction values in the **more_ext** as the optimal choice for the DES+VIDEO catalogue.

The remaining analysis of the most suitable extinction values will now focus on the **COSMOS** templates. When looking at the results for the **default** [$E(B - V) \leq 0.5$], **ext** [$E(B - V) \leq 1.0$] and **more_ext** [$E(B - V) \leq 2.0$] configurations, the conclusion is fairly straightforward: the photo-z metrics tend to worsen slightly with increased extinction (σ and $f_{3\sigma}$ being the only exceptions). Overall this makes the extinction in the **default** setup the best choice for the DES+VIDEO data, although it must be noted that the differences between the three configurations are quite small (< 2% for all metrics). The fact that the **default** setup performs best suggests that the extinction values of $E(B - V) \leq 0.5$ in the default setup are largely sufficient, corroborating the values from Ilbert et al. (2009) and the **LePHARE README** files. The optimal dust extinction required for the **COSMOS** SEDs is therefore less than for the **AVEROI_NEW** templates, where values of up to $E(B - V) \leq 1.0$ are needed to achieve the best performance. A likely explanation is that the **COSMOS** template set is either more complete or more accurate at the red end. In the latter case, this could again be related to the fact that the **COSMOS** templates provide a better joining of the UV to the IR than the **AVEROI_NEW** templates, which leads to more accurate red SEDs.

3.6.1.3 Emission lines

To investigate the effect of adding emission lines, the metrics for the **no_emlines** setups are compared to the **default** results. For the **COSMOS** templates, the statistics agree with the original findings by Ilbert et al. (2009), which indicated that emission lines are important to the photo-z performance. The improvement in photometric redshifts is especially clear when comparing Figures 3.3b and 3.3h. The most prominent enhancement occurs for the main body of objects below $z_{\text{spec}} \lesssim 1.2$, where the distribution is much narrower with emission lines. In terms of metrics, the superior performance of the **default** setup is reflected in Δz_{50} (lower by 0.0026); σ (lower by 1.3%); and particularly in σ_{68} (36.8%), which encapsulates the narrower distribution noted earlier. The **no_emlines** configuration does perform better for $f_{2\sigma}$ (lower by 0.9%) and $f_{3\sigma}$ (8.8%), but because the 2σ and 3σ bounds are widened by the higher scatter σ , these lower outlier fractions do not directly correspond to better photometric redshifts. The **no_emlines** results are also lower by 0.0124 in $\bar{\Delta}z$, and indicate one more good high-z galaxy through N_{good} , and one fewer low-redshift contaminant through N_{int} . However, as the latter two metrics improve only by a single object, these advantages were deemed insignificant compared to the overall worse performance. Emission lines are therefore applied in the final **COSMOS** configuration.

The conclusions are more mixed for the **AVEROI_NEW** templates. The most prominent result here is that there are 4 fewer low-redshift interlopers when emission lines are included. The **default** setup also shows slightly lower values of σ (lower by 0.2%) and $f_{3\sigma}$ (0.30%), but it performs worse in terms of $\bar{\Delta}z$ (higher by 0.0015), Δz_{50} (0.0055), σ_{68} (higher by 2.36%) and $f_{2\sigma}$ (0.32%; although the last result again does not necessarily indicate worse emission line performance due to the narrower value of σ). Altogether, the **AVEROI_NEW** results are therefore not clearly conclusive for photometric redshift research in general. However, the lower values of σ , $f_{3\sigma}$, and particularly the lower number of low-redshift contaminants, make the application of emission lines the most suitable choice at high redshifts.

3.6.2 Photo-z code options

3.6.2.1 Prior

When comparing the **default** and **no_prior** setups, it is clear that the $N(z)$ prior indeed functions as expected. Based on Table 3.1, the use of such a prior improves the scatter σ and σ_{68} by 9.2% and 4.8% respectively for the **AVEROI_NEW** templates, and by 8.9% and 3.8% for the **COSMOS** SEDs. It is noteworthy that a fall in scatter indeed agrees with previous studies by Ilbert et al. (2006) and Tanaka (2015). Regarding outlier fractions, even though the $f_{2\sigma}$ and $f_{3\sigma}$ values come out lower without a prior, those outcomes probably do not reflect a true worsening in redshifts because the lower values are likely caused by the higher scatter σ . There seems to be no clear trend in improvement with regards to the mean and median bias $\bar{\Delta}z$ and Δz_{50} . For the **AVEROI_NEW** templates, the inclusion of a prior improves $\bar{\Delta}z$ by 0.0122, but Δz_{50} gets slightly worse by 0.0009. On the other hand, for the **COSMOS** templates the prior improves Δz_{50} by 0.001, but $\bar{\Delta}z$ gets worse by 0.0132.

More insights into the effects of a prior can be obtained by direct comparisons between Figure 3.3a and Figure 3.3i (for **AVEROI_NEW**), and between Figure 3.3b and Figure 3.3j (for **COSMOS**). In both cases, the plots without a prior show an outlier population of galaxies with $z_{\text{spec}} < 1$ and $z_{\text{phot}} > 2.5$, which is strongly reduced with the inclusion of a prior. In particular, the prior decreases the number of problematic fits at $z_{\text{phot}} \sim 3$ for which the 4000 Å break at $z \sim 0$ is confused with a Lyman break at $z \sim 3$, in agreement with predictions by Benítez (2000). The reduction of outliers is also reflected numerically in the lower values of N_{int} : compared to the **no_prior** configurations, the **default** setups contains 71 fewer interlopers for the **AVEROI_NEW** templates and 54 fewer for the **COSMOS** templates. This corresponds to a reduction by a factor 1.5 and 2 respectively. The result that a prior cuts down extreme outliers mirrors similar findings by e.g. Benítez (2000), Ilbert et al. (2006), and Tanaka (2015).

The improved accuracy provided by the prior does come at a cost, as its application is found to reduce the number of good high-redshift galaxies N_{good} . The difference is relatively small for the **AVEROI_NEW** configuration (N_{good} is 4 less), but it is significant for the **COSMOS** templates, where N_{good} is reduced by 12 objects. This suggests that the inclusion of a prior excludes true high-redshift sources from a photometrically selected sample, in agreement with predictions

by Benítez (2000). Despite this disadvantage, it was nevertheless decided to include a prior in the final photo-z setup. The reasons for this decision are two-fold. Firstly, the fractional reduction in contamination by including a prior is larger than the loss of good sources, and given the rarity of true high-redshift galaxies it is important to reduce the contamination. Secondly, the spectroscopically confirmed sources in the photo-z tests are biased towards bright sources for which spectra can be measured easily. The effect of low-redshift interlopers in a real catalogue with many faint objects is expected to be much larger. Including a prior will help to limit this contamination.

3.6.2.2 Adaptive offsets

The adaptive offsets also appear to be working as expected, since the tests show that they generally improve the photo-z statistics. These improvements particularly concentrate in the scatter and outlier fractions, a result which was previously also found by Dahlen et al. (2013). For the **AVEROI_NEW** templates, the **default** setup performs better than the **no_adapt** setup in terms of $\overline{\Delta z}$ (although only marginally by 0.00001), σ (by 5.8%), σ_{68} (23.3%), $f_{2\sigma}$ (0.6%), $f_{3\sigma}$ (4.7%), N_{int} (14 fewer interlopers) and N_{good} (1 more high-z galaxy). The only exception is Δz_{50} , which improves by 0.0019 without the adaptive offsets. Although less prominent, a general improvement is also seen for the **COSMOS** templates, where the **default** configuration shows improvement compared to the **no_adapt** setup in σ (by 3.6%), σ_{68} (39.2%) and $f_{3\sigma}$ (2.5%). On the other hand, $\overline{\Delta z}$ and Δz_{50} are better without adaptive offsets, by 0.0066 and 0.0096 respectively. $f_{2\sigma}$ is also 0.9% lower, although this result is inconclusive as to the real redshift improvement because of the higher value of σ . A further advantage of the **no_adapt COSMOS** setup, especially relevant to this thesis, is that without the adaptive offsets, the number of interlopers N_{int} goes down by 3, and N_{good} goes up by 3. Given the goal of producing optimal results at high redshifts, this could *prima facie* be considered a reason not to use the offsets with the **COSMOS** templates. However, the noted improvements are fairly small, and the lower scatter values, together with a comparison between Figure 3.3l and Figure 3.3b, all indicate that the adaptive offsets do generally improve the redshifts. The plots show slight improvement around $z_{\text{phot}} \sim 3$, whilst the main enhancement is concentrated around $z_{\text{phot}} < 0.5$. Even though those lower redshifts are not directly relevant for the high-redshift performance, the positive effect of the offsets suggests that the adaptive correc-

tions track real properties, which would generally be expected to translate to high redshifts too. For all the reasons mentioned above, it was therefore decided to include the adaptive offsets in the final selected configuration for both setups.

It must be borne in mind, however, that improvements in photo-z performance when applying adaptive offsets are trivially expected if these offset corrections have been calibrated via the same set of spectroscopic objects that was used in the photo-z accuracy tests (as is the case in this thesis). When using such corrections to estimate photometric redshifts for a larger superset of survey objects, the measured advantage does therefore not necessarily carry over. Because offset values can be affected by template incompleteness (see Section 3.2.3), they may be biased towards the types of galaxies present within the spectroscopic subsample. If these sources are not representative of the galaxy population in the full survey, the offsets may not improve the photo-z performance for the total DES+VIDEO catalogue. Fortunately, the spectroscopic sample in this thesis is large and originates from a varied range of sources, making it likely that the spectra are indeed varied enough to avoid strong biases. Furthermore, the adaptive offsets also correct for errors in the zero-point calibration as well as aperture corrections, both of which are independent of the spectroscopic subsample. At least some of the improvements provided by the adaptive offsets are therefore expected to apply directly to the full survey, justifying the decision to include this feature in the final configurations.

3.6.3 Input photometry

3.6.3.1 Auto vs aperture fluxes

The analysis will now turn to the question of whether auto or aperture fluxes produce more accurate photometric redshifts. For the **COSMOS** templates, Table 3.1 shows that the aperture fluxes in the **default** setup perform better than the auto fluxes in the **auto** setup for almost all metrics. Improvements were observed in $\overline{\Delta z}$ (lower by 0.0286), Δz_{50} (0.0080), σ (lower by 4.2%), σ_{68} (12.4%), $f_{2\sigma}$ (8.3%), $f_{3\sigma}$ (1.3%) and N_{good} (5 more objects). The only exception is N_{int} , which is lower by 12 objects when adopting auto photometry.

For the **AVEROI_NEW** templates, aperture fluxes also produce better results for $\overline{\Delta z}$ (lower by 0.0139), σ_{68} (lower by 12.7 %), $f_{2\sigma}$ (1.0%; although this is insignificant due to the worse performance in σ , see below) and N_{good} (9 more objects).

However, the auto fluxes perform better in terms of Δz_{50} (by 0.0067), σ (2.2%), $f_{3\sigma}$ (1.0%) and N_{int} (33 fewer interlopers). A possible explanation for why some metrics improve under the **auto** setup may be to do with the sizes of some objects. For extended sources larger than 1.95 arcsec, 1.95 arcsec apertures cannot capture their full emission, whereas auto apertures — which are scaled individually to the size of each source — are expected to produce more accurate estimates of their total flux. Since photometric redshifts predominantly⁹ rely not on the absolute flux measurement but on the colour *difference* between filters, using fixed aperture photometry is generally not a problem for most sources (as reflected in the fact that the photo-zs still work well for the vast majority of large objects). Nevertheless, it can occasionally be an issue for extended objects where different regions have different colours, or if the DES and VIDEO astrometry do not line up perfectly¹⁰. The use of aperture fluxes could thus occasionally lead to catastrophic photo-z failures, which would particularly manifest in higher σ scatter and larger outlier fractions compared to the auto statistics, while leaving the σ_{68} scatter in the very core of the distribution mostly unaffected. This behaviour is indeed exactly what is detected in the measurements of σ , σ_{68} , and $f_{3\sigma}$ of the **AVEROI_NEW** configuration. It is not clear to the author why the partial improvement with auto photometry is only observed for that template set. It is speculated that the answer probably lies in the specifics of the SEDs, possibly combined with the effect of the adaptive offsets (which are different for each setup). Nevertheless, for the **COSMOS** setup the aperture fluxes do show smaller improvements for σ , $f_{2\sigma}$, and $f_{3\sigma}$ than for σ_{68} compared to the auto photometry, in line with the hypothesis that adopting auto fluxes has a larger impact on the catastrophic failure rate than on the main body of photo-zs.

The reasoning above can also explain why for both template sets the **auto** results show lower values for N_{int} . The author visually inspected¹¹ the imaging and best-fit SEDs of low-redshift interlopers in the **default** setups, and found that the majority of these objects are very red extended galaxies. As explained,

⁹For the sake of conceptual clarity, the effect of the $N(z)$ prior is set aside in this line of reasoning.

¹⁰Even though the astrometry of the two surveys was found to line up very well (see Section 2.3.1), for extended objects that are bright outside the aperture, even a small offset could make a difference.

¹¹The general process for the visual inspection of photo-z fits will be described in more detail in Section 4.2.4.

auto magnitudes capture more of the total flux for such large sources, altogether making it less likely that the code assigns them erroneous (in this case high) redshifts. However, one must note that with values of 72 (**AVEROI_NEW**) and 40 (**COSMOS**), N_{int} is not close to zero even when applying auto photometry, so using these fluxes still cannot properly eliminate low-redshift contamination from extended objects. This is probably partly caused by template incompleteness at the very red end, and partly by the fact that the forced photometry in this thesis bases the position of the auto apertures on the shape of an object in the $r + i + z$ detection image. For galaxies that are more extended in VIDEO than in DES, such apertures would still be too small to capture the full flux.

Despite creating less contamination at high redshifts (and at low redshifts for the **AVEROI_NEW** setup), auto magnitudes perform considerably worse in terms of the number of correctly measured high-redshift objects N_{good} . This is almost certainly caused by the underestimation of **SExtractor** auto fluxes for faint objects, an effect which has been well documented in the literature (Labbé et al. 2003; Quadri et al. 2007). Very faint measurements tend to miss an appreciable fraction of light, because the scaled auto apertures often become too small. Since high-redshift galaxies are on the whole faint as a population, this issue disproportionately affects their photo-zs in particular. The demonstrated significant reduction of N_{good} for the **auto** setups, which is fractionally larger than the drop in N_{int} for both template sets, is the most important reason why this thesis concludes that aperture photometry is more suitable for high-redshift photo-z research. Moreover, for the full DES+VIDEO catalogue the benefit of aperture fluxes is anticipated to be even larger than expected purely from N_{good} , because the total catalogue contains many sources that are several magnitudes fainter than the spectroscopically confirmed galaxies used in the photo-z accuracy tests. Based on all these reasons, it was decided to use aperture magnitudes in the final configurations for both template sets.

3.6.3.2 Near-IR data

As described when outlining the aims of this thesis in Section 1.2.4, a one of its core objectives is to investigate the benefits of adding near-infrared information to optical surveys. Table 3.1 indicates that for the DES+VIDEO dataset, the addition of VIDEO near-IR photometry indeed strongly improves the redshift estimates. This result applies to both template sets. A comparison between

the **default** (DES+VIDEO) and **des** (DES-only) metrics shows that for the **AVEROI_NEW** templates, the DES+VIDEO photo-zs are better than the DES-only results by 0.0018 for $\bar{\Delta}z$, 1.1% for σ , 45.9% for σ_{68} , 12.9 % for $f_{2\sigma}$, 14.3% for $f_{3\sigma}$, and by 7 more objects for N_{good} . The trend is similar but stronger for the **COSMOS** templates, which show superior results for DES+VIDEO over DES by 0.0057 for Δz_{50} , 25.5% for σ , 56.4% for σ_{68} , 25.3% for $f_{2\sigma}$, 33.5% for $f_{3\sigma}$, and 6 more objects for N_{good} . The only metrics where the DES+VIDEO statistics perform worse are Δz_{50} (by 0.0056) for the **AVEROI_NEW** SEDs, $\bar{\Delta}z$ (by 0.0021) for the **COSMOS** templates, and N_{int} for both setups. The latter metric goes up enormously when including the VIDEO data — from 5 to 105 interlopers for the **AVEROI_NEW** templates, and from 5 to 52 for **COSMOS**.

Near-infrared data in high-redshift research

This sharp increase in low-redshift interlopers is slightly surprising; because the majority of metrics demonstrate that near-IR data strongly improves the photo-zs, one may expect a priori that this general result also carries over to $z_{\text{phot}} > 4$. A low contamination rate is crucial to the search for high-redshift galaxies, so it is important to understand the reasons behind this rise in N_{int} . Investigating the SEDs of individual low-redshift contaminants revealed that it is very likely caused by the combined effect of template incompleteness and the prior, as will now be explained. As described earlier in Section 3.6.3.1, most low-redshift interlopers in the **default** configuration are IR-luminous galaxies whose extremely red spectra are not fitted well by any low-redshift templates. Subsequently, their red colours are best captured by a high-redshift solution instead. However, because the DES-only data excludes the near-IR colours that are not well matched to the available low-redshift templates, low-redshift solutions are more compatible with the DES input photometry without any VIDEO data. Provided these low-redshift fits are sufficiently probable, the prior will then cause **LePHARE** to favour them over high-redshift options. As a result, the **des** configurations show a smaller number of low-redshift interlopers. To frame the argument more generally: without near-IR data there are fewer fluxes to constrain a (true or fictitious) high-redshift solution, and the code is then more likely to select a low photometric redshift because of the prior. Following this reasoning, the **des** output is expected to contain fewer high-redshift photo-zs across the board, leading to a

reduction in the number of good high-redshift solutions as well. This effect is indeed observed in the lower values of N_{good} .

The lower values of N_{int} for the **des** setups are thus most likely caused by the combined effect of template incompleteness and the prior, rather than because exclusively optical fluxes truly capture the redshifts better. On top of that, it is reiterated that the other metrics in Table 3.1 do indicate that including near-IR data improves the photo-zs. In particular, the DES+VIDEO photometry performs better in terms of N_{good} , which is one of the reasons why it was decided to include near-IR data in the final configuration for this thesis. Moreover, the positive effect on the number of correctly retrieved high-redshift galaxies is expected to increase further for the full DES+VIDEO catalogue, which contains many sources fainter than the spectroscopic testing sample. These objects have less reliable photometry, causing that their best-fit solutions to be less strongly constrained. For some faint true high-redshift galaxies, if only DES photometry is included, the prior could then lead the photo-z code to favour low-redshift solutions. Finally, the last argument in favour of using near-IR data for the high-redshift search involves reducing contamination by sources other than low-redshift galaxies. In contrast to the spectroscopic test sample, which is comprised only of galaxies, the full dataset contains many other types of objects that can contaminate a high-redshift sample. In Chapter 4, it will be shown that contamination by transients is a major issue in the $z \sim 6$ sample, and that near-IR data can be used to eliminate these problematic detections. The VIDEO data is also of major importance to reduce contamination by stars later; Section 3.7.1 will demonstrate the value of the near-IR data for star-galaxy separation.

Near-infrared data in general photo-z research

In addition to the search for distant galaxies, it is also interesting to focus on the effect of near-IR data on the lower-redshift bulk of test objects, as the DES+VIDEO dataset can offer valuable insights for photo-z science in general. As mentioned in the beginning of the current discussion on near-IR data, Table 3.1 demonstrates that almost all general (i.e. non-high-redshift) metrics are significantly more accurate with the inclusion of VIDEO data. This is in agreement with results in the literature – both simulations (e.g. Banerji et al. 2008; Abdalla et al. 2008) and observational studies (Rudnick et al. 2001; Jarvis et al. 2013; Banerji et al. 2015; Salvato et al. 2019).

Plots can provide a more detailed understanding of the effects of adding near-IR data. First of all, some insight can be gained by considering the z_{spec} vs z_{phot} plots in Figure 3.3, where the effect of the VIDEO data can be observed by comparing the **default** distributions (Figure 3.3a for **AVEROI_NEW**, Figure 3.3b for **COSMOS**) to the **des** equivalents (Figure 3.3o, **AVEROI_NEW**, Figure 3.3p). To analyse the effect of the VIDEO data further, the author created two more figures to show the statistical metrics for these results as a function of redshift. For the range $0 < z < 6.5$ this plot is shown in Figure 3.4, and a low redshift close-up for $0 < z < 1.5$ is provided in Figure 3.5.

Together, all the plots show more precisely how the near-IR photometry improves the redshifts estimates. The discussion will concentrate on the most notable features. The first of these involves the scatter in the $\Delta z = z_{\text{phot}} - z_{\text{spec}}$ distribution in Figure 3.3 (3.3a vs 3.3o and 3.3b vs 3.3p). For both template sets, it is clear that the DES+VIDEO distributions are narrower at the core than the DES-only ones. This is the behaviour that causes the observed strong decrease in the σ_{68} metric (45.9% for **AVEROI_NEW** and 56.5% for **COSMOS**). For the **COSMOS** plots, the addition of near-IR data leads to considerably less scatter in the tails of the distribution as well. This is reflected in the value of σ , which is lower for DES+VIDEO by 25.6% compared to DES-only (at 1.1% the equivalent improvement is less prominent for the **AVEROI_NEW** templates). In general, the larger scatter with DES-only photometry is most notable between $0.3 \lesssim z_{\text{phot}} \lesssim 2$. Looking at the scatter as a function of z_{phot} in Figure 3.4, it becomes apparent that over most of the redshift range the values of σ and σ_{68} are indeed smaller for the DES+VIDEO configurations than for DES-only. The most prominent exception is the region around $2.7 \lesssim z_{\text{phot}} \lesssim 3.1$, where both scatter metrics come out lower without VIDEO data. However, a quick look at the corresponding distributions in Figure 3.3 demonstrates this reduced scatter does not correspond to better photo-z performance. Instead it is simply a result of a consistently stronger bias; above $z_{\text{phot}} \gtrsim 2.7$ almost all objects in the DES-only dataset lie at $z_{\text{spec}} \sim 0$, leading to low scatter centred around a false result. The second exception is the region around $z_{\text{phot}} \sim 0$ for the **AVEROI_NEW** templates. In this region, most notable in Figure 3.5, the DES+VIDEO configuration shows slightly higher scatter metrics. Figure 3.3a demonstrates that this effect originates from the large number of objects at any z_{spec} that are assigned $z_{\text{phot}} \sim 0$, which most likely occurred as a result of template incompleteness. This problem is localised to $z_{\text{phot}} \sim 0$. Apart from

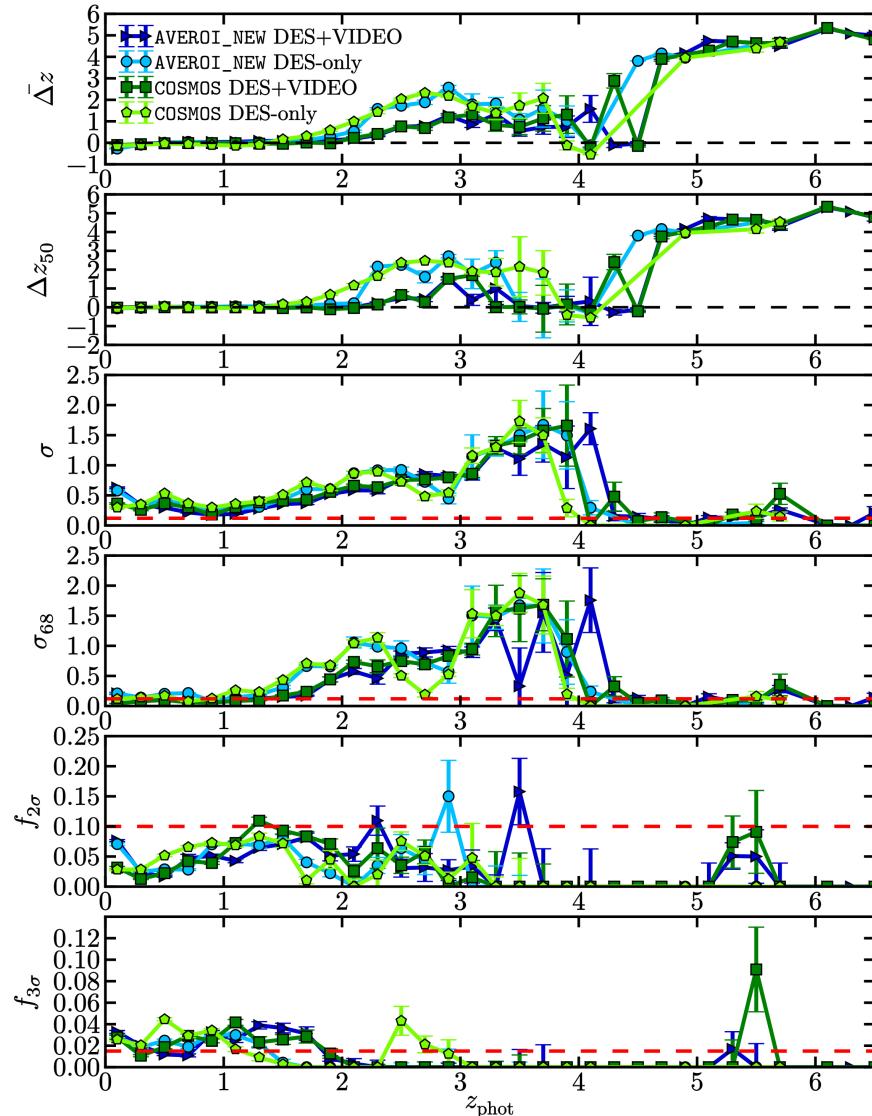


Figure 3.4 The statistical metrics defined in Equations 3.16–3.21 as a function of photometric redshift, calculated in redshift bins of $\delta z = 0.2$. Plots have been created via a custom script, adapted by the author based on code written by Matias Carrasco Kind for Sánchez et al. (2014). Results for the **AVEROI_NEW** templates are plotted as dark blue triangles for the **default** (DES+VIDEO) configuration, and light blue circles for the **des** configuration. Results for the **COSMOS** templates are shown as dark green squares for the **default** configuration, and light green pentagons for the **des** configuration. The red dashed lines in the σ , σ_{68} , $f_{2\sigma}$ and $f_{3\sigma}$ plots indicate the DES collaboration requirements on these metrics (Sánchez et al. 2014). When interpreting all six plots, it is important to consider the interplay between the different metrics. Around $z_{\text{phot}} \approx 3$ for instance, the scatter metrics σ and σ_{68} for the DES-only configurations fall below those of the DES+VIDEO setups. However, a closer look at the z_{spec} vs z_{phot} plots (Figures 3.3a, 3.3o, 3.3b and 3.3p) shows that this is due to the majority of galaxies in those redshift bins having very low $z_{\text{spec}} < 0.5$. This is reflected in the bias metrics, where the DES-only setups show much higher values for $\bar{\Delta}z$ and Δz_{50} . Similarly, there is a connection between the value of the outlier fractions $f_{2\sigma}$ and $f_{3\sigma}$ and the value of σ , where a higher value of σ leads to fewer objects being classed as $>2\sigma$ and $>3\sigma$ outliers.

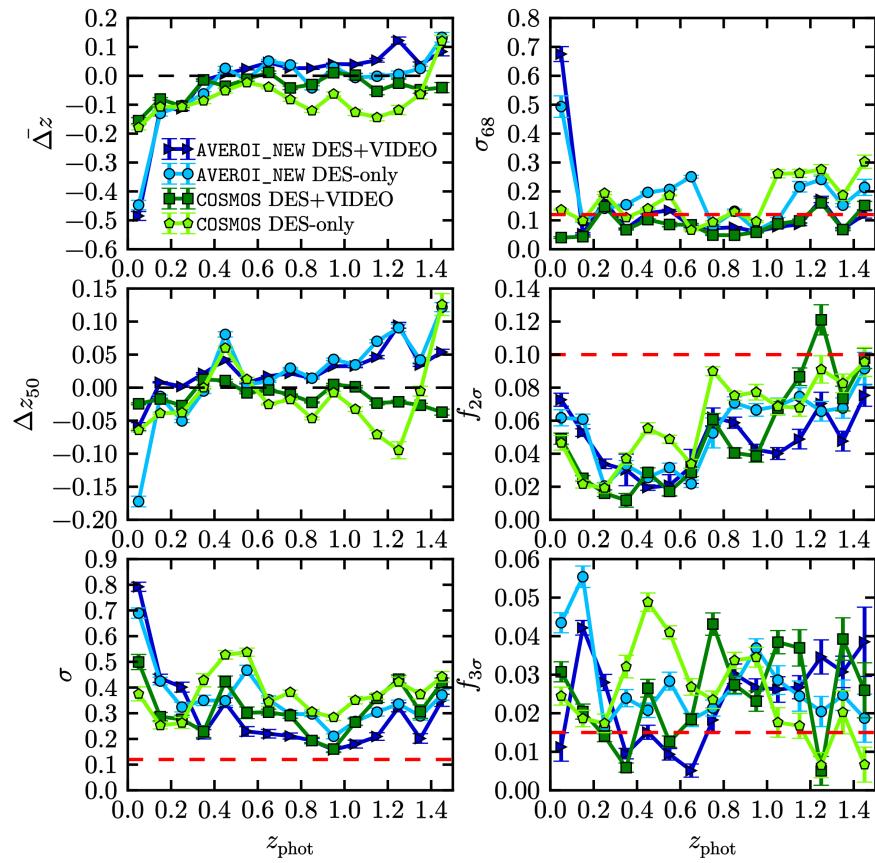


Figure 3.5 Results from Figure 3.4 zoomed in on the low-redshift region $0 < z_{\text{phot}} < 1.5$, this time using redshift bins of $\delta z = 0.1$. In all other aspects, the description is identical to that of Figure 3.4.

the two exceptions stated above, the inclusion of near-IR data generally leads to lower scatter at almost all redshifts. This is an important result for photometric redshift studies, as low scatter is often desirable for cosmology (Sánchez et al. 2014; Banerji et al. 2008; Crocce et al. 2011; Nakajima et al. 2012).

The second main improvement that comes with the addition of near-infrared data concerns the reduction of outliers. Photometric redshifts heavily rely on strong spectral features, such as the Lyman and 4000 Å breaks. With only *griz* DES data, the photo-zs enter the so-called ‘redshift desert’ (Rudnick et al. 2001; Ilbert et al. 2006) at $z \gtrsim 1.5$, as the 4000 Å break shifts out of the *z*-band. Because the Lyman break only enters the *g*-band between¹² $z = 2.3$ and $z = 3.4$, without near-IR data there are no strong continuum features available to constrain the redshift of objects between $1.5 \lesssim z \lesssim 3$. As a result, the DES-only results in Figures 3.3o and 3.3p show a large number of outliers between $2 \lesssim z_{\text{phot}} \lesssim 3$, mainly with $z_{\text{spec}} \lesssim 0.5$. Many of these are galaxies without a strong 4000 Å break, for which the code erroneously assumes that the absence of such a feature must mean the Lyman break lies unobserved at $2 \lesssim z_{\text{phot}} \lesssim 3$. The addition of near-infrared data removes this degeneracy, as the absence of a 4000 Å break can now be confirmed in the VIDEO *YJH*-bands. The effectiveness of near-infrared data at overcoming problems to do with the redshift desert is also illustrated by Figure 3.4. From $z \gtrsim 1.5$ onwards, the bias metrics Δz and Δz_{50} start to deviate more strongly from zero for the DES-only plots (compared to their DES+VIDEO counterparts). Similarly, the scatter as indicated by σ and σ_{68} also increases sharply. The importance of the VIDEO data is observed the most in the value of Δz_{50} , which stays close to zero for most of the $1.5 < z < 3$ redshift range when near-IR data is included.

Figure 3.5 focuses on the value of near-infrared data for the local ($z < 1.5$) end of the spectroscopic subsample. This redshift range contains the bulk of the available spectra (see Figure 2.10), and is of particular interest for cosmological applications of photo-zs. The plots demonstrate that in most redshift bins, the values for the DES+VIDEO metrics are lower than for DES-only. It is also interesting to consider the red dashed lines for σ , σ_{68} , $f_{2\sigma}$, and $f_{3\sigma}$, which show the DES collaboration requirements on these metrics (Sánchez et al. 2014). The im-

¹²Depending on the amount of absorption in the Lyman- α forest, the rest-frame Lyman break lies between 912 Å (the Lyman limit) and 1216 Å (Lyman- α), see e.g. Section 1.3.1.2 and Madau (1995).

portance of near-infrared data is especially apparent for σ_{68} , where the values in most bins only lie below the specified requirement when VIDEO data is added.

The discussion on near-IR data will now conclude with some final remarks. As mentioned at the beginning of this chapter, the reader must bear in mind that the configurations in this thesis have been designed primarily with high-redshift performance in mind. Regarding the conclusions for general photo-z research, this affects several aspects. Most importantly, the choice of code was informed by the fact that template fitting codes are more suitable in this regime. Results are therefore presented with the caveat that they are not explicitly proven to hold for empirical methods. Naturally, a related limitation is that the research in this chapter has only been performed using **LePHARE**. Extending this work to other codes is therefore presented as a topic for further study (although at low redshifts several comparative studies already exist, see e.g. Hildebrandt et al. 2010; Abdalla et al. 2011; Sánchez et al. 2014). Furthermore, the setups used in the near-IR analysis all include emission lines, even though Section 3.6.1.3 indicated that emission lines do not conclusively improve the non-high-redshift results for the **AVEROI_NEW** templates. However, the small worsening in some metrics between the **default** and **no_emlines** setups is considerably less than the improvement with VIDEO data, so it is expected that near-IR data will still be beneficial if emission lines are omitted. Lastly, the third possible high-redshift bias lies in the use of fixed aperture magnitudes, which Section 3.6.3.1 found to perform slightly worse for the **AVEROI_NEW** templates in some metrics. However, a comparison between the **auto** and **des_auto** metrics in Table 3.1 confirms that using near-IR data leads to considerable improvements with auto photometry as well.

3.6.3.3 Template sets revisited

Let us now return to the question of template set choice raised in Section 3.6.1.1. The preceding analyses of the **AVEROI_NEW** setups have concluded that for those SEDs the **more_ext** configuration is the most suitable for this thesis (i.e. when judged primarily by the expected high-redshift performance). For the **COSMOS** templates, the **default** configuration emerged as the best-performing set of options, at high redshifts as well as over the full (low-z dominated) redshift range.

For the sake of versatility, it was decided to run both these best-performing setups on the full DES+VIDEO catalogue. Anyone who wishes to use the catalogue can then choose which template set is most appropriate for their particular

science goals. For instance, the high-redshift search later in this thesis makes use of the **AVEROI_NEW** templates, for reasons that will be explained in Section 4.2.2. Lower redshift applications, on the other hand, may benefit from using the **COSMOS** photo-zs instead, as Section 3.6.1.1 demonstrated that these templates performed better in almost all non-high-redshift metrics.

3.7 Full DES+VIDEO catalogue

After it was decided to use the **AVEROI_NEW more_ext** and **COSMOS default** setups as the final configurations, photometric redshifts were generated for all 2 443 576 sources in the full DES+VIDEO catalogue. As before, these redshifts have been calculated by running **LePHARE** per VIDEO tile, calibrating the adaptive offsets in the same way as for the spectroscopic subsample in Section 3.5.1. The results were then added to the DES+VIDEO catalogue from Chapter 2.

Before any meaningful scientific inferences can be made from the photo-zs, it is necessary to eliminate sources that are stars, as these of course have nonsensical redshift estimates. The current section will therefore firstly address finding a suitable star-galaxy separation strategy. Along the way, the analysis will also discuss the extent to which the VIDEO data helps to distinguish between galaxies and stellar sources. The final part of this section will then present the photometric redshifts for all galaxies in the DES+VIDEO dataset, obtained after filtering out stars via the established method.

3.7.1 Star-galaxy separation

3.7.1.1 Background

Finding a reliable way of filtering out stars is crucial for ensuring that the DES+VIDEO catalogue is suitable for future studies of galaxy evolution and cosmology. Such a star-galaxy separation strategy is also necessary for the $z \gtrsim 5$ galaxy search later in this thesis, as stars are known contaminants among high-redshift samples. Furthermore, the study of how to differentiate between galaxies and stellar sources is of general interest to the scientific community, so it is important to investigate what insights the DES+VIDEO catalogue can offer.

Common star-galaxy separation methods in the literature often employ morphological strategies. In recent years, use of the **spread_model** classifier has

become popular (Desai et al. 2012; Soumagnac et al. 2015; Banerji et al. 2015). This parameter is computed by **SExtractor** as a discriminant between the best-fitting PSF model and an extended disk model. Other morphology-based options in the literature have included the use of neural networks, such as the **CLASS_STAR** parameter in **SExtractor** (Toft et al. 2003; De Jong et al. 2015; Sevilla-Noarbe et al. 2018) or custom machine learning methods (Soumagnac et al. 2015; Sevilla-Noarbe et al. 2018). However, at $z \sim 6$, galaxies are frequently small in size (Ono et al. 2012; Curtis-Lake et al. 2016), which means that they are not reliably resolved in ground-based surveys with similar seeing to DES+VIDEO. While some objects in ground-based high-redshift samples show resolved light-profiles (Bowler et al. 2012), overall roughly half the candidates are unresolved (Willott et al. 2013; Bowler et al. 2014), and the apparent sizes are generally dominated by the seeing (Bowler et al. 2015). Subsequently, morphological criteria would likely not be effective in distinguishing between stars and galaxies at high redshifts.

Alternative methods based on colour information are therefore more appropriate for the purpose of this thesis. Some authors have performed star-galaxy separation via simple colour-colour cuts based on optical and near-IR photometry (e.g. Baldry et al. 2010; Jarvis et al. 2013). Others have employed an SED-based approach using the full range of photometry from all available filters, by fitting star SEDs and rejecting objects from a high- z sample if they show good stellar fits (e.g. Bowler et al. 2014, 2015).

3.7.1.2 A χ^2_v -based method

This thesis has opted to investigate an SED-based approach. **LePHARE**'s inbuilt capacity to fit stellar SEDs makes such a strategy straightforward to implement, since the code assigns every object a best-fit galaxy, QSO, and stellar template with corresponding χ^2 in the output files. This χ^2 can be converted¹³ straightforwardly to a reduced χ^2_v via Equation 3.5. Star-galaxy separation can then be performed by requiring that any object designated as a galaxy must have a lower galaxy reduced $\chi^2_{v,\text{gal}}$ compared to its star fit $\chi^2_{v,\text{star}}$:

¹³This is necessary in order to enable meaningful comparisons between objects with varying numbers of observed filters.

$$\chi_{v,\text{gal}}^2 - \chi_{v,\text{star}}^2 < \Delta\chi_v^2. \quad (3.24)$$

Here, $\Delta\chi_v^2$ is a threshold that can be adjusted according to the desired strength of the separation cut. A value of $\Delta\chi_v^2 = 0$ dictates that any object designated as a galaxy must simply show a better galaxy fit than a star fit. Negative values of $\Delta\chi_v^2$ produce a stricter selection, with more negative values generating increasingly conservative cuts. The star-galaxy separation method in the rest of the current chapter employs a threshold of $\Delta\chi_v^2 = 0$. For the selection of a high-redshift sample in Chapter 4, a lower value of $\Delta\chi_v^2 = -1.5$ is imposed.

3.7.1.3 Method validation

In order to validate the above method, it has been trialled on the DES+VIDEO **LePHARE** output, using the fits from both the **AVEROI_NEW** and **COSMOS** SEDs. The results were then compared to a truth catalogue based on spectroscopic and space-based observations, provided by Nacho Sevilla-Noarbe on behalf of the DES collaboration¹⁴. A similar truth catalogue is described in Sevilla-Noarbe et al. (2018). The dataset contains a total of 226 611 sources with classifications. Of these, 73 551 matched to the (non-duplicate) DES+VIDEO catalogue within a 1 arcsec radius. Sources flagged as galaxies (**TRUE_CLASS** = 1) account for 64 326 of this total, and stars (**TRUE_CLASS** = 0) for 982. The 9225 remaining objects are either classified as QSOs or uncertain. Of the 982 objects listed as stars, 132 also occur in the DES galaxy spectroscopic redshift catalogue (introduced in Section 2.1.3); these sources were removed from the list of stars in order to achieve maximum confidence that the truth sample indeed consists solely of stars. Finally, to reduce the chance of any biases through faulty photometry, any sources that contain non-zero values for the **FLAGS_WEIGHT** parameters were also removed (the process and justification for this is the same as in Section 3.5). The final truth sample then consists of 847 stars and 62 951 galaxies, leading to a total truth catalogue containing 63 798 sources.

Using the **AVEROI_NEW** SED fits, the star-galaxy separation method proposed in Section 3.7.1.2 (applying Equation 3.24 with a loose cut of $\Delta\chi_v^2 = 0$ to fits based on DES+VIDEO photometry) classifies 2895 of the total 63 798 sources as stars

¹⁴For the sake of producing reproducible research, it should be mentioned that this catalogue was titled **round5_101_testing_master.fits**.

and 60 903 as galaxies. Out of the 847 real stellar sources, 834 (98%) are indeed retrieved as stars. Only 15 (1.8%) actual stars are falsely designated as galaxies. This implies that the *stellar contamination* — defined as the ratio between the number of stars misidentified as galaxies to the total number of sources classified as galaxies (Soumagnac et al. 2015) — equals¹⁵ 0.025%. Moreover, 60 888 true galaxies (97%) are confirmed as galaxies. This percentage — i.e. the ratio of the number of true well-classified galaxies to the number of total true galaxies — is sometimes referred to as the *galaxy completeness* (Soumagnac et al. 2015). Finally, 2063 (3.3%) real galaxies are wrongly classified as stars. When repeating the above analysis with the **COSMOS** fits, the proposed separation strategy finds 2275 stars and 61 523 galaxies. It correctly identifies a total of 828 (98%) true stars and misjudges 19 (2.2%) as galaxies, implying a stellar contamination rate of 0.031%. In the same vein, 61 504 (98% galaxy completeness) true galaxies are correctly classified and 1447 (2.3%) are wrongly classified as stars. Figures 3.6a and 3.6c further illustrate the separation results, showing the full $\chi_{v,\text{gal}}^2 - \chi_{v,\text{star}}^2$ distributions of the true stars and galaxies for the two template sets. When comparing all the results for the two configurations, it is apparent that the **COSMOS** SEDs perform slightly better in terms of galaxy completeness. On the flip side, however, they also show higher stellar contamination. A likely explanation for these observations is the lower galaxy χ_v^2 found among the **COSMOS** fits, leading to lower values of $\chi_{v,\text{gal}}^2 - \chi_{v,\text{star}}^2$ (because the two configurations use the same set of stellar templates, $\chi_{v,\text{star}}^2$ remains almost¹⁶ entirely the same). Therefore, more objects satisfy the $\chi_{v,\text{gal}}^2 - \chi_{v,\text{star}}^2 < 0$ criterion for galaxy classification. As a result, more true galaxies are assigned galaxy status with the **COSMOS** SEDs, leading to higher galaxy completeness, but more true stars are classified as galaxies as well, increasing stellar contamination.

Regardless, it can be concluded that for both template sets the suggested star-galaxy separation strategy is highly effective, with a failure rate of around

¹⁵It is stressed that because the ratio of stars to galaxies in the truth catalogue is not representative of the true numbers in the DES+VIDEO fields, the stellar contamination does not give an accurate view of the absolute expected fraction of misclassified stars. However, the values are nevertheless valuable for a comparison between different photo-z configurations; because all the verification tests use the same truth catalogue, they give an indication of which setup is likely to contain the most contamination from stars.

¹⁶The two configurations produce different adaptive offsets, so the $\chi_{v,\text{star}}^2$ may also vary ever so slightly.

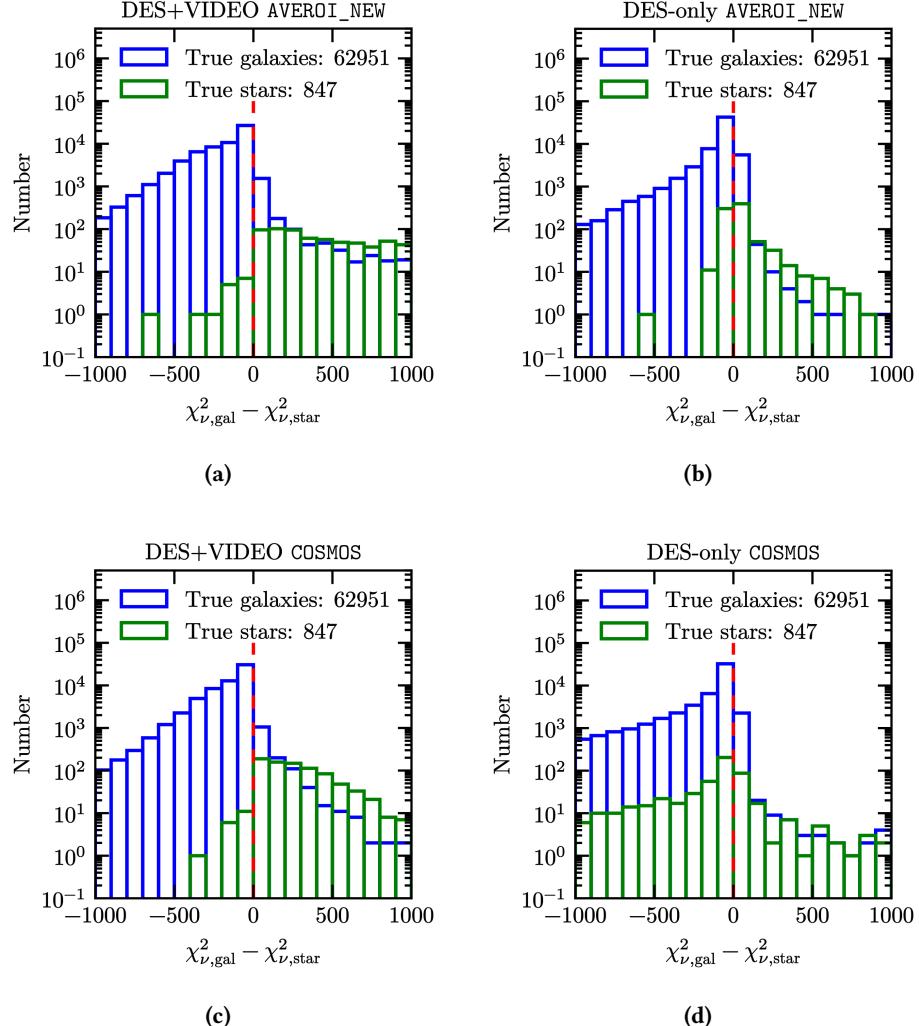


Figure 3.6 The star-galaxy separation strategy proposed in this thesis, which is based on reduced χ^2_ν values for the best-fit star and galaxy templates computed by **LePHARE**. Objects consist of all non-duplicate sources in the DES+VIDEO footprint matched to a truth catalogue (see the text), comprising a total of 847 true stars (green bars) and 62 951 true galaxies (blue bars). The red line shows the separation threshold $\chi^2_{\nu,\text{gal}} - \chi^2_{\nu,\text{star}} = 0$. The technique in this chapter classifies objects to the left of this line as galaxies, and objects to the right as stars. Figures (a) and (c) show the results for template fitting using both DES and VIDEO data for the **AVEROI_NEW** and **COSMOS** templates respectively. Figures (b) and (d) show the results using DES data only. Further discussion is provided in the text.

2–3% for both stars and galaxies. Despite this excellent prognosis though, it must be remarked that most of the objects in the validation truth catalogue are relatively bright. Such sources are generally expected to have the highest quality photometry, and it is possible that the current method is less accurate at the faint magnitudes that are typical in high-redshift research, where photometric uncertainties are higher. However, this issue affects any star-galaxy separation technique based on broadband imaging alone, including the morphology-based strategies mentioned in Section 3.7.1.1. All things considered, the good performance of the proposed χ_v^2 -based technique on the validation truth catalogue instills confidence that the technique will be competitive at high redshifts as well.

3.7.1.4 The effect of near-IR data

It is interesting to investigate to what extent the star-galaxy separation is helped by the inclusion of near-IR data. To answer this question, **LePHARE** was run on the truth catalogue sources using only DES *griz* photometry (via the **des** setups introduced in Section 3.4.2), and the separation procedure from the previous section was repeated. This time, classification based on the **AVEROI_NEW** templates selects 6128 (supposed) stars and 57 670 galaxies. The selection correctly identifies only 531 (64%) true stars, and falsely classifies 316 (36%) as galaxies, leading to a stellar contamination rate of 0.55%. Regarding true galaxies, 57 354 sources are correctly retrieved, implying a galaxy completeness of 91%. The remaining 5597 (8.9%) true galaxies are misclassified as stars. For the **COSMOS** templates, which find 2515 stars and 61 283 galaxies, most numbers are even less favourable; only 190 (22%) of the actual stars are correctly classified, and a majority of 657 (78%) true stars are wrongly identified as galaxies, leading to a substantially worse stellar contamination of 1.1%. The galaxy completeness is slightly better than for the **AVEROI_NEW** templates, with 60 626 (96%) of the true galaxies classified properly. The other 2325 (3.6%) true galaxies are wrongly judged to be stars. Figures 3.6b and 3.6d provide a more detailed view of the performance, plotting the $\chi_{v,\text{gal}}^2 - \chi_{v,\text{star}}^2$ distributions for both template sets.

A direct comparison between these DES-only results and the earlier DES+VIDEO outcomes clearly illustrates that the χ_v^2 star-galaxy separation method proposed in this thesis is significantly more successful when near-IR data is included. This result holds regardless of what template set is used. The effect on stellar contaminants is particularly strong; when removing the near-IR data,

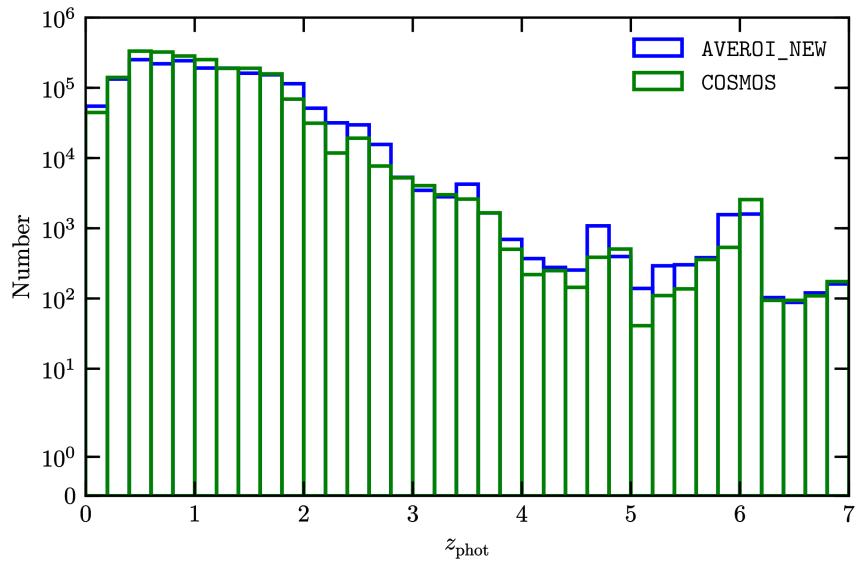


Figure 3.7 Photometric redshift distributions for all galaxies with good photometry in the final DES+VIDEO catalogue, plotted in bins of $\delta z = 0.2$. The blue and green histograms show the photo-zs for the **AVEROI_NEW** and **COSMOS** templates respectively. The plot exclusively includes objects classified as galaxies, selected by applying the star-galaxy separation strategy from Section 3.7.1.2 to the $\chi^2_{v,\text{gal}}$ output from the respective template sets. Sources with unreliable photometry have been removed via their **FLAGS_WEIGHT** values as described in the text.

the stellar contamination rate increases 22 times for **AVEROI_NEW** and 35 times for **COSMOS**. Additionally, the galaxy completeness drops by a factor of 1.06 for **AVEROI_NEW** and for 1.01 for **COSMOS**. These findings highlight the importance of near-IR data for SED-based star-galaxy separation methods. The positive effect on stellar contamination is expected to be especially important at high redshifts, where stars are a common contaminant in candidate galaxy samples.

3.7.2 Total galaxy redshift distribution

The star-galaxy separation classifier proposed in Section 3.7.1.2 was applied to the full DES+VIDEO dataset. Of the 2 443 576 objects (including duplicates) in the catalogue, 1 976 201 are classified as galaxies with the **AVEROI_NEW** templates, and 2 191 030 with the **COSMOS** SEDs.

Figure 3.7 presents the redshift distribution of all the selected galaxies with good photometry, showing the results for both template sets. To reduce the

chances of unreliable photometry causing inaccurate photo-zs, sources with non-zero **FLAGS_WEIGHT** values have been removed from the plot (using the same procedure as before in Sections 3.5 and 3.7.1.3), leaving 1 919 510 galaxies for **AVEROI_NEW** and 2 129 053 for **COSMOS**. The histograms show that the two configurations produce a broadly similar distribution, with several noteworthy peaks at high redshifts. These features originate from transitions between filters, and correspond to changes in the dropout bands of Lyman-break galaxies. For instance, the strong peak at $z_{\text{phot}} \equiv 6$ is situated around the $z_{\text{phot}} = 5.99$ point where r -dropouts end and i -dropouts begin. The similar behaviour around $z \equiv 4.8$ is caused by the changeover between g -dropouts and r -dropouts at $z_{\text{phot}} = 4.84$. The prior is likely to be a contributing factor as to why galaxies tend to cluster around these peaks; because this weighting factor leads **LePHARE** to favour the lowest redshift solution that still provides a good fit, it will tend to push the photo-z to the lowest allowable value. For Lyman-break galaxies, this lowest value is dictated by the dropout bands (e.g. for an i -dropout, solutions below $z_{\text{phot}} = 5.99$ are not compatible with the observed lack of i -band flux). Another contributing factor to the feature at $z \equiv 6$ in particular consists of (transient) contaminants that are only observed in the z -band, and thus partly mimic the Lyman-break behaviour of an i -dropout. Section 4.2.4.1 will discuss such objects in detail.

The final galaxy photo-z distribution in Figure 3.7 concludes the photometric redshift part of this thesis. Equipped with all the necessary photo-z fitting data, the DES+VIDEO catalogue is ready for the selection of high-redshift galaxies, which is the topic of the next chapter.

Chapter 4

High-redshift Candidates

4.1 Motivation

The previous chapter concentrated on the task of optimising photometric redshift estimates for the DES+VIDEO catalogue, focused especially on optimal results at high redshifts. Furthermore, it also proposed a method for separating stars and galaxies based on the best-fit stellar and galaxy SEDs obtained during the photo-z computation. Together, these results can be used to identify a collection of $z \gtrsim 5$ high-redshift candidates. The current chapter now describes the process of isolating such a sample. The purpose of this research is to investigate an optimal selection strategy, and to select a preliminary collection of high-redshift candidates. In future work, such a sample could be used to constrain the properties of galaxies in the early stages of evolution. Examples of such properties have been outlined in the introduction (see Section 1.3.2.2).

Although a strategy based on photometric redshifts could in principle retrieve any high-redshift galaxy with magnitudes brighter than the DES+VIDEO depths, the high photometric errors of objects close to the survey’s limiting magnitude can lead to uncertain redshift estimates. To make sure redshifts for the candidate sample are maximally secure, this thesis concentrates on galaxies with i -band or z -band magnitudes below $m_{\text{AB}} < 25$. These bright sources are particularly interesting, because as explained in Section 1.3.2.1, the known samples of such objects are currently small. Section 1.3.3 made the case how, thanks to the unprecedented size of a deep search area in the DES+VIDEO catalogue, the current work can thus provide a substantial increase in potential candidates.

Moreover, this thesis focuses predominantly on Lyman-break galaxies, which are most likely to dominate high-redshift samples. Despite the fact that a selection method based on photo-zs can theoretically also recover sources of other spectral types (provided these are represented in the template set), such objects are expected to be much less bright in the rest-frame UV (observed-frame optical), so they are unlikely to satisfy the brightness criteria. Furthermore, their spectral features are much less distinct in the optical to near-IR wavebands, which makes them almost impossible to separate from stars or red low-redshift contaminants. For all the reasons listed above, the targets in this thesis thus consist of *bright Lyman-break galaxies*.

4.2 High-redshift galaxy candidate selection process

4.2.1 Brief overview

After a careful investigation of possible selection criteria for high-redshift sources, the eventual strategy that was settled on consists of multiple stages. The task of the current part of this chapter is to describe this procedure. In order to clarify the structure of the upcoming discourse, a brief summary is listed below.

- Step 1** The first stage is made up of a series of broad initial cuts, which are presented in Section 4.2.3. The aim of this round is to extract a preliminary sample of bright high-redshift galaxy candidates.
- Step 2** The next stage then involves visual examination of the preliminary candidates. Each object is assessed individually by looking at its imaging, photometry, and best-fit SEDs. Section 4.2.4 reports on this inspection, which has uncovered several major sources of contamination.
- Step 3** Lastly, a second selection round consists of another round of cuts to address each type of contaminant identified in step 2, resulting in a final sample of secure candidates. Section 4.2.5 describes this second selection stage.

4.2.2 Template choice

Before any candidate selection could occur, a choice had to be made regarding which template set to use for the photometric redshifts. To explain this decision

process, the current section revisits the issue of comparing the **AVEROI_NEW** and **COSMOS** template sets, this time specifically in terms of high-redshift performance. Section 3.6.1.1 discussed how in general, the plots and metrics from Figure 3.3 and Table 3.1 indicate that the **COSMOS** templates produce better accuracy over the full redshift range. However, the sample in those tests is almost entirely dominated by galaxies at $z_{\text{spec}} < 1.5$, so this general conclusion may not translate to higher redshifts. The only quantities to test the high-redshift performance exclusively are N_{int} and N_{good} . For these metrics, there appears to be a trade-off between contamination and true high-redshift candidates when it comes to template choice. Regarding contamination, the **COSMOS** SEDs show only 51 low-redshift galaxies in a high-redshift¹ sample (N_{int}), whereas there are 104 for the **AVEROI_NEW** templates. But that higher purity comes at a cost: the **COSMOS** configuration only retrieves 14 galaxies with accurate high redshifts (N_{good}), compared to 23 for **AVEROI_NEW**. These quantities suggest that the number of true high-redshift galaxies in a photo-z selected sample is higher for the **AVEROI_NEW** configuration. This counts strongly in favour of those templates, especially since it is also anticipated that the higher contamination can be managed in another way. Figure 3.3 clearly shows that the number of galaxies at $z_{\text{phot}} \gtrsim 5$ is roughly equal between the two template sets. As a result, visual inspection (step 2 of the candidate selection procedure) is not considerably harder with the **AVEROI_NEW** templates, and targeted cuts can be applied in step 3 to remove any low-redshift interlopers. Besides, the star-galaxy separation validation in Section 3.7.1.3 suggested that a galaxy sample selected based on the **AVEROI_NEW** SEDs may also contain less stellar contamination. As stars are known to be a significant contaminant in bright high-redshift samples (Willott et al. 2013; Bowler et al. 2015), this provides another edge over the **COSMOS** configuration. Together, all the above reasons led to the decision to base the high-redshift search in this thesis on the **AVEROI_NEW** templates.

4.2.3 Broad initial cuts

After the **AVEROI_NEW** templates were established as the choice of template set, the corresponding photometric redshifts could be used to commence the search

¹The reader is reminded that the definition of high-redshift for N_{int} and N_{good} is slightly different from the $z \gtrsim 5$ definition used in the rest of this thesis, see Section 3.5.2.

for high-redshift candidates. It was decided to use the full set of 2 443 576 DES+VIDEO sources for this search. This choice to retain duplicates was deliberate, to inspect whether multiple instances of the same candidate (with slightly different photometry) make it into the final sample. As outlined in the overview from Section 4.2.1, the first step of the candidate search consists of an initial selection round to narrow the full DES+VIDEO catalogue down to a preliminary sample of high-redshift galaxies. The following exposition will describe in detail the five cuts that make up this first stage.

4.2.3.1 Weight flag cut

The DES and VIDEO images contain some regions with unreliable measurements, due to e.g. bad pixels around chip gaps, faulty CCDs, saturated bright stars, or transients. These issues have been described in Sections 2.1.1.5 and 2.1.2.4, where it was also explained that objects in or directly adjacent to these bad regions are flagged by **SExtractor** with a non-zero value for the **FLAGS_WEIGHT** parameter. To remove candidates that could be biased by unreliable photometry, it was therefore required that **FLAGS_WEIGHT** = 0 in all observed bands², the same strategy as for the assembly of the spectroscopic testing sample described in Section 3.5.1. The resulting weight flag selection can be summarised as follows³:

$$\begin{aligned} \text{FLAGS_WEIGHT_G} &= 0 \wedge \text{FLAGS_WEIGHT_R} = 0 \wedge \text{FLAGS_WEIGHT_I} = 0 \\ \wedge \text{FLAGS_WEIGHT_Z} &= 0 \wedge \text{FLAGS_WEIGHT_Y} = 0 \wedge \text{FLAGS_WEIGHT_Z2} = 0 \\ \wedge \text{FLAGS_WEIGHT_Y2} &= 0 \wedge \text{FLAGS_WEIGHT_J} = 0 \wedge \text{FLAGS_WEIGHT_H} = 0 \\ \wedge \text{FLAGS_WEIGHT_K} &= 0, \quad (4.1) \end{aligned}$$

where **FLAGS_WEIGHT_Z** and **FLAGS_WEIGHT_Y** denote the DES *z*-band and *Y*-band, and **FLAGS_WEIGHT_Z2** and **FLAGS_WEIGHT_Y2** refer to the VIDEO *Z*-band

²This includes the DES *Y*-band, which was not included in the photo-z fitting. The motivation for this is the same as in Section 3.5.1.

³Naturally, this form of the expression is for objects that are observed in all filters. For objects that are not observed in all bands, the weight flag criterion is not imposed on the unobserved bands.

and Y -band. After applying this cut to the DES+VIDEO catalogue, 2 375 871 objects remained.

4.2.3.2 Redshift selection cut: definition of $z \sim 5$ and $z \sim 6$ samples

The next step consists of a redshift cut, and the discussion now turns to the task of establishing the appropriate photo-z boundaries for a high-redshift sample. Firstly, this involves specifying lower and upper bounds on the redshift. It was decided to impose limits that ensure candidates are as secure as possible. As discussed in Section 1.3.1.2, photo-z estimates at high redshifts use the Lyman break as the main discerning spectral feature. Therefore, such measurements are generally most reliable when this break falls between two filters (which is when it shows up most distinctively in the photometry). The most secure lower bound for $z \geq 5$ galaxies in the DES+VIDEO catalogue thus lies at $z_{\text{phot}} = 4.83$, the redshift at which the Lyman break moves from the r -band into the i -band. This cutoff also guarantees that all candidates in the sample show dropout behaviour in at least 2 DES bands, increasing the confidence that they are true high-redshift galaxies.

Regarding an upper redshift bound, the principle of maximum security entails a threshold of $z = 6.5$. Even though i -dropouts can technically be detected out to $z_{\text{phot}} = 7.77$, it is unlikely that the DES+VIDEO data can reliably identify credible candidates above $z = 6.5$ given the survey depths. Indeed, the vast majority of studies that have found $z \geq 6.5$ candidates have either used *HST* data, or ground-based imaging significantly deeper than DES+VIDEO (Oesch et al. 2010b; McLure et al. 2010; Bouwens et al. 2015; Bowler et al. 2014). Bowler et al. (2012) did find ~ 10 possible very bright $z \sim 7$ candidates in ground-based data of similar limiting magnitudes as DES+VIDEO, but their selection method was specifically geared to this redshift range. In particular, the forced photometry in their study was extracted based on a combined $Y + J$ detection image, as opposed to the riz data used for the DES+VIDEO catalogue. Because all secure Bowler et al. (2012) candidates showed z -band emission fainter than the DES 5σ depths, those objects would not be reliably detected in the DES+VIDEO catalogue. For these reasons, any photometric redshifts for $z \geq 6.5$ galaxies were deemed too uncertain. It may still be possible to search for bright $z \sim 7$ galaxies in DES+VIDEO using a similar strategy to Bowler et al. (2012) based on VIDEO Y - and J -band detections, but this is left as a possibility for future research.

To be able to compare the $4.83 \leq z_{\text{phot}} < 6.5$ sources in this thesis to samples in the literature, the candidates must be divided into $z \sim 5$ and $z \sim 6$ bins. The remainder of this section discusses how to define the optimal limits for this split. A simplistic first suggestion would be to identify the $z \sim 5$ sample with r -dropouts (i.e. $4.83 \leq z_{\text{phot}} < 5.99$) and the $z \sim 6$ sample with i -dropouts (i.e. $5.99 \leq z_{\text{phot}} < 6.5$). However, it will now be argued that this choice is undesirable given the magnitude cut imposed in a later step. That cut will be described in Section 4.2.3.5 and consists of a $m_{\text{AB}} < 25.0$ limit in the chosen detection band. It is natural to choose the i -band as the required detection filter at $z \sim 5$ and the z -band at $z \sim 6$, given the range of these two filters. However, if one were to follow the earlier suggestion where the $z \sim 5$ sample would be directly equated with r -dropouts, this would correspond to an $i_{\text{AB}} < 25.0$ limit for all sources out to $z \leq 5.99$. This is unideal, given that any r -dropouts at $z \gtrsim 5.5$ are redshifted halfway out of the i -band, causing significantly dimmer i -band fluxes. In fact, it was verified that only a handful of $z_{\text{phot}} \geq 5.5$ objects in the catalogue would make it into the high-redshift sample if the $i_{\text{AB}} < 25.0$ criterion were to be imposed on all $4.83 \leq z_{\text{phot}} < 5.99$ objects. The suggested simplistic definition of the $z \sim 5$ sample would thus omit⁴ a significant number of possible candidates between $5.5 \leq z < 5.99$. Therefore, it was considered undesirable to equate the $z \sim 5$ sample directly with the r -dropout redshift range. Instead, the split has been defined as follows:

$$\begin{aligned} 4.83 \leq z_{\text{phot}} < 5.5 &\quad \text{for the } z \sim 5 \text{ sample,} \\ 5.5 \leq z_{\text{phot}} < 6.5 &\quad \text{for the } z \sim 6 \text{ sample.} \end{aligned} \tag{4.2}$$

For r -dropouts between $5.5 \leq z_{\text{phot}} < 5.99$, the magnitude selection thus amounts to $z_{\text{AB}} < 25.0$. This is a more suitable choice, since any emission that is redshifted out of the i -band instead boosts the z -band flux. Beyond $z \gtrsim 5.5$, the z -band is thus likely to contain more of the flux (on the plausible assumptions that r -dropout LBGs show a relatively flat observed-frame continuum throughout the i -band⁵, and that any Lyman- α emission line is not so strong that it severely biases the

⁴The issue is even more urgent given the fact that Lyman-break galaxies enter the VIDEO Z -band at $z > 5.81$. In the regions where Z -band data is available, a method that is not sensitive to galaxies at $5.5 \leq z < 5.99$ could remove a significant number of reliable candidates from the high-redshift sample.

⁵Even though LBGs are known to show blue rest-frame UV emission, their observed-frame SEDs are stretched out along the wavelength-axis as their redshift increases. Since $\lambda_{\text{obs}} = (1+z)\lambda_{\text{em}}$ (see Equation 1.1), the flux in a wavelength interval Δ is stretched by a factor of 6.5 by $z=5.5$. Over

magnitudes). All in all, the boundaries in Equation 4.2 ensure that a maximal number of good $5.5 \leq z_{\text{phot}} < 5.99$ candidates are included in the final sample⁶. After the cuts in Equation 4.2 were applied, the $z \sim 5$ and $z \sim 6$ samples consisted of 14 479 and 9899 objects respectively.

4.2.3.3 Star-galaxy separation

The next selection step aims to make sure that any high-redshift candidate is indeed truly a galaxy and not a star. Section 3.7.1 described how this thesis has explored a χ^2_v -based approach for star-galaxy separation. This method isolates galaxies via Equation 3.24, which is restated below for the sake of clarity:

$$\chi^2_{v,\text{gal}} - \chi^2_{v,\text{star}} < \Delta\chi^2_v.$$

Section 3.7.1.3 found excellent separation performance with this strategy when using a straightforward value of $\Delta\chi^2_v = 0$ (which simply requires that any galaxy must show a better fit to a galaxy template than to a star template). Naturally, a negative threshold of $\Delta\chi^2_v$ imposes an even stricter selection. Because stars can be a major contaminant among bright high-redshift candidates (Willott et al. 2013; Bowler et al. 2015), it was deemed important to reduce the chances of stars entering the sample. The threshold has therefore been set to $\Delta\chi^2_v = -1.5$, and the resulting star-galaxy separation cut can be summarised as follows:

$$\chi^2_{v,\text{gal}} - \chi^2_{v,\text{star}} < -1.5. \quad (4.3)$$

the ~ 1500 Å width of the i -filter, this leads to a relatively flat observed spectrum, as can be noted in the spectrum of a model $z=6$ galaxy in Figure 1.2.

⁶The detail-oriented reader may wonder why the $z \sim 5$ sample was not instead equated with $4.83 \leq z_{\text{phot}} < 5.99$ r -dropouts, while imposing a looser magnitude cut of $i_{\text{AB}} < 25.0 \vee z_{\text{AB}} < 25.0$. This option was indeed investigated, but it was found that such a criterion led to the inclusion of roughly twice as many galaxies at $z \lesssim 5.5$. These additional galaxies did not pass the $i_{\text{AB}} < 25.0$ requirement, but were instead selected due to their $z_{\text{AB}} < 25.0$ emission. Given that this phenomenon was observed all the way out to $z \sim 4.83$, it is highly unlikely that many of these galaxies are true Lyman-break sources. This is because true $z \sim 4.83$ Lyman-break galaxies are expected to show blue SEDs redwards of the Lyman break, with the i -band emission comparable to or stronger than the z -band flux. Instead, the additional candidates are likely low-redshift contaminants or stars with more gradually sloping red spectra. The idea of imposing a $i_{\text{AB}} < 25.0 \vee z_{\text{AB}} < 25.0$ magnitude cut on an r -dropout $z \sim 5$ sample was therefore rejected, as it would drastically increase the contamination.

After applying this criterion, the $z \sim 5$ and $z \sim 6$ samples contained 572 and 3504 galaxies respectively.

4.2.3.4 Galaxy fit

In addition to eliminating stars, it is also important to remove objects that do not show an adequate fit to any galaxy template. Such sources are either not galaxies at all, or they are galaxies that cannot be fitted well by any galaxy SED in the template set (e.g. due to inaccurate photometry or template incompleteness). Objects in the former category naturally do not belong in an LBG sample, and sources in the latter category are at risk of spurious photometric redshifts. Therefore, it was required that any candidate must have a low galaxy $\chi^2_{\nu,\text{gal}}$:

$$\chi^2_{\nu,\text{gal}} < 10. \quad (4.4)$$

After this selection step, 385 and 1011 candidates were left in the respective $z \sim 5$ and $z \sim 6$ samples.

4.2.3.5 Magnitude cut

The search in this thesis specifically concentrates on bright (i.e. $m_{\text{AB}} < 25.0$) galaxies, as explained at the very start of the current chapter. To select only the most luminous sources, a $m_{\text{AB}} < 25.0$ magnitude cut was therefore imposed on the detection band for each high-redshift sample. As laid out previously in Section 4.2.3.2, this detection band was chosen to be the i -band for the $z \sim 5$ sample and the z -band for the $z \sim 6$ sample. For both samples, the imposed $m_{\text{AB}} < 25.0$ limit is ≈ 1.5 mag AB brighter than the 5σ depths of the very deepest regions (see Figure 2.7).

For the purpose of isolating maximally secure candidates, it was decided that the desired $m_{\text{AB}} < 25.0$ magnitude must be measured to a significance of at least 7.5σ , which is similar to the 6σ – 10σ thresholds used in comparable studies (McLure et al. 2009; Willott et al. 2013; Bowler et al. 2015). A 7.5σ significance corresponds to a signal-to-noise ratio of $S/N \equiv F/\sigma_F = 7.5$, where F and σ_F symbolise the flux and flux error respectively. The required significance can thus be expressed as a magnitude error σ_m in the following way:

$$\sigma_m = \frac{2.5}{\ln 10} \frac{\sigma_F}{F} = \frac{2.5}{\ln 10} \frac{1}{S/N} = 0.1448. \quad (4.5)$$

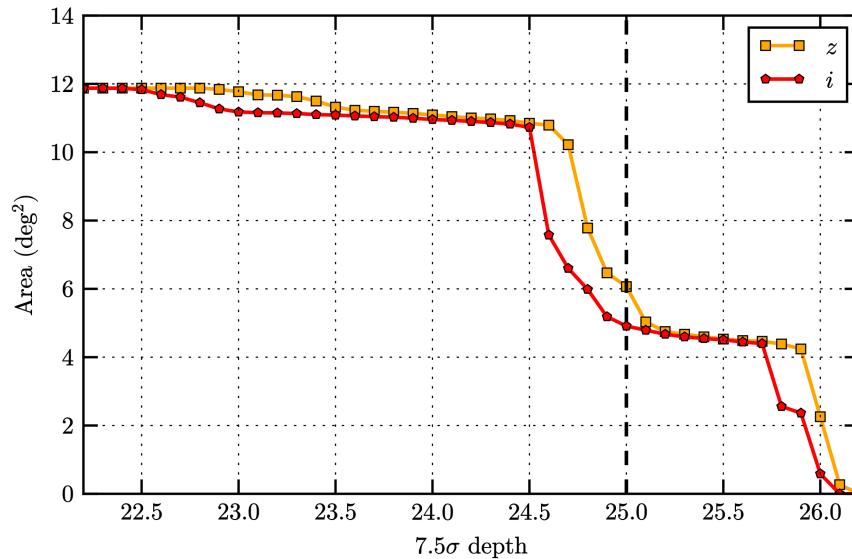


Figure 4.1 The area (with non-zero weight flags, see Section 4.2.3.1) which is covered to a given 7.5σ depth as a function of that depth, for both the i -band and z -band data. The dashed vertical line signifies the $m_{AB} < 25.0$ magnitude limit imposed in the candidate selection process.

Putting together all the requirements described above, the magnitude cuts amount to:

$$\begin{aligned} i_{AB} < 25.0 \quad \wedge \quad \sigma_i < 0.1448 &\quad \text{for the } z \sim 5 \text{ sample,} \\ z_{AB} < 25.0 \quad \wedge \quad \sigma_z < 0.1448 &\quad \text{for the } z \sim 6 \text{ sample.} \end{aligned} \quad (4.6)$$

After these limits were applied, the $z \sim 5$ and $z \sim 6$ samples consisted of 119 and 139 candidates respectively.

For later reference, it is useful to point out that the DES 7.5σ depths are shallower than $m_{AB} = 25.0$ in large regions of the DES+VIDEO footprint. Therefore, the magnitude/depth cut in Equation 4.6 has essentially restricted the high-redshift search to a smaller effective area — i.e. the area with a 7.5σ limiting magnitude of $m_{7.5\sigma} > 25.0$. This is illustrated in Figure 4.1, which plots the available area to a given depth against depth (after excluding areas with bad weight flags defined in Section 4.2.3.1). There is 6.1 deg^2 available to a 7.5σ limiting magnitude of $i_{AB} < 25.0$, and 4.9 deg^2 to $z_{AB} < 25.0$.

CANDIDATE CUTS			
$z \sim 5$ SAMPLE		$z \sim 6$ SAMPLE	
cut definition	remaining	cut definition	remaining
Total catalogue ⁱ	2 443 576	Total catalogue ⁱ	2 443 576
FLAGS_WEIGHT ⁱⁱ = 0	2 375 871	FLAGS_WEIGHT ⁱⁱ = 0	2 375 871
$4.83 \leq z < 5.5$	14 479	$5.5 \leq z < 6.5$	9899
$\chi^2_{v,\text{gal}} - \chi^2_{v,\text{star}} < -1.5$	572	$\chi^2_{v,\text{gal}} - \chi^2_{v,\text{star}} < -1.5$	3504
$\chi^2_{v,\text{gal}} < 10$	386	$\chi^2_{v,\text{gal}} < 10$	1011
$i_{\text{AB}} < 25.0 \wedge \sigma_i < 0.1447 (7.5\sigma)$	119	$z_{\text{AB}} < 25.0 \wedge \sigma_z < 0.1447 (7.5\sigma)$	316

ⁱThis catalogue consists of all sources in the DES+VIDEO footprint, including duplicates from overlapping DES or VIDEO tiles. ⁱⁱThe requirement on **FLAGS_WEIGHT** was imposed in all (observed) filters, so that ‘**FLAGS_WEIGHT** = 0’ is short for ‘**FLAGS_WEIGHT_G**=0**FLAGS_WEIGHT_R**=0**FLAGS_WEIGHT_I**=0**FLAGS_WEIGHT_Z**=0 \wedge **FLAGS_WEIGHT_Y**=0 \wedge **FLAGS_WEIGHT_Z2**=0 \wedge **FLAGS_WEIGHT_Y2**=0 \wedge **FLAGS_WEIGHT_J**=0**FLAGS_WEIGHT_H**=0**FLAGS_WEIGHT_K**=0’, where Z and Y refer to the DES z-band and Y-band, and Z2 and Y2 refer to the VIDEO Z-band and Y-band.

Table 4.1 An overview of the selection criteria that have been applied to the total DES+VIDEO catalogue to obtain a preliminary sample of high-redshift candidate galaxies for visual inspection and further selection.

4.2.3.6 Conclusion of first round cuts

The magnitude cuts above concluded the first selection round for the high-redshift search. Table 4.1 summarises all five selection criteria from this stage. Together, these broad initial cuts have generated a preliminary sample of 119 $z \sim 5$ and 319 $z \sim 6$ candidates, to be narrowed down further by visual inspection and a second round of cuts described in the following sections.

The observant reader may have noticed that the number of $z \sim 6$ candidates in the preliminary sample is significantly higher than the number of $z \sim 5$ candidates, which conflicts with the known behaviour of cosmic structure formation (as reflected, for instance, in the luminosity functions in Figure 1.3, which show that the abundance of bright galaxies increases with decreasing redshift). This number disparity is partly caused by the fact that the $z \sim 6$ sample includes a slightly wider redshift range than the $z \sim 5$ sample. However, it was found during visual inspection that the main cause of the number difference is contamination by transients and artefacts in the $z \sim 6$ sample. This issue will be addressed further in Section 4.2.4.1.

4.2.4 Visual inspection

The next step in the selection process visually inspected each preliminary candidate. The aim of this stage, which is the topic of the current section, is to isolate any remaining contaminants. For this purpose, the author created a custom piece of code to produce an inspection image for each candidate object. Figure 4.2 shows an example of such an image.

Each inspection image consists of two parts. The top half is based on code written by Sophie Reed, and shows an array of fourteen 30 arcsec by 30 arcsec postage stamp cutouts of the object in various types of imaging. Nine of these cutouts show the nine DES+VIDEO bands used for computing photo-zs ($grizZYJHK_s$), with images left blank in bands where no data exists. The cutouts array also displays the DES $r + i + z$ detection image, as well as two false-colour images constructed from $g + r + i$ and $r + i + z$ images. The final two slots in the postage stamp array are filled with mid-infrared imaging from the *Spitzer* Extragalactic Representative Volume Survey (SERVS; Mauduit et al. 2012) in the IRAC CH1 (3.6 μ m) and CH2 (4.5 μ m) wavebands. This survey was conducted by the *Spitzer* Space Telescope in its post-cryogenic phase and covers

an area of 18 deg^2 largely overlapping the DES+VIDEO footprint. With depths around $m_{\text{AB}} \sim 23.1$ and a PSF FWHM of 2 arcsec, its image quality is considerably poorer than DES+VIDEO. Because of this, the SERVS data has not been used for the catalogue and photometric redshifts in this thesis. Evidence that the inclusion of IRAC fluxes does not improve photo-zs for many codes (around half of codes tested by Hildebrandt et al. 2010, including **LePHARE**) contributed to this decision. However, since the raw imaging does contain information that reflects the amount and extent of mid-IR emission, SERVS cutouts were considered valuable for the purpose of visual inspection.

The second half of the inspection image consists of an SED plot that illustrates the photo-z fitting results for the candidate object in question. Measured magnitudes and magnitude errors are plotted in black for every (observed) filter, with the best-fit galaxy, star, and QSO templates overlaid in blue, yellow, and green respectively. For some candidates, **LePHARE** also found a secondary solution. As explained in Section 3.2.1.1, this occurs if the probability of the second-best galaxy redshift is above a certain user-specified threshold (chosen to be 0.1 in this thesis). If such a solution exists, it is overlaid in red. Magenta arrows show the 3σ depths in the area around the candidate object, as computed from Equations 2.6 and 2.8 (the top of the arrow corresponds to the 3σ depth). Inside the SED plot, a smaller inlay shows the probability distribution $P(z)$ as a function of redshift, computed by **LePHARE** during photometric redshift fitting (see Equation 3.3).

Together, the combination of postage stamp images and SEDs contains a wealth of photometric and morphological clues. This information was then used to identify possible contaminants in the preliminary sample. The process unveiled several major types of contamination, which will be discussed below.

4.2.4.1 Contamination from transient emission and imaging artefacts

The most easily distinguishable type of contamination consists of objects that are neither stars nor galaxies of any sort. Specifically, this category is comprised of imaging artefacts and transient emission. The latter kind was found to occur the most often, due to cosmic rays as well as moving extended sources such as asteroids, airplanes, or satellites. Transients were easily identified in the postage stamp cutouts via two main distinguishing properties. Firstly, the shape of the emission provided crucial evidence: spikes from fast-moving objects show up as

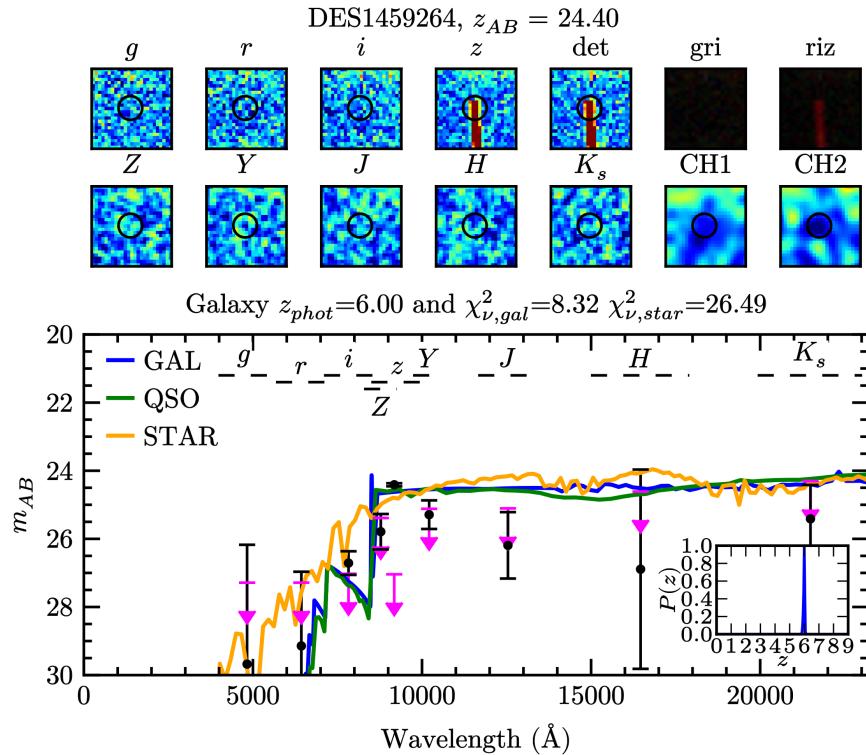


Figure 4.2 Verification image showing an example of contamination from a moving object. As the object happened to be in the field-of-view while a z -band exposure was taken, it only appears in that filter, mimicking a Lyman break between i and z . On top of that, the errors in the redder (VIDEO) filters are large enough that the random background flux is compatible with galaxy continuum emission so that its best galaxy fit passed the required $\chi_{\nu,\text{gal}} < 10.0$ preliminary cut.

bright elongated trails, and slower asteroids or satellites often look like resolved spheroids that are bright all over (as opposed to galaxies or stars whose light profiles look like ellipsoids, disks, or PSFs). Secondly, transients only show up in one filter, because they temporarily appeared in the field-of-view while that waveband was being exposed. The reason why such time-variable objects can infiltrate the preliminary sample is to do with this second property. If a bright transient happens to appear in the detection band, its observed SED can mimic the spectrum of high-redshift Lyman-break galaxies. This is because the lack of flux bluewards of the detection band can replicate the characteristic LBG dropout behaviour shortwards of the Lyman break, while the strong emission in the detection band can imitate the expected strong continuum flux in the rest-frame

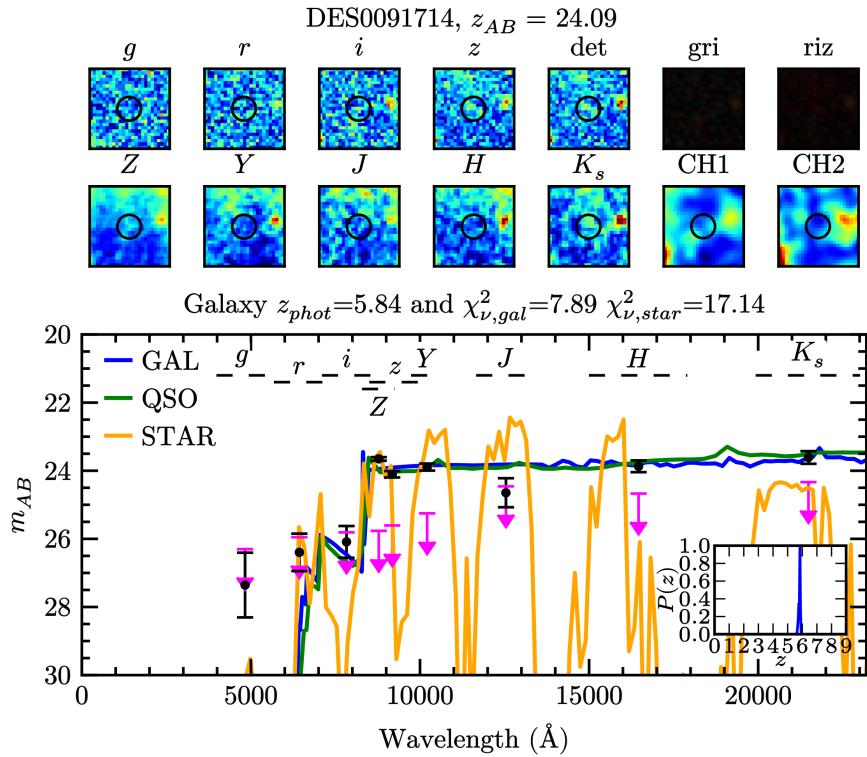


Figure 4.3 An example of contamination caused by imaging artefacts. Zooming out on the imaging for this contaminant revealed that the diffuse background emission in the postage stamps is caused by the extended PSF of a very bright nearby star. The random z -band spike in this PSF flux turns out to be bright enough to pass the **SExtractor** detection threshold, so that it shows up as a phantom detection in the DES+VIDEO catalogue. Because the nearby star is particularly luminous in the near-IR, its colours fairly closely replicate those of true high-redshift galaxies, which caused this phantom detection to make it into the preliminary sample.

UV. In some of these cases, the photometry redwards of the detection band is consistent with LBG continuum emission, for instance if the data contains sufficiently large photometric errors in these redder bands, or if the contaminant overlaps with a real object, a random spike in background emission, or an artefact (see below). **LePHARE** may assign such transients a high-redshift solution, which, if the errors in the redder bands are large enough, can have a $\chi^2_{\nu,gal}$ within the range from the preliminary cuts. An example of a transient contaminant is shown in Figure 4.2.

Imaging artefacts form another category of contaminant that was identified during visual inspection. Since imaging faults do not represent any external light source at all, these sorts of objects are ‘phantom’ detections. The most common type of artefact contamination was found to be extended emission in the PSF of very bright stars (including diffraction spikes). Occasionally, random fluctuations in this stellar emission are bright enough to pass the object detection threshold in **SExtractor**, causing them to appear as phantom objects in the DES+VIDEO catalogue. In some cases, their colours are sufficiently like true high-redshift galaxies that these sources entered the preliminary sample. Figure 4.3 shows an example of contamination from the extended PSF of a bright star.

Contamination from transients and artefacts was found exclusively in the $z \sim 6$ sample. This is the primary cause of the discrepancy between the number of $z \sim 5$ and $z \sim 6$ preliminary candidates noted previously in Section 4.2.3.6. The discussion will now turn to why this contamination specifically occurs at $z \sim 6$. Two main reasons can be identified. Firstly, the DES data is generally deeper than VIDEO, so the VIDEO photometric uncertainties are typically larger than those in DES. On top of that, near-infrared imaging also suffers from higher atmospheric background noise, which can lead to random bright spots within the VIDEO data, especially if the background subtraction was not performed perfectly. As a result, compared to DES, the VIDEO photometry is more likely to be consistent with the expected continuum emission from a high-redshift galaxy, even if there is no real object present. At $z \sim 6$, all imaging redwards of the detection band lies within VIDEO, so a z -band transient can mimic a true $z \sim 6$ LBG due to this lower VIDEO data quality. On the other hand, any i -band transient would also need to show bright continuum emission in the DES z -band to be compatible with a $z \sim 5$ LBG, which is less likely thanks to the higher DES image quality. The second reason why contamination from transients and artefacts is less probable among $z \sim 5$ candidates, is that a $z \sim 5$ candidate requires continuum emission to be detected in one additional filter compared to $z \sim 6$ objects (i.e. $izZYJHK_s$ instead of $zZYJHK_s$). If the object is not a true high-redshift LBG, the inconsistency between the measured contaminant flux and the high-redshift template flux in this additional filter will increase its $\chi^2_{v,\text{gal}}$, which significantly decreases the chances of this object making it into the preliminary candidate sample.

4.2.4.2 Contamination from low-redshift galaxies and QSOs

Visual inspection also uncovered low-redshift galaxies as another major type of contaminant. While these were not always quite as straightforward to identify as the earlier ‘junk’ objects, the inspection images nevertheless contain various key indicators. These distinguishing characteristics will be described below.

The first way to identify low-redshift contamination is via flux bluewards of the Lyman break. As explained in Section 1.3.1.2, true Lyman-break galaxies show virtually no emission shortwards of 912 Å, and a strong (redshift dependent) lack of flux between 912 Å and 1216 Å. Objects which contain blue flux incompatible with these limits can therefore not be true high-redshift galaxies. In addition to low-redshift interlopers, blue light could theoretically also originate from stars. However, this is much less likely in a high-redshift sample, since the types of stars that are compatible with the colours of $z \gtrsim 5$ galaxies are M, L, and T (brown) dwarfs, which are undetectable in blue filters. Regardless, even if it is stellar in origin, excess blue flux indicates contamination. For the $z \sim 5$ sample, such an excess consists of any emission in the g -band, and any r -band flux that is brighter than the typical $r_{\text{AB}} - i_{\text{AB}}$ drop values of around 1 to 2 (or more) expected for r -dropouts based on observational studies (Songaila 2004; Van der Burg et al. 2010). Similarly, contamination in the $z \sim 6$ sample can be confirmed through the presence of g -band emission, a 912 Å drop in r -band flux smaller than the expected $r_{\text{AB}} - z_{\text{AB}} \gtrsim 3$ break (such a drop manifests as a non-detection in r in almost all cases), or a 1216 Å drop less than the expected $i_{\text{AB}} - z_{\text{AB}}$ break of at least 2 or 3 (all the above colours have been derived from observations; Songaila 2004; Bowler et al. 2015; Willott et al. 2013). During visual inspection, these requirements exposed several preliminary candidates as contaminants, either because they show g -band emission in the cutouts, or because their Lyman breaks are not strong enough in the SED plots. On top of that, a further set of preliminary candidates was marked as highly uncertain on the basis that their imaging is too noisy⁷ in the designated blue filters to rule out an excess of blue flux. To ensure maximum purity of the final high-redshift sample, these dubious sources were also classified as contaminants. An example is shown in Figure 4.4.

A second strategy to identify low-redshift contaminants looked at the shape of their observed SEDs. Since their photometry must be sufficiently similar to

⁷These noisy regions lie mainly in the proximity of chip gaps.

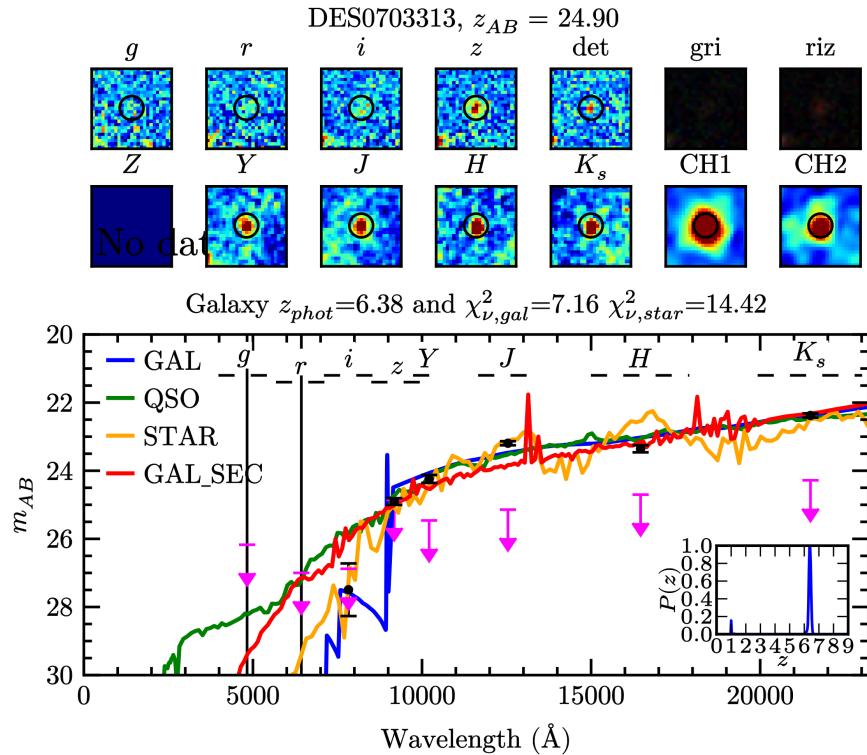


Figure 4.4 An example of low-redshift contamination likely caused by noisy photometry in the g - and r -band (due to close proximity to the chip gaps). On the whole, the object shows clear signs of a low-redshift origin, such as a gently sloping SED and bright *Spitzer* emission. Due to the high noise in the g - and r -band imaging, the object did not pass the **SExtractor** detection threshold in these bands, mimicking the flux measurements of an i -dropout. However, visual examination of the g - and r -band cutouts suggests that these wavebands do contain some flux that was missed due to the high noise.

true high-redshift galaxies to have passed the first round of preliminary cuts, their intrinsic spectra need to be exceptionally red, implying that they must be extremely dusty. The SEDs of intrinsically red sources generally show a more gently sloping increase than typical Lyman-break galaxies, and their observed colours often continue getting redder into the near-infrared regime (whereas Lyman-break galaxies show blue colours longwards of the Lyman break). Because they are so dusty, they also often display bright mid-IR emission in the *Spitzer* CH1 and CH2 cutouts. During visual inspection, several preliminary candidates were found to show such photometry indicative of a red low-redshift galaxy. It is suspected that this failure by **LePHARE** to correctly identify some

very red galaxies may be due to inaccurate photometry, or template incompleteness at the extreme red end. Regarding the former issue, the inspection image SED plots of many low-redshift contaminants indeed sometimes indicate that $\chi^2_{v,\text{gal}}$ might have been biased by one or two bands with very small errors, while the other bands look more compatible with a low-redshift fit. This hypothesis is supported by the fact that the best-fit solution for most (89%) objects in the preliminary sample has a reduced chi-squared of $\chi^2_v > 1$, which suggests that the photometric uncertainties in at least some bands may be slightly too small compared to the template set. For another class of low-redshift contaminants, it was found that the false high-redshift outcome has likely been biased by data in noisy regions (as previously seen in Figure 4.4) or nearby sources. Figure 4.5 presents an example of the latter cause.

Inbuilt features in **LePHARE** can provide further ways to distinguish potential low-redshift contaminants. One such strategy utilised secondary redshift solutions, which, if present⁸, are plotted in the inspection images. Several preliminary candidates were found to have secondary redshifts between $0 < z_{\text{sec}} < 2$, indicating that their photometry is compatible with a low-redshift origin. The contaminant previously introduced in Figure 4.4 is an example of such an object with a secondary solution around $z_{\text{sec}} \approx 1$. Because it is important that all final candidates are maximally secure, any preliminary candidates with a low-redshift secondary solution were considered contaminants.

A further useful feature in **LePHARE** is the ability to fit QSO templates (see Section 3.3.2), which helped to identify preliminary candidates that could be low-redshift quasars. Since each inspection image contains the best-fit QSO template, potential low-redshift QSOs could easily be recognised in the SED plots. During visual inspection, it was found that several preliminary candidates have a reasonable quasar solution at low-redshift. In many cases, the photometry for these sources indeed shows signs of low-redshift emission, such as gently sloping SEDs similar to those of the low-redshift galaxies found earlier. Again, to ensure maximal purity, objects with a good low-redshift QSO fit were flagged as contaminants.

⁸As explained in Sections 3.2.1.1 and 4.2.4, a secondary solution exists if the probability function $P(z)$ is larger than 0.1 at the second most probable photometric redshift.

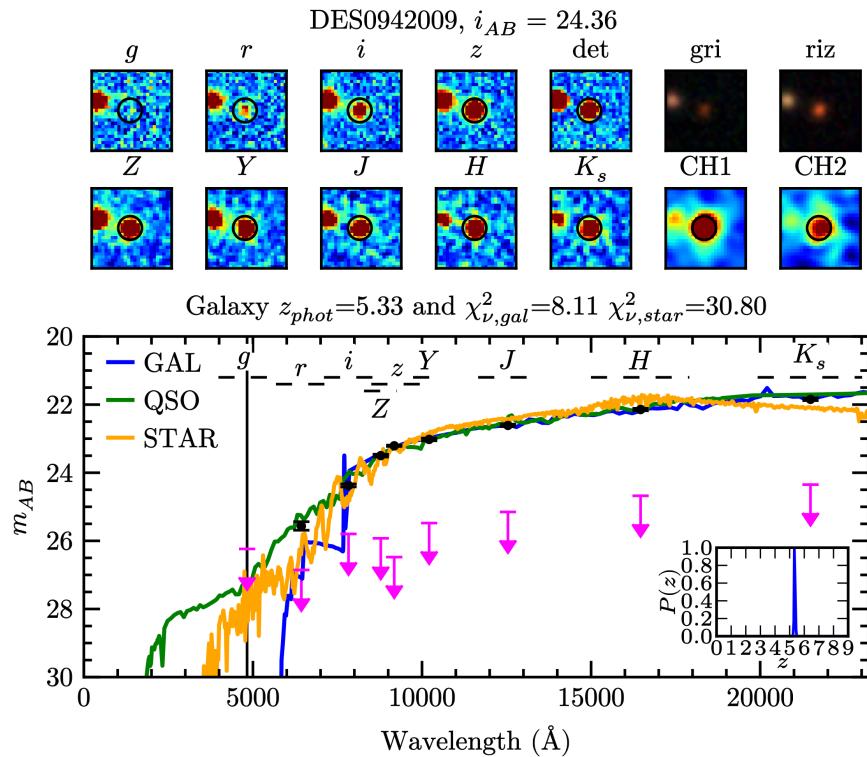


Figure 4.5 An example of low-redshift contamination likely caused by flux from a nearby source to the left. Like the low-redshift interloper in Figure 4.4, this object shows a gently sloping SED (note the lack of a strong $r - i$ break) and luminous **Spitzer** emission, both indicating a low-redshift origin. It is likely that the galaxy fit was biased by very small errors in some bands where the flux was boosted by a bright neighbour.

4.2.4.3 Contamination from stars

Stars constitute the final contaminant type that was revealed by visual inspection. Out of all the different categories of contamination, these sources were the most challenging to identify, because their inspection images contain the least distinctive features. Nevertheless, there are a few possible indicators. First of all, some stellar contaminants can in theory show blue emission, as explained during the discussion of low-redshift interlopers in Section 4.2.4.2. Secondly, the morphology in the postage stamps can provide some clues, since stars always appear as point sources, whereas a substantial fraction of bright galaxies are likely resolved. Unfortunately, it is expected that some of the $z \sim 5$ galaxies targeted in this thesis are also unresolved in the DES+VIDEO imaging, as

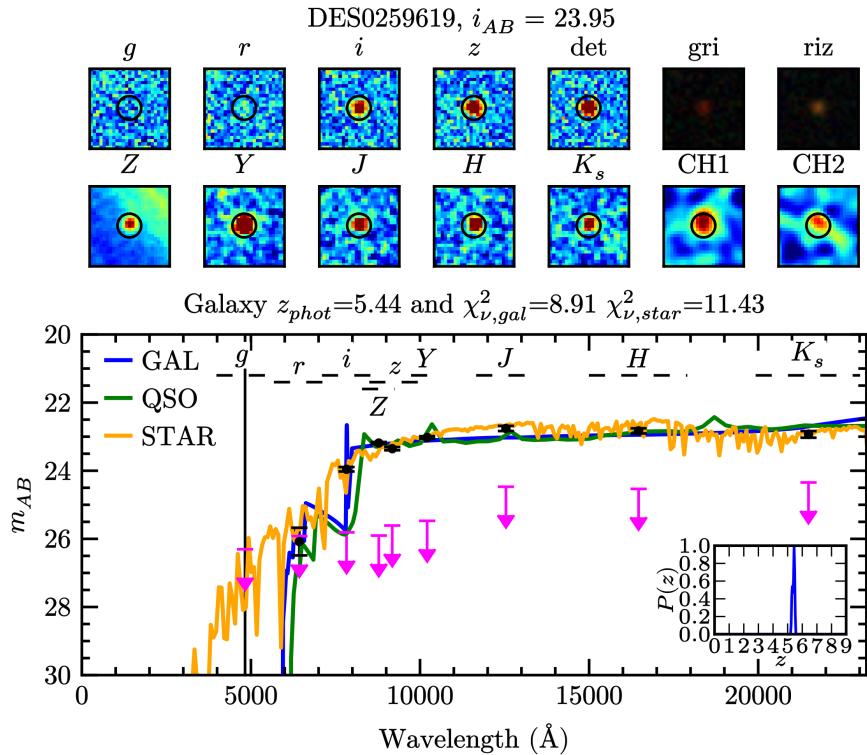


Figure 4.6 An example of a suspected stellar contaminant caused by biased photometry. The postage stamps for this object indicate point source emission in all filters, which, for an object this bright ($i_{AB} = 23.95$) suggests that the source is likely a star. The SED plots support this hypothesis, as they show photometry that is on the whole (more) compatible with the stellar template. The high photometric redshift is likely strongly influenced by extremely small error bars in the Z -filter. These small uncertainties, in turn, are caused by boosting of the Z -band flux due to background emission from the extended PSF of a very bright star (which was discovered upon zooming out on the imaging).

previously described in Sections 3.4.1.3 and 3.7.1.1. Nonetheless, if a very bright preliminary candidate shows PSF-shaped emission in all filters, this increases the likelihood that it is a stellar contaminant. The third and most reliable indicator for finding stars, however, is their photometry. During visual inspection, it was found that some preliminary candidates have an observed SED that appears for the most part more compatible with the best-fit stellar template than the best-fit galaxy template (despite the fact that the preliminary star-galaxy separation cut required that the galaxy template has a lower χ^2_ν). Similarly to the analogous issue with low-redshift interlopers, this suggests the possibility of template in-

completeness or inaccurate photometry. Again, it was observed that there are some cases where a high-redshift galaxy fit might be favoured due to one or two bands with very small errors, while the other bands are more compatible with a stellar fit. In other cases, the postage stamps indicate poor or biased imaging in one or more filters. In the end, it was decided to classify preliminary candidates as a potential star if they display a PSF-like morphology in all filters, and either a) they have an observed SED that appears on the whole (more) compatible with a stellar fit or b) they show an excess of blue flux (according to the requirements laid out previously in Section 4.2.4.2). An example of a suspected star is presented in Figure 4.6. This object appears stellar in shape and shows a good fit to the stellar template in most filters, but its fit has been biased by flux from one of the rings in the extended PSF of an extremely bright nearby star.

4.2.5 Second round of cuts

After visual inspection identified the types of contamination in the preliminary sample, a series of targeted cuts was made to address each of these categories. This second selection round narrowed the sample down to a final set of secure high-redshift galaxy candidates. The following sections will describe the cuts in this final stage.

4.2.5.1 Artefacts and transients

The first cut addressed artefacts and transients. Because this type of contamination was exclusively found among $z \sim 6$ candidates (as discussed in Section 4.2.4.1), this step was only applied to these objects. All transients, as well as most artefacts, only show large fluxes in one band, and artefacts that are present in multiple bands are usually only bright in some filters. Therefore, the vast majority could be removed by requiring that any final candidate must be detected to a significance of at least 3σ in a second filter redwards of the Lyman break. Bearing in mind the intrinsic blue colours of LBGs, the optimal filter for this would be the Z -band, but because data in this band does not exist for all VIDEO tiles, Y or J detections were also permitted. According to Equation 4.5, a $> 3\sigma$ significance corresponds to a magnitude error of $\sigma_m < 0.3619$, so that the cut in question amounts to the following:

$$\sigma_Z < 0.3619 \quad \vee \quad \sigma_Y < 0.3619 \quad \vee \quad \sigma_J < 0.3619 \quad \text{for the } z \sim 6 \text{ sample. (4.7)}$$

After this step, there were 179 candidates left in the $z \sim 6$ sample.

4.2.5.2 Low-redshift galaxies and QSOs

Contamination from low-redshift galaxies and QSOs was tackled via a combination of five cuts that addressed the issues identified in Section 4.2.4.2. These cuts are listed and described below:

- As established in Section 4.2.4.2, the presence of g -band flux is a definite indicator of contamination, most likely from low-redshift galaxies (including QSOs), or possibly from stars. To tackle such contaminants, only with objects g -band flux below the local 2σ magnitude limit were retained in the sample:

$$g_{AB} < m_{\text{lim},2\sigma} \quad (4.8)$$

This cut resulted in a total of 126 and 153 remaining objects in the $z \sim 5$ and $z \sim 6$ samples respectively.

- The next step targeted contaminants whose observed photometry appears overall more consistent with a low-redshift template, despite a good high-redshift χ^2_v . When describing these objects in Section 4.2.4.2, it was suggested that inaccurate photometry is one of the causes of this issue. For many such low-redshift interlopers, the photometric redshift was found to be skewed towards a high-redshift result by small errors in one or more filters. To eliminate any bias from bands with very low errors, a second photo-z run was performed on all preliminary candidates, this time imposing a floor⁹ on the uncertainties such that any error below $\sigma_m < 0.1$ was increased to $\sigma_m = 0.1$. Any candidates that switched to a low-redshift

⁹It is interesting to note that the imposed threshold of $\sigma_m = 0.1$ is roughly equal to the maximum adaptive offset in this thesis (see Section 3.2.3), which is 0.08 in the H -band of the cdfs1 tile. Therefore, the error floor also insures the final candidate sample against any contamination that could result from these offsets.

solution with these higher errors were then eliminated by imposing the following requirement:

$$z_{\text{phot},\text{high_error}} > 4.73. \quad (4.9)$$

This new redshift threshold is 0.1 lower than the previous lower bound at $z_{\text{phot}} = 4.83$ in order to allow for a slight shift in photometric redshift. For the same reason, objects were also permitted to swap between the $z \sim 5$ or $z \sim 6$ bins. After this process, there were 90 candidates left at $z \sim 5$ and 115 at $z \sim 6$.

- Candidates were also rejected if **LePHARE** found a secondary solution at low redshift, either with the original errors or the higher errors introduced in the previous step. Like in the previous step, ‘low-redshift’ has been defined as $z \leq 4.73$ for the purpose of this cut. Since **LePHARE** assigns a secondary redshift of $z_{\text{sec}} = -99.0$ to objects without any secondary solution, the cut can be formulated as follows:

$$\begin{aligned} (z_{\text{sec}} < 0 \quad \vee \quad z_{\text{sec}} > 4.73) \quad \wedge \\ (z_{\text{sec},\text{high_error}} < 0 \quad \vee \quad z_{\text{sec},\text{high_error}} > 4.73). \end{aligned} \quad (4.10)$$

After this step¹⁰, the samples contained a total of 70 and 101 candidates at $z \sim 5$ and $z \sim 6$ respectively.

- To remove any further low-redshift galaxy stragglers, the next step placed a restriction on the likelihood of a $z \leq 2.0$ solution. To achieve this, **LePHARE** was run for a third time in order to compute the $\chi^2_{v,z \leq 2.0}$ value of the best possible $z \leq 2.0$ galaxy solution for each candidate. This third run used the higher errors introduced above, and imposed a maximum permitted redshift of $z = 2.0$ during photo-z fitting. It was required that the resulting low-redshift solutions must be a significantly worse fit than the earlier high-redshift solutions, with the ratio set to:

¹⁰It must be noted that the $z > 4.73$ parts of the criterion turned out to be inconsequential since all secondary redshifts in the preliminary sample happen to fall below $z = 2$.

$$\frac{\chi_{v,\text{high_error}}^2}{\chi_{v,\text{high_error},z \leq 2.0}^2} < 0.4. \quad (4.11)$$

The relatively restrictive threshold of 0.4 was chosen to ensure maximum purity of the final sample, so that any final candidate must have a high-redshift χ_v^2 just over twice as good as the best low-redshift alternative. In the end, imposing Equation 4.11 did not eliminate any $z \sim 5$ candidates, leaving the size of that sample at 70. The number of $z \sim 6$ candidates was cut down to 95.

- Possible contamination by low-redshift quasars was addressed by removing any candidates with an acceptable low-redshift QSO solution. An acceptable fit has been defined as $\chi_v^2 < 10.0$, the same threshold as in Equation 4.4 of the preliminary cuts. ‘Low-redshift’ has again been defined as $z \leq 4.73$. The requirements were applied to the QSO fits from both the original and higher errors, so that any candidate must satisfy:

$$(z_{\text{QSO}} > 4.73 \quad \vee \quad \chi_{v,\text{QSO}}^2 \geq 10.0) \quad \wedge \quad (z_{\text{QSO,high_error}} > 4.73 \quad \vee \quad \chi_{v,\text{QSO,high_error}}^2 \geq 10.0). \quad (4.12)$$

This cut did not eliminate any $z \sim 5$ objects, leaving 70 candidates in that sample. For the $z \sim 6$ sample, 87 objects remained.

4.2.5.3 Stars

Contamination from stars is the last major issue that had to be confronted. As explained in Section 4.2.4.3, visual inspection indicated that the SEDs for many of these potential stellar contaminants appear more compatible with a star than galaxy templates in most filters, an issue also noted among low-redshift interlopers. Similarly, inaccurate photometry was identified as a likely cause. Therefore, contamination from stars was addressed by imposing a second star-galaxy separation criterion, this time using the aforementioned higher photometric errors. Analogous to the low-redshift equivalent in Equation 4.11, this cut consists of a maximum on the ratio between the star and galaxy $\chi_{v,\text{high_error}}^2$:

$$\frac{\chi^2_{v,\text{gal,high_error}}}{\chi^2_{v,\text{star,high_error}}} < 0.4. \quad (4.13)$$

After this step, there were 39 and 54 candidates left in the $z \sim 5$ and $z \sim 6$ samples respectively.

4.2.5.4 Manual removal

Despite the specialised cuts described so far, at this point the $z \sim 6$ sample still contained 13 contaminants. Among these are 10 imaging artefacts and transient objects that accurately replicate high-redshift SEDs, for instance because they overlap with a true object, or with diffuse background flux from e.g. bright stars. The other 3 contaminants contain g -band imaging in a noisy region around the chip gaps, so that their dropout behaviour could not be reliably confirmed. All these 13 sources were removed from the sample by hand. This final step narrowed the $z \sim 6$ sample down to 41 objects.

4.2.5.5 Summary and verification of the second selection round

The second round of cuts described in the preceding sections has thus resulted in a final sample of 39 $z \sim 5$ and 41 $z \sim 6$ candidates. Table 4.2 summarises this round by listing all the selection steps that have been applied.

To investigate the effectiveness of these cuts, the author has investigated the completeness and purity of the sample via the inspection images. Regarding purity, the inspection images of all the final candidates reassuringly indicate no signs of contamination in both the $z \sim 5$ and $z \sim 6$ samples. This suggests that this thesis has succeeded in its aim of isolating a highly secure sample of high-redshift galaxies. The completeness part of the verification procedure assessed the inspection images of all the objects rejected at each step of the second round, to check whether any of them could be real high-redshift galaxies. The remainder of this section describes the results of this completeness check.

For the $z \sim 6$ sample, the selection proved highly successful. Most of the rejected objects are either artefacts, transients, or low-redshift galaxies. In fact, with the exception of the cut in Equation 4.13, none of the steps removed preliminary candidates that would not have been rejected in a subsequent step. Equation 4.13 did, due to their adequate stellar fits, remove three objects which in reality

CANDIDATE CUTS

<i>z</i> ~ 5 SAMPLE		<i>z</i> ~ 6 SAMPLE	
cut definition	remaining	cut definition	remaining
Preliminary sample	131	Preliminary sample	320
—	131	$\sigma_Z < 0.3619 \vee \sigma_Y < 0.3619 \vee \sigma_J < 0.3619$	179
$g_{AB} < m_{\text{lim},2\sigma}$	126	$g_{AB} < m_{\text{lim},2\sigma}$	153
$z_{\text{phot,high_error}} > 4.73$	90	$z_{\text{phot,high_error}} > 4.73$	115
$(z_{\text{sec}} < 0 \vee z_{\text{sec}} > 4.73) \wedge (z_{\text{sec,high_error}} < 0 \vee z_{\text{sec,high_error}} > 4.73)$	70	$(z_{\text{sec}} < 0 \vee z_{\text{sec}} > 4.73) \wedge (z_{\text{sec,high_error}} < 0 \vee z_{\text{sec,high_error}} > 4.73)$	101
$\frac{\chi^2_{v,\text{gal,high_error}}}{\chi^2_{v,\text{high_error},z \leq 2.0}} < 0.4$	70	$\frac{\chi^2_{v,\text{gal,high_error}}}{\chi^2_{v,\text{high_error},z \leq 2.0}} < 0.4$	95
$(z_{\text{QSO}} > 4.73 \vee \chi^2_{v,\text{QSO}} \geq 10.0) \wedge (z_{\text{QSO,high_error}} > 4.73 \vee \chi^2_{v,\text{QSO,high_error}} \geq 10.0)$	70	$(z_{\text{QSO}} > 4.73 \vee \chi^2_{v,\text{QSO}} \geq 10.0) \wedge (z_{\text{QSO,high_error}} > 4.73 \vee \chi^2_{v,\text{QSO,high_error}} \geq 10.0)$	87
$\frac{\chi^2_{v,\text{gal,high_error}}}{\chi^2_{v,\text{star,high_error}}} < 0.4$	39	$\frac{\chi^2_{v,\text{gal,high_error}}}{\chi^2_{v,\text{star,high_error}}} < 0.4$	54
—	39	Manual removal	41

Table 4.2 An overview of the selection criteria that have been applied to the preliminary subsample to produce a final sample of high-redshift galaxy candidates. The samples of 39 $z \sim 5$ and 41 $z \sim 6$ candidates each contain five duplicates (see Section 4.3), so there are 34 and 36 unique candidates respectively.

cannot be stars, because their postage stamps indicate extended emission¹¹. It is therefore possible that these sources may be true high-redshift galaxies. However, because their photometry is not accurate enough to indicate a strong preference for a galaxy template over a stellar SED, the fit for these candidates was considered too unreliable.

The selection for the $z \sim 5$ sample was found to be slightly less efficient than at $z \sim 6$, with some cuts removing probable real $z \sim 5$ galaxies. The photometry of these rejected preliminary candidates shows a strong 1216 Å break as well as blue continuum colours, both of which are incompatible with the gently sloping red SEDs of dusty low-redshift galaxies. In many cases the postage stamps also display extended emission, ruling out a stellar origin as well. It is postulated that this unnecessary rejection, which did not occur in the $z \sim 6$ sample, is partly caused by the minimum $\sigma_m > 0.1$ error floor (introduced in Section 4.2.5.2 and used in Equations 4.9, 4.10, 4.11, 4.12, and 4.13). Because the magnitude uncertainty σ_m scales proportionally with the flux uncertainty σ_F and inversely with the flux F (see Equation 4.5), this error floor only affects filters with sufficiently deep data and adequately bright emission. Galaxies at $z \sim 5$ are on average brighter than their $z \sim 6$ counterparts in all filters, and they are also bright in more filters, since they emit (rest-frame UV) continuum flux in both the i - and z -band rather than the z -band alone. Emission from the i -band is comparatively more affected by the error floor, because this filter is deeper than the z -band and VIDEO bands. On the whole, all these factors entail that the imposition of minimum errors has a larger effect on the $z \sim 5$ photometry. This effect could lead to the rejection of real candidates, because the higher errors produce a lower χ^2_v for all templates, increasing the possibility of a low-redshift secondary solution, as well as the likelihood of a low-redshift primary solution (due to the prior vastly favouring low-redshift fits). In addition to the error floor, another potential cause for the removal of probable $z \sim 5$ candidates is that the second round of cuts has been made conservative by design. Strong cuts have been chosen in order to reduce contamination, which could come at the cost of rejecting some true candidates. This issue affects the $z \sim 5$ sample more strongly, because $z \sim 5$ galaxies are more compatible with a larger range of star and low-redshift galaxy templates, due to dropout behaviour being displayed in one fewer filter. As a result, $z \sim 5$ fits may

¹¹These objects also have high values of **spread_model**, exceeding the galaxy threshold of $\text{spread_model} > 0.0028$ posited by Banerji et al. (2015) for another DES dataset.

be less likely to pass the conservative second round cuts. One could argue that the use of the higher errors is therefore not appropriate at $z \sim 5$. However, it was found that several stellar and low-redshift contaminants enter the final sample when the error floor is removed. It was therefore decided to retain the minimum errors, in order to ensure a maximally secure sample. An additional advantage of this decision is that it also enables a more direct comparison with the $z \sim 6$ candidates. Nevertheless, it would be interesting to investigate whether it would be possible to formulate alternative selection criteria for the $z \sim 5$ sample that reject fewer good candidates. This task is left as a suggestion for future work.

4.3 Final sample

Now the process for isolating a secure collection of high-redshift candidates has been discussed, the current section focuses on these final candidates in further detail. Table 4.3 presents the 39 objects in the $z \sim 5$ sample, listing their position, photometric redshift and magnitudes in the r , i , z and J filters. Table 4.4 lists the corresponding results for the 41 $z \sim 6$ candidates. It is worth noting that these samples include duplicates. The total number of unique galaxies is 34 at $z \sim 5$, and 36 at $z \sim 6$, as both samples contain 5 pairs of duplicates (indicated in the aforementioned tables). The existence of such duplicates is reassuring, because it suggests that the selection method retrieves similar results for flux measurements in tiles with slightly different depths.

False-colour images for the candidates are presented in Figures 4.7 and 4.8 for the $z \sim 5$ and $z \sim 6$ samples respectively. These images have been generated using a custom code created by the author, described in detail in Appendix B. It is reassuring to note that several candidates very clearly display extended emission, definitively confirming their status as galaxies. Some of these galaxies even show signs of irregular, ellipsoid, or disk-like shapes. Furthermore, many more (around half the objects in both samples) show a morphology with at least some indications of an extended shape. This is consistent with results from Bowler et al. (2015) and Willott et al. (2013), who found that over half, but not all, bright $z \sim 6$ galaxies are resolved in ground-based imaging (with a slightly lower seeing of 0.7–0.8 arcsec).

$z \sim 5$ CANDIDATE SAMPLE

ID	RA and Dec (J2000)	z_{phot}	r_{AB}	i_{AB}	z_{AB}	J_{AB}	Notes
DES0261588	00:40:13.35-44:00:37.26	5.03	$25.15^{\pm 0.18}$	$23.41^{\pm 0.04}$	$22.99^{\pm 0.04}$	$23.25^{\pm 0.12}$	a
DES0334028	00:40:13.35-44:00:37.26	4.98	$25.15^{\pm 0.18}$	$23.41^{\pm 0.04}$	$22.99^{\pm 0.04}$	$23.20^{\pm 0.11}$	a
DES0773209	02:22:28.00-04:51:37.10	4.95	> 28.24	$25.00^{\pm 0.06}$	$24.67^{\pm 0.05}$	$24.94^{\pm 0.30}$	
DES0892423	02:24:43.46-04:33:16.77	4.98	$26.97^{\pm 0.33}$	$24.89^{\pm 0.05}$	$24.61^{\pm 0.05}$	$24.84^{\pm 0.27}$	
DES0924984	02:25:10.25-04:12:22.94	4.97	$26.19^{\pm 0.18}$	$24.57^{\pm 0.04}$	$24.29^{\pm 0.04}$	$24.63^{\pm 0.22}$	
DES0927703	02:25:17.30-04:10:47.95	4.88	$26.32^{\pm 0.20}$	$24.73^{\pm 0.04}$	$24.57^{\pm 0.05}$	$24.59^{\pm 0.22}$	
DES1008140	02:23:39.65-04:38:34.73	4.96	$26.53^{\pm 0.23}$	$24.68^{\pm 0.05}$	$24.46^{\pm 0.04}$	$24.58^{\pm 0.22}$	
DES1027267	02:25:31.35-05:10:25.29	4.90	$27.06^{\pm 0.36}$	$24.87^{\pm 0.05}$	$24.83^{\pm 0.06}$	$25.25^{\pm 0.40}$	
DES1053297	02:25:18.60-04:57:41.88	4.90	$27.28^{\pm 0.45}$	$24.87^{\pm 0.06}$	$24.61^{\pm 0.05}$	$24.68^{\pm 0.23}$	
DES1069344	02:24:26.12-04:49:48.35	5.04	$26.33^{\pm 0.19}$	$24.92^{\pm 0.06}$	$24.53^{\pm 0.05}$	$24.73^{\pm 0.25}$	
DES1099438	02:24:23.91-05:28:06.61	4.90	$26.03^{\pm 0.16}$	$24.33^{\pm 0.03}$	$24.09^{\pm 0.03}$	$24.11^{\pm 0.14}$	
DES1158545	02:27:55.10-04:23:57.81	4.89	$27.12^{\pm 0.40}$	$24.82^{\pm 0.05}$	$24.72^{\pm 0.06}$	$24.56^{\pm 0.21}$	
DES1230744	02:28:01.00-03:59:37.89	4.88	$26.26^{\pm 0.15}$	$24.49^{\pm 0.03}$	$24.36^{\pm 0.03}$	$24.18^{\pm 0.15}$	
DES1240477	02:28:02.95-05:17:21.98	4.97	$27.07^{\pm 0.37}$	$24.86^{\pm 0.05}$	$24.56^{\pm 0.05}$	$24.74^{\pm 0.25}$	
DES1248943	02:27:26.10-05:13:47.74	4.91	$27.11^{\pm 0.38}$	$24.65^{\pm 0.04}$	$24.45^{\pm 0.04}$	$24.38^{\pm 0.18}$	
DES1266621	02:28:40.91-05:08:12.87	5.01	$25.92^{\pm 0.13}$	$24.44^{\pm 0.04}$	$24.09^{\pm 0.03}$	$24.48^{\pm 0.19}$	
DES1391206	03:26:57.50-27:47:26.78	4.89	$26.70^{\pm 0.21}$	$24.49^{\pm 0.04}$	$24.58^{\pm 0.04}$	$24.62^{\pm 0.23}$	
DES1429846	03:27:26.41-28:11:04.15	5.03	$26.28^{\pm 0.14}$	$24.59^{\pm 0.04}$	$24.23^{\pm 0.03}$	$24.36^{\pm 0.26}$	b
DES1453274	03:29:25.84-27:19:34.34	4.89	$26.10^{\pm 0.12}$	$24.47^{\pm 0.04}$	$24.30^{\pm 0.03}$	$24.43^{\pm 0.19}$	
DES1480785	03:30:09.41-28:10:13.10	4.86	$25.40^{\pm 0.06}$	$23.93^{\pm 0.02}$	$23.75^{\pm 0.02}$	$23.75^{\pm 0.10}$	c
DES1506707	03:29:09.48-28:01:35.17	5.29	$26.75^{\pm 0.22}$	$24.81^{\pm 0.05}$	$24.17^{\pm 0.03}$	$24.20^{\pm 0.16}$	
DES1510338	03:30:21.63-28:00:28.02	4.97	$26.36^{\pm 0.15}$	$24.66^{\pm 0.04}$	$24.39^{\pm 0.03}$	$24.65^{\pm 0.24}$	
DES1513918	03:30:29.44-27:59:21.69	4.91	$26.87^{\pm 0.25}$	$24.96^{\pm 0.05}$	$24.82^{\pm 0.05}$	$25.19^{\pm 0.39}$	d
DES1516307	03:28:44.72-27:58:41.09	5.29	$27.15^{\pm 0.32}$	$24.98^{\pm 0.06}$	$24.38^{\pm 0.03}$	$24.68^{\pm 0.24}$	
DES1613457	03:27:26.41-28:11:04.15	5.02	$26.28^{\pm 0.14}$	$24.58^{\pm 0.04}$	$24.22^{\pm 0.03}$	$24.31^{\pm 0.25}$	b
DES1615885	03:30:09.41-28:10:13.10	4.84	$25.40^{\pm 0.06}$	$23.93^{\pm 0.02}$	$23.75^{\pm 0.02}$	$23.74^{\pm 0.15}$	c
DES1683336	03:29:18.72-28:40:21.16	4.88	$26.76^{\pm 0.22}$	$24.77^{\pm 0.04}$	$24.50^{\pm 0.04}$	$24.59^{\pm 0.32}$	
DES1798803	03:30:27.12-27:19:41.66	4.88	$26.71^{\pm 0.22}$	$24.91^{\pm 0.05}$	$24.69^{\pm 0.05}$	$24.61^{\pm 0.23}$	

continued

$z \sim 5$ CANDIDATE SAMPLE (CONTINUED)

ID	RA and Dec (J2000)	z_{phot}	r_{AB}	i_{AB}	z_{AB}	J_{AB}	Notes
DES1802305	03:32:45.06-27:18:38.68	4.89	$26.82^{\pm 0.24}$	$24.95^{\pm 0.06}$	$24.78^{\pm 0.05}$	$24.88^{\pm 0.29}$	
DES1870784	03:30:54.31-28:00:20.08	4.83	$25.88^{\pm 0.10}$	$24.34^{\pm 0.03}$	$24.34^{\pm 0.03}$	$24.34^{\pm 0.18}$	
DES1873583	03:30:29.44-27:59:21.77	4.91	$26.89^{\pm 0.26}$	$24.96^{\pm 0.05}$	$24.84^{\pm 0.05}$	$25.07^{\pm 0.35}$	d
DES1929684	03:30:54.69-27:39:17.44	4.96	$26.67^{\pm 0.21}$	$24.79^{\pm 0.05}$	$24.54^{\pm 0.04}$	$24.50^{\pm 0.21}$	
DES1941550	03:32:54.04-27:34:44.68	4.97	$26.72^{\pm 0.21}$	$24.70^{\pm 0.04}$	$24.47^{\pm 0.04}$	$24.41^{\pm 0.19}$	
DES2022830	03:32:07.78-28:53:22.34	4.84	$25.65^{\pm 0.08}$	$24.21^{\pm 0.03}$	$23.96^{\pm 0.02}$	$24.15^{\pm 0.21}$	e
DES2023522	03:32:41.96-28:53:06.73	4.96	$26.58^{\pm 0.19}$	$24.49^{\pm 0.04}$	$24.11^{\pm 0.03}$	$24.05^{\pm 0.19}$	
DES2048158	03:31:36.51-28:43:14.73	5.28	$26.73^{\pm 0.21}$	$24.91^{\pm 0.05}$	$24.27^{\pm 0.03}$	$24.04^{\pm 0.19}$	
DES2122745	03:31:45.95-28:12:02.69	4.95	$26.99^{\pm 0.24}$	$24.83^{\pm 0.07}$	$24.75^{\pm 0.05}$	$25.08^{\pm 0.50}$	
DES2177455	03:31:32.36-29:01:48.78	4.96	$26.94^{\pm 0.28}$	$24.45^{\pm 0.03}$	$24.08^{\pm 0.02}$	$24.05^{\pm 0.19}$	
DES2191223	03:32:07.78-28:53:22.21	4.83	$25.63^{\pm 0.08}$	$24.20^{\pm 0.03}$	$23.95^{\pm 0.02}$	$24.14^{\pm 0.21}$	e

Notes:

The sample contains five pairs of duplicates, originating from the tile overlap listed below. The parentheses after each tile enclose the catalogue ID of the duplicate object in that tile.

^a Duplicates from the es1-north (DES0261588) and es1-south (DES0334028) VIDEO tiles.

^b Duplicates from the DES0236-2832 (DES1429846) and DES0328-2749 (DES1613457) DES tiles.

^c Duplicates from the cdfs1 (DES1480785) and cdfs2 (DES1615885) VIDEO tiles.

^d Duplicates from the DES0328-2749 (DES1513918) and DES0332-2749 (DES1873583) DES tiles.

^e Duplicates from the DES0332-2832 (DES2022830) and DES0333-2915 (DES2191223) DES tiles.

Table 4.3 The final sample of $z \sim 5$ candidates, with for each candidate its catalogue ID, position, photometric redshift z_{phot} , and r -, i -, z -, and J -band magnitude. The sample contains five pairs of duplicates, which are flagged in the notes. The value of z_{phot} is the photometric redshift computed with the original uncertainties (i.e. without the error floor imposed in Section 4.2.5.2). Magnitudes are given for a 1.95 arcsec diameter aperture. If the magnitude of a candidate falls below the 1σ depth in that area, the 1σ limiting magnitude is shown instead.

$z \sim 6$ CANDIDATE SAMPLE

ID	RA and Dec (J2000)	z_{phot}	r_{AB}	i_{AB}	z_{AB}	J_{AB}	Notes
DES0027723	00:34:29.87-43:43:19.09	5.93	>27.62	$26.86^{\pm 0.70}$	$24.69^{\pm 0.11}$	$25.21^{\pm 0.72}$	
DES0688180	02:22:28.32-04:33:08.94	5.88	>28.22	$26.49^{\pm 0.25}$	$24.76^{\pm 0.06}$	$24.79^{\pm 0.26}$	a
DES0825517	02:23:00.37-04:33:25.15	6.09	>28.29	>28.14	$24.84^{\pm 0.06}$	$24.76^{\pm 0.26}$	
DES0826107	02:22:28.33-04:33:08.91	5.88	>28.22	$26.41^{\pm 0.23}$	$24.75^{\pm 0.06}$	$24.77^{\pm 0.26}$	a
DES0837948	02:23:00.03-04:27:56.15	5.99	>28.29	$27.69^{\pm 0.72}$	$25.00^{\pm 0.07}$	$24.95^{\pm 0.30}$	
DES0853065	02:24:15.10-04:20:47.05	6.08	>28.25	>28.14	$24.96^{\pm 0.07}$	$25.61^{\pm 0.56}$	b, f
DES0911243	02:24:15.10-04:20:47.05	6.07	>28.25	>28.14	$24.96^{\pm 0.07}$	$25.02^{\pm 0.32}$	b, f
DES0938949	02:24:51.15-04:03:29.42	6.07	>28.23	>28.12	$24.78^{\pm 0.06}$	$24.83^{\pm 0.27}$	g, i
DES0975555	02:23:52.26-04:59:27.69	5.90	$27.94^{\pm 0.82}$	$27.15^{\pm 0.46}$	$24.85^{\pm 0.06}$	$24.75^{\pm 0.25}$	c
DES1020121	02:25:43.56-05:16:02.62	6.18	>28.22	>28.08	$24.98^{\pm 0.09}$	$24.93^{\pm 0.30}$	
DES1026848	02:25:15.61-05:10:36.95	5.56	$27.72^{\pm 0.66}$	$25.65^{\pm 0.11}$	$24.66^{\pm 0.05}$	$24.55^{\pm 0.21}$	
DES1049468	02:23:52.26-04:59:27.69	5.92	$27.94^{\pm 0.82}$	$27.15^{\pm 0.46}$	$24.85^{\pm 0.06}$	$25.23^{\pm 0.39}$	c
DES1170574	02:26:55.55-04:20:05.77	5.85	>28.29	$26.70^{\pm 0.28}$	$24.96^{\pm 0.07}$	$25.39^{\pm 0.45}$	
DES1182031	02:27:57.37-04:15:44.16	6.04	$26.74^{\pm 0.34}$	>28.12	$24.78^{\pm 0.08}$	$25.37^{\pm 0.44}$	
DES1251416	02:26:50.96-05:13:01.16	5.99	>28.29	$27.31^{\pm 0.51}$	$24.71^{\pm 0.06}$	$24.67^{\pm 0.23}$	
DES1272155	02:26:12.78-05:06:28.34	5.82	>28.23	$27.35^{\pm 0.53}$	$24.74^{\pm 0.06}$	$23.59^{\pm 0.09}$	
DES1299781	02:27:15.34-04:56:59.94	5.71	$26.89^{\pm 0.30}$	$25.39^{\pm 0.09}$	$24.04^{\pm 0.03}$	$24.14^{\pm 0.14}$	
DES1310137	02:26:27.02-04:52:38.14	6.15	>28.29	>28.17	$24.54^{\pm 0.05}$	$24.55^{\pm 0.21}$	h
DES1361409	02:28:20.94-04:27:48.13	5.68	>28.24	$26.22^{\pm 0.19}$	$24.79^{\pm 0.06}$	$24.74^{\pm 0.25}$	i
DES1362410	02:28:23.81-04:25:23.24	5.79	>28.18	$25.43^{\pm 0.13}$	$23.98^{\pm 0.03}$	$23.96^{\pm 0.12}$	i
DES1380979	03:27:00.73-28:05:17.32	6.17	>27.56	>28.00	$24.94^{\pm 0.06}$	$25.19^{\pm 0.39}$	
DES1431780	03:30:06.95-27:28:01.38	5.60	$28.14^{\pm 0.83}$	$25.57^{\pm 0.10}$	$24.42^{\pm 0.04}$	$24.90^{\pm 0.30}$	d
DES1436416	03:28:34.65-27:25:50.74	5.64	$27.51^{\pm 0.47}$	$25.71^{\pm 0.12}$	$24.44^{\pm 0.04}$	$24.33^{\pm 0.18}$	
DES1491906	03:28:51.96-28:06:49.50	5.74	>28.54	$26.42^{\pm 0.21}$	$24.98^{\pm 0.06}$	$25.40^{\pm 0.47}$	
DES1494058	03:28:37.42-28:05:31.32	5.87	>28.45	$26.67^{\pm 0.35}$	$24.67^{\pm 0.06}$	$24.51^{\pm 0.21}$	e
DES1499635	03:27:30.56-28:03:47.75	5.58	$28.53^{\pm 1.09}$	$25.68^{\pm 0.10}$	$24.59^{\pm 0.04}$	$24.58^{\pm 0.22}$	
DES1510928	03:29:56.17-28:00:18.36	5.82	$27.72^{\pm 0.55}$	$26.34^{\pm 0.19}$	$24.63^{\pm 0.04}$	$24.48^{\pm 0.20}$	
DES1535579	03:27:40.63-27:52:13.53	6.03	>28.52	$27.71^{\pm 0.68}$	$24.90^{\pm 0.05}$	$24.88^{\pm 0.29}$	

continued

$z \sim 6$ CANDIDATE SAMPLE (CONTINUED)

ID	RA and Dec (J2000)	z_{phot}	r_{AB}	i_{AB}	z_{AB}	J_{AB}	Notes
DES1578557	03:28:07.48-27:38:33.71	5.78	> 28.49	$26.34^{\pm 0.19}$	$24.79^{\pm 0.05}$	$24.69^{\pm 0.25}$	
DES1591044	03:30:05.75-27:34:17.97	5.58	> 28.47	$25.95^{\pm 0.14}$	$24.82^{\pm 0.05}$	$24.93^{\pm 0.31}$	
DES1608291	03:28:16.90-27:28:56.88	5.93	> 28.50	$25.22^{\pm 0.07}$	$23.14^{\pm 0.01}$	$22.94^{\pm 0.05}$	
DES1611505	03:30:06.95-27:28:01.33	5.59	$28.07^{\pm 0.78}$	$25.55^{\pm 0.10}$	$24.42^{\pm 0.04}$	$24.89^{\pm 0.30}$	d
DES1627764	03:28:37.42-28:05:31.32	5.85	> 28.45	$26.67^{\pm 0.35}$	$24.67^{\pm 0.06}$	$24.11^{\pm 0.21}$	e
DES1649871	03:30:33.69-28:51:22.66	5.97	> 27.44	$26.61^{\pm 0.37}$	$24.26^{\pm 0.05}$	$24.42^{\pm 0.27}$	
DES1650997	03:30:32.24-28:51:01.41	5.87	> 28.56	$26.73^{\pm 0.28}$	$24.71^{\pm 0.05}$	$24.13^{\pm 0.21}$	
DES1666316	03:29:08.46-28:46:17.75	5.83	> 28.51	$25.48^{\pm 0.13}$	$24.05^{\pm 0.04}$	$23.89^{\pm 0.17}$	
DES1870647	03:32:45.87-28:00:23.37	5.52	$27.84^{\pm 0.60}$	$25.35^{\pm 0.08}$	$24.09^{\pm 0.03}$	$23.42^{\pm 0.08}$	
DES1873886	03:30:34.07-27:59:15.40	5.98	> 28.50	$27.35^{\pm 0.49}$	$24.96^{\pm 0.06}$	$24.85^{\pm 0.29}$	
DES1875662	03:33:09.77-27:58:37.63	5.90	> 28.49	$27.05^{\pm 0.37}$	$24.84^{\pm 0.05}$	$25.19^{\pm 0.39}$	
DES2015850	03:33:33.85-27:33:03.83	5.57	> 28.56	$25.44^{\pm 0.08}$	$24.35^{\pm 0.03}$	$24.24^{\pm 0.25}$	
DES2131269	03:33:00.13-28:39:38.94	5.89	> 28.59	$27.58^{\pm 0.56}$	$24.84^{\pm 0.05}$	$23.85^{\pm 0.18}$	

Notes:

The sample contains five pairs of duplicates, originating from the tile overlap listed below. The parentheses after each tile enclose the catalogue ID of the duplicate object in that tile.

^a Duplicates from the DES0221-0416 (DES0688180) and DES0223-0416 (DES0826107) DES tiles.

^b Duplicates from the xmm2 (DES0853065) and xmm3 (DES0911243) VIDEO tiles.

^c Duplicates from the xmm2 (DES0975555) and xmm3 (DES1049468) VIDEO tiles.

^d Duplicates from the DES0328-2706 (DES1431780) and DES0328-2749 (DES1611505) DES tiles.

^e Duplicates from the cdfs1 (DES1494058) and cdfs2 (DES1627764) tiles.

Three candidates correspond to the following objects from a sample by Willott et al. (2013):

^f WMH2

^g WMH3

^h WMH5

Spectroscopy:

ⁱ Candidate selected for spectroscopic follow-up.

Table 4.4 The final sample of $z \sim 6$ candidates, with for each candidate its catalogue ID, position, photometric redshift z_{phot} , and r -, i -, z -, and J -band magnitude. The sample contains five pairs of duplicates, which are flagged in the notes. Objects that were found in a $z \approx 6$ study by Willott et al. (2013) are also flagged, as well as the three candidates that were submitted for spectroscopic follow-up. Further details on z_{phot} and the magnitude measurements are identical to those in Table 4.3.

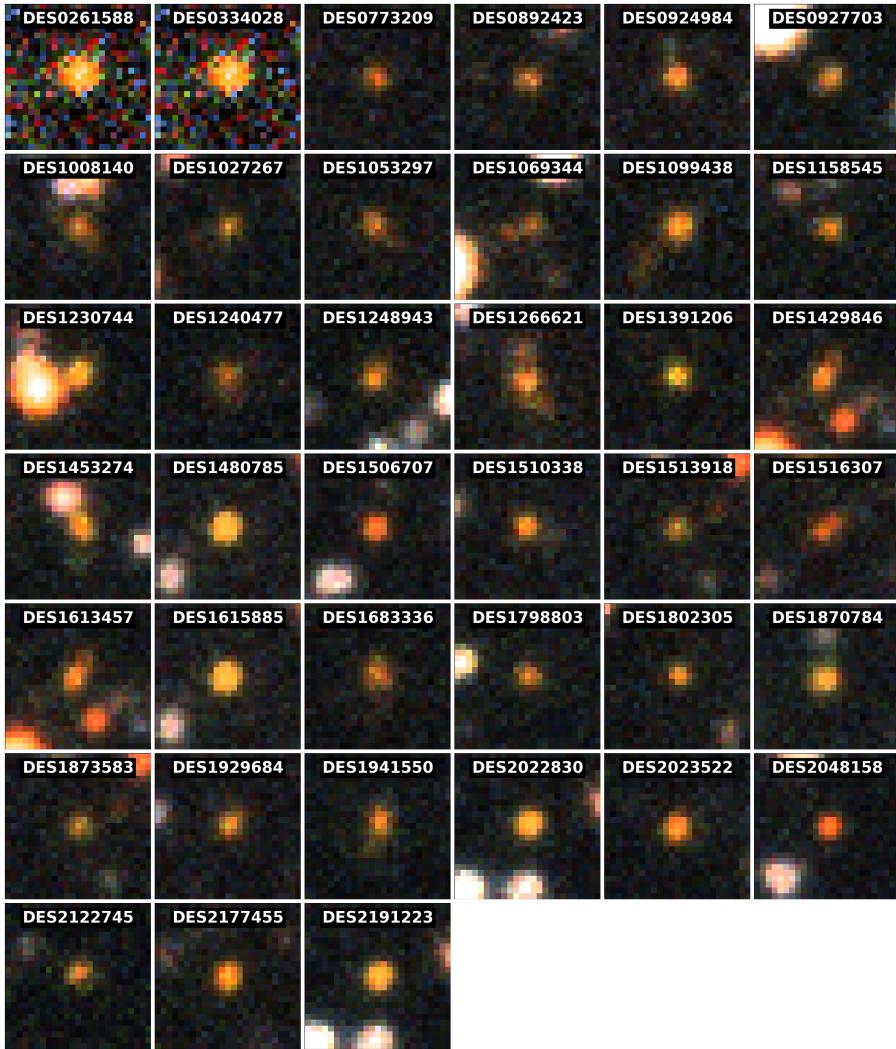


Figure 4.7 False-colour images of the 39 candidates in the final $z \sim 5$ sample. These have been generated from cutouts of the DES g -, r -, i -, and z -band data using a custom code created by the author, described in Appendix B. Each image covers an area of 6.6 arcsec \times 6.6 arcsec. It is clear that several pairs of candidates have identical false-colour images; those are duplicates originating from DES tile overlap. It is also interesting to observe that the morphologies of many candidates show signs of extended emission.

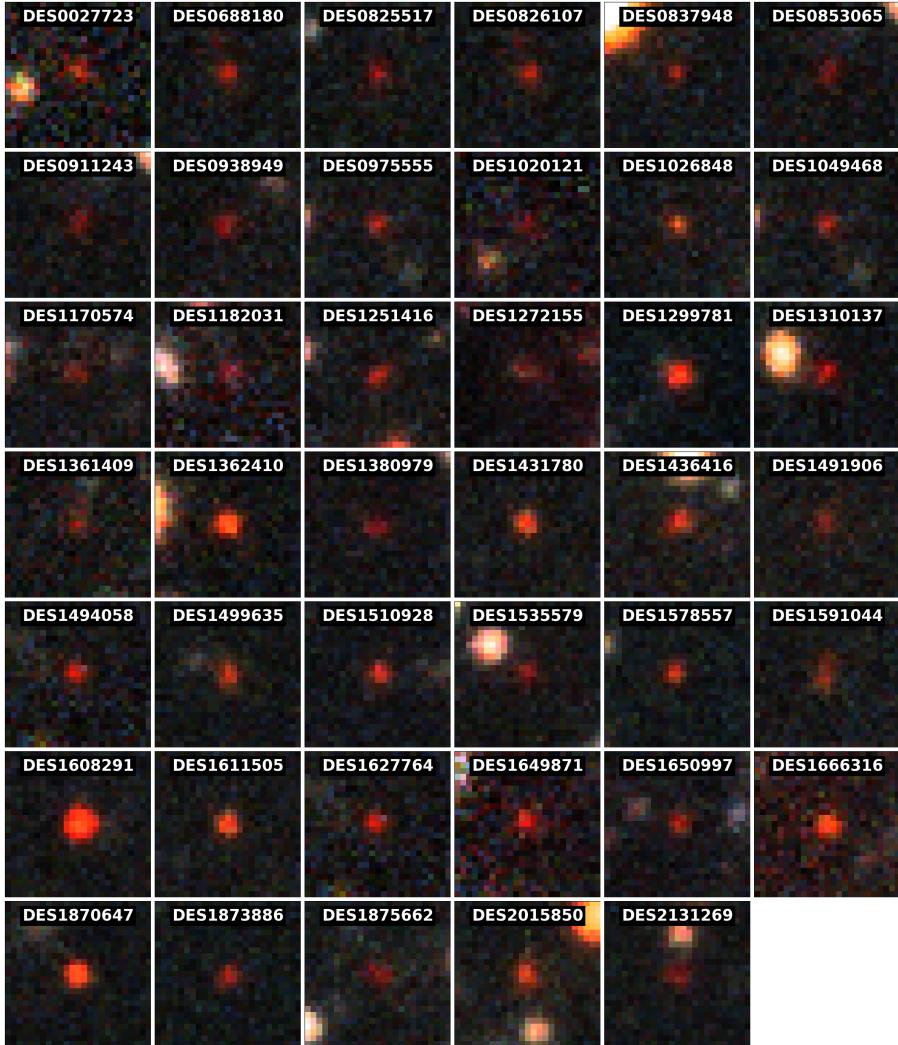


Figure 4.8 False-colour images of the 41 final $z \sim 6$ candidates, obtained in the same way as for the $z \sim 5$ sample in Figure 4.7. The colour calibration and scale are identical, with each cutout covering $6.6 \text{ arcsec} \times 6.6 \text{ arcsec}$. Similarly to Figure 4.7, some pairs of candidates have identical false-colour images due to the presence of duplicates. Several objects show morphologies indicative of extended emission. It is reassuring to note that the $z \sim 6$ candidates generally appear fainter and smaller than their $z \sim 5$ counterparts, as would be expected from their greater distances.

4.3.1 Comparison with previous studies

4.3.1.1 Number of expected candidates

It is interesting to estimate the expected number of DES+VIDEO high-redshift candidates based on other research in the literature. In recent years, several authors have found samples of bright $z \gtrsim 5$ galaxies. For instance, McLure et al. (2009) found 1621 possible¹² galaxies with magnitudes of $z'_{\text{AB}} < 26.0$ and redshifts of $z > 4.5$ over a search area of 0.63 deg^2 with 5σ depths of $z'_{\text{AB}} \approx 26.2$. In a comparable study restricted only to $z \sim 6$, Bowler et al. (2015) searched a deep 1.36 deg^2 region with 5σ depths of $z'_{\text{AB}} = 26.5\text{--}26.8$, finding 263 candidates with magnitudes of $z'_{\text{AB}} < 26.0$. These two studies suggest that there may be several hundreds, possibly thousands, of high-redshift $z_{\text{AB}} < 26.0$ objects in the deepest 4 deg^2 of DES+VIDEO data. However, the search in this thesis is limited to the brightest galaxies with magnitudes of $m_{\text{AB}} < 25.0$, and these luminous sources have been found to be much rarer. Comparable studies in this regime have mainly been conducted at $z \sim 6$. The study by Bowler et al. (2015) found nine bright $z' < 25.0$ candidates in their 1.65 deg^2 search area (situated within the COSMOS and SDXS fields). These candidates, selected via photometric redshifts, span a redshift range of $5.7 < z < 6.3$. For a 0.29 deg^2 sub-region¹³ the authors also reported three $z' < 25.0$ galaxies when they imposed $5.5 < z < 6.5$ limits identical to those used in this thesis. Willott et al. (2013) conducted another study that found 14 $z \approx 6$ candidate Lyman-break galaxies with magnitudes of $z' < 25.0$ in the four CFHTLS Deep Fields. Their search covered an area of 4 deg^2 with a 5σ depth of $z' \approx 26.5$, and it made use of colour-colour selection criteria, which were most sensitive to galaxies in the range $5.7 < z < 6.0$.

The samples from Bowler et al. (2015) and Willott et al. (2013) can be used to extrapolate the rough number of expected DES+VIDEO $z \sim 6$ galaxies. For this it is necessary to know the effective search area, which was measured to be 4.9 deg^2 as previously described in Section 4.2.3.5. Using this quantity, the numbers from

¹²It must be remarked that this sample was probably not entirely pure. McLure et al. (2009) did not exclude objects with a possibility of a secondary solution at low redshift, instead choosing to weight their objects by the redshift probability density function $P(z)$ when computing their luminosity function.

¹³Bowler et al. (2015) refer to this area as the ULTRAVISTA/COSMOS DR1 region. This region has shallower near-IR data than the other 1.36 deg^2 of their footprint, but has identical $z'_{\text{AB}} = 26.5\text{--}26.8$ limiting magnitudes.

previous studies were then simply scaled by area to obtain the estimated number of DES+VIDEO catalogues. Those calculations produced an expected total of 27 $z \sim 6$ candidates based on the Bowler et al. (2015) study, and a lower estimate of 17 from Willott et al. (2013).

Clearly, both these values are smaller than the actual DES+VIDEO total of 36 unique candidates. There are several possible explanations for this. Firstly, the Willott et al. (2013) estimate of 17 is likely too low, as Bowler et al. (2015) demonstrated that the Willott et al. (2013) selection method excluded many genuine LBGs (possibly by as much as a factor of 3–6). Additionally, both studies restricted their final estimates to a narrower redshift range than DES+VIDEO, so their projected number counts are inherently expected to be lower than those in this thesis. To emphasise this point, if one considers only the 0.29 deg^2 region where Bowler et al. (2015) quoted three objects within the DES+VIDEO range of $5.5 < z < 6.5$, the number of estimated candidates rises to 51 — a little more than were found in this thesis. A further factor is that the number counts and search area in the Bowler et al. (2015) sample are small, which makes any extrapolations from their numbers strongly susceptible to random chance, both from selection effects¹⁴ and from variations in the underlying number of actual bright $z \sim 6$ galaxies in the fields. Finally, another possible factor may be cosmic variance, although this is *prima facie* unlikely given the fields in question. The studies by Willott et al. (2013) and Bowler et al. (2015) both include the COSMOS field, which according to either study contains an overdensity of galaxies at $z \sim 6$. The DES+VIDEO sample, on the other hand, contains a roughly equal number of objects in the equally-sized¹⁵ XMM-LSS and CDF-S fields. In the former region, which is also within the Willott et al. (2013) search area, Willott et al. (2013) found 8 candidates — similar to the number of objects in their other three fields apart from COSMOS. This implies that none of the DES+VIDEO footprint is likely to be particularly overdense. It is nevertheless possible that cosmic variance did contribute to why the Bowler et al. (2015) estimate falls below DES+VIDEO. Their paper raised the possibility that the SDXS field (which supplied 45% of their footprint) may be underpopulated around $z \approx 5.7$, which could cause

¹⁴It is certainly possible that the photo-zs computations or star-galaxy separation by Bowler et al. (2015) happened to exclude a few real sources, or that uncertainties in the photometry caused some of their magnitude measurement to be just over $z' = 25.0$.

¹⁵The effective area in both these fields is approximately equal, and largely corresponds to the deep regions in Figures 2.5 and 2.6.

their estimate to be biased slightly low. All things considered, all the mitigating factors in this paragraph suggest that there are several highly plausible reasons why the candidate estimates from previous studies have come out low, which means that these comparisons provide no reason to believe that the DES+VIDEO sample is substantially too large. Consequently, there is no compelling evidence for significant contamination within the sample presented in this thesis.

4.3.1.2 Comparison with Willott et al. (2013)

Conveniently, as mentioned previously, the imaging used by Willott et al. (2013) partially overlaps with the DES+VIDEO footprint in the XMM-LSS field. It is therefore interesting to compare their candidates to the $z \sim 6$ sample in this thesis. Their selection method used both a different dataset (optical imaging from the CFHTLS-Deep and near-IR data from an earlier VIDEO release) and a different selection method (based on colour-colour cuts instead of photometric redshifts). Even though most of the candidates in their sample are not spectroscopically confirmed, the Willott et al. (2013) objects can therefore be used as an independent means to verify the candidates in this thesis.

Out of the 40 $z \sim 6$ Lyman-break candidates found by Willott et al. (2013), eight are located within the DES+VIDEO footprint. Of these eight objects, six unique sources exist in the DES+VIDEO catalogue (WMH2, WMH3, WMH4, WMH5, WMH6, WMH8). The two others, WMH1 and WMH7, are too faint in the $r + i + z$ detection image to be included in the DESDM catalogues, so they are not present in the DES+VIDEO tables. WMH2 is found twice in the DES+VIDEO data, as it is a duplicate that was observed in both the xmm2 and xmm3 VIDEO tiles.

Of the six unique common candidates, only three (WMH2, WMH3 and WMH5) are brighter than the $z_{AB} < 25.0$ magnitude limit imposed in Equation 4.6. It is reassuring to note that all three of these bright objects appear in the final DES+VIDEO candidate sample. Especially encouraging is the fact that this includes WMH5, which was spectroscopically confirmed to be a high-redshift galaxy at $z_{\text{spec}} = 6.068$. Moreover, both duplicate instances of WMH2 passed the selection criteria, confirming that the VIDEO photometry in both the xmm2 and xmm3 tiles is sufficient in quality to ascertain its high-redshift status. The DES+VIDEO candidates that correspond to a Willott et al. (2013) object are flagged in Table 4.4.

The remaining three objects (WMH4, WMH6 and WMH8) were not included in the DES+VIDEO sample due to their faint magnitudes. Besides, WMH6 was also assigned a low DES+VIDEO redshift of $z_{\text{phot}} = 0.8$, despite it being spectroscopically confirmed at $z_{\text{spec}} = 5.645$. This suggests that the selection method in this thesis is indeed less effective in the faint $z_{\text{AB}} > 25.0$ regime, and that it is appropriate to impose this magnitude limit. Nevertheless, WMH4 and WMH8 did get assigned DES+VIDEO photometric redshifts of $z_{\text{phot}} = 6.1$ and $z_{\text{phot}} = 6.0$ respectively, suggesting that the photo-z method in this thesis does retrieve accurate results for some faint sources at (presumed) high redshifts. Furthermore, these two candidates also pass all the other selection criteria in both rounds of Section 4.2, which means that they would have made it into the final sample were it not for the magnitude limit.

The conclusion of the comparison in this section is highly encouraging – the fact that all three bright Willott et al. (2013) galaxies (including the spectroscopically confirmed WMH5) were retrieved as high-redshift candidates provides additional confidence in the selection method used in this thesis. Furthermore, the good performance for 2 out of 3 of the fainter objects is also encouraging. Thirdly, the two DES+VIDEO candidates that correspond to the WMH2 and WMH3 candidates are very likely to be true high-redshift galaxies, as they have independently been found via both the Willott et al. (2013) colour-cuts and the photometric redshift method used in this thesis.

4.4 Attempted spectroscopic follow-up

Despite the good level of confidence in the search strategy employed in this thesis, spectroscopic redshifts constitute the only way of confirming the high-redshift status of candidates with certainty. Therefore, the author attempted to obtain spectra for some of the most promising DES+VIDEO candidates. Observing time for this spectroscopic follow-up was kindly secured on the Large Binocular Telescope (LBT) via Paul Martini, who volunteered to observe two candidates on the Multi Object Double Spectrographs (MODS; Pogge et al. 2010). This instrument also facilitated the observation of a third $z \sim 6$ candidate close to one of the original candidates, thanks to its large field of view and capacity to observe multiple objects at once. All in all, candidates DES0938949 (WMH3; $z_{\text{phot}} = 6.07$), and DES1361409 + DES1362410 (observed simultaneously; $z_{\text{phot}} =$

5.68 and $z_{\text{phot}} = 5.79$ respectively) were submitted for spectroscopic follow-up and observed in October 2016. Unfortunately, the observations yielded no useful spectra, most likely due to bad weather conditions. Further spectroscopic follow-up is therefore highly recommended as an objective for future research.

Chapter 5

Summary and conclusion

A core part of the research presented in this thesis has concentrated around applications of multiwavelength survey datasets. As we saw in Chapter 1, modern advancements in imaging technology have created highly sophisticated instrumentation capable of conducting surveys that efficiently capture the night sky in digital imaging. Over the last 30 years, such survey datasets have become an integral part of observational astronomy, revolutionising many parts of cosmology and galaxy evolution studies. This thesis has capitalised on those developments, producing a combined optical+near-IR catalogue over the largest area to comparable depths so far. That data has been used to address a number of important astronomical challenges, specifically the computation of photometric redshifts and the identification of a sample of candidate $z \gtrsim 5$ galaxies. The current chapter summarises and concludes all this work, revisiting the core findings in each of the three parts of this thesis.

5.1 Summary of Chapter 2: combined DES+VIDEO catalogue

Chapter 2 described the process of merging the optical *griz* data from the Dark Energy Survey (DES) deep fields with the near-infrared *ZYJHK_s* VISTA Deep Extragalactic Observations Survey (VIDEO). The two datasets were added together via a specially created automated data processing pipeline, which extracted forced photometry directly from the VIDEO imaging using a deeper DES *r + i + z* image for source detections. The pipeline then added the measured

VIDEO fluxes to pre-existing DES catalogues, producing a catalogue that contains 2 443 576 total detections originating from 2 152 575 unique sources, altogether spanning a total footprint of 12.1 deg^2 . A large portion covering 10.8 deg^2 reaches 5σ limiting magnitudes of at least $z_{5\sigma} = 24.9$, of which 4.3 deg^2 reaches at least $z_{5\sigma} = 26.2$.

While creating the DES+VIDEO pipeline, this thesis produced two main results that are important to survey science in general. For starters, it has provided further evidence for a well-supported (e.g. Labb   et al. 2003; Gawiser et al. 2006; Whitaker et al. 2011; Muzzin et al. 2013) vital requirement for processing survey datasets: if imaging has been resampled to a different pixelscale, it is necessary to correct photometric errors to account for correlated noise. Section 2.2.4.3 showed that without such corrections, the VIDEO uncertainties measured by **SExtractor** would be underestimated by as much as a factor of 2 in some filters. Because accurate uncertainties are important to many applications in astronomy (not in the least to photometric redshifts), it is imperative that anyone who wishes to compile catalogues from resampled imaging must thus take great care to measure the photometric noise properties correctly. A second key finding was presented in Section 2.3.3, and concerns the power of forced photometry when combining different surveys. By extracting VIDEO forced photometry via DES detection image deeper than VIDEO, this thesis managed to increase the total number of (unique) VIDEO detections by 22% compared to the number that would have been obtained by simply position matching the existing DES and VIDEO tables. In a sub-region where the DES data is especially deep compared to VIDEO, this increase was as high as 34%. This thesis therefore further adds to the existing evidence for the power of combining surveys through forced photometry. It has demonstrated clearly that a deeper survey can be leveraged to extract more flux information from a shallower one, increasing the number of secure detections in the shallower dataset, while at the same time adding more wavebands to the deeper. In this way, forced photometry unlocks a form of synergy between surveys at different wavelengths. This thesis therefore advocates that observatories with different wavelength coverage coordinate surveys in the same fields. It also recommends that forced photometry be used to take full advantage of any overlapping datasets that already exist.

The most important result of Chapter 2, however, is probably the DES+VIDEO catalogue itself. As mentioned above, this dataset contains VIDEO fluxes

from more (generally fainter) objects than can be achieved with VIDEO alone, boosting the amount of information available from this survey. Besides, Section 1.1.4 outlined the great potential of multiwavelength datasets, which have a wide range of applications in modern astronomy. Together, the ten *grizZYJHK_s* filters in the DES+VIDEO catalogue provide a wealth of multiwavelength information over optical and near-infrared frequencies. Since both the DES and VIDEO datasets used in this thesis have recently become publicly available, the DES+VIDEO catalogue thus forms a useful resource for the astronomical community.

5.1.1 Suggestions for future work

One of the ways that future work could improve the DES+VIDEO catalogue is by measuring aperture corrections for the photometry. As explained in Section 2.2.5.2, such corrections were not strictly necessary for the work in this thesis, and due to the challenging and time-consuming nature of accurately measuring the inhomogenous DES PSF this was considered beyond the scope of this thesis. However, including such corrections would further add to the value of the catalogue as a resource for the general astronomical community.

In a similar vein, the scientific value of the catalogue could be increased by the addition of *Spitzer* IRAC fluxes from SERVS (see Section 4.2.4). Such measurements would likely not have been useful for photometric redshifts (Hildebrandt et al. 2010), but would further broaden the scope of the catalogue for general use. Forced photometry with DES detection images can increase the amount of information extracted from the SERVS imaging, and at the same time add mid-IR wavebands to the catalogue. It is worth noting that, due to the large difference in PSF size between the DES and IRAC imaging, this endeavour would require a method for accurately deblending IRAC sources, such as that used by e.g. Labb   et al. (2006) and Muzzin et al. (2013).

Finally, in future, improvements to the catalogue can be made by adding data from later DES and VIDEO releases, which are expected to have higher limiting magnitudes and include imaging for the full set of VIDEO filters in all tiles. Thanks to the automated nature of the pipeline created for this thesis, this suggestion should be relatively straightforward to implement.

5.2 Summary of Chapter 3: photometric redshifts

Chapter 3 focused on computing photometric redshifts for the DES+VIDEO catalogue created in Chapter 2. The first point of action was to test 18 configurations of the template fitting code **LePHARE** on a sample of 35 596 DES+VIDEO sources with spectroscopic redshifts. These configurations were based on two sets of templates (the **AVEROI_NEW** SEDs, and the **COSMOS** SEDs), and nine different options for photo-z code features and input photometry were trialled for each of the two sets. To assess the performance of each setup, the quality of the output photometric redshifts was quantified via a series of eight accuracy metrics capturing the bias, scatter, outlier fractions, and high-redshift accuracy, measured using the known spectroscopic redshifts.

Section 3.6 then compared and analysed the metrics of all the setups, in order to find the most suitable configuration (for each template set) to use on the full DES+VIDEO catalogue. The primary aim here was to achieve optimal performance at high redshifts. For both template sets, it was concluded that the most suitable high-redshift setups add emission lines to the SEDs, apply adaptive offsets, apply an $N(z)$ prior, use fixed aperture photometry, and include near-IR VIDEO fluxes. Regarding the extinction parameters, it was established that values of up to $E(B - V) = 0.5$ perform best for the **COSMOS** templates, while the most appropriate values for the **AVEROI_NEW** templates include extinction up to $E(B - V) = 2.0$.

Section 3.6 also investigated what the DES+VIDEO tests imply for photometric redshift studies in general. This analysis omitted the high-redshift accuracy metrics, and looked only at those statistics that apply to the full testing sample – i.e. the bias, scatter, and outlier fractions. These primarily reflect the accuracy for the low-redshift ($z_{\text{spec}} \lesssim 1.5$) majority of galaxies in the sample, which is also approximately the range that most photo-z studies tend to focus on. The DES+VIDEO tests established the following conclusions:

- The **COSMOS** templates perform best with extinction values up to $E(B - V) = 0.5$. For the **AVEROI_NEW** SEDs, the best accuracy is instead achieved with values up to $E(B - V) = 1.0$. Notably, this is higher than the values of $E(B - V) \leq 0.6$ used in the Arnouts et al. (2007) study that produced the **AVEROI_NEW** templates, and substantially higher than the $E(B - V) \leq 0.3$

recommendation in the **README** files of the **LePHARE** installation. Future users of these SEDs may want to bear this in mind.

- For the **COSMOS** templates, enabling the option to let **LePHARE** add emission lines to the templates generally improves photometric redshifts. The data did not conclusively establish whether this is also the case for the **AVEROI_NEW** SEDs.
- The DES+VIDEO tests also support the usefulness of the $N(z)$ prior and adaptive offset features in **LePHARE**. It was demonstrated that the use of the prior improves the scatter for both template sets. Applying adaptive offsets was similarly found to be helpful, particularly with regard to the scatter and outlier fractions.
- Fixed aperture magnitudes constitute the best way to supply input photometry for the **COSMOS** templates. For the **AVEROI_NEW** SEDs, it was inconclusive whether auto or fixed aperture photometry yield better results.
- An especially important part of Section 3.6 consisted of an in-depth analysis of the impact of near-IR photometry on the DES+VIDEO redshifts. This thesis resoundingly showed that adding near-IR VIDEO fluxes to the optical DES data leads to considerably lower scatter and outlier fractions. For the **AVEROI_NEW** and **COSMOS** templates respectively, the near-IR data reduced σ by 1.1% and 25.5%, σ_{68} by 45.9% and 56.4%, $f_{2\sigma}$ by 12.9% and 25.3%, and $f_{3\sigma}$ by 14.3% and 33.5%. While the beneficial effect of near-IR data was generally noted at all redshifts, the VIDEO photometry was found to be especially important in the area around $2 \leq z_{\text{phot}} \leq 3$, where the 4000 Å break has redshifted out of the optical filters but can still be constrained in the near-IR. Altogether, the data thus presents strong evidence for the importance of the near-IR waveband for photometric redshift studies.

This leaves the template comparison as the final result from Section 3.6 to be summed up here. When considering the outcome for the (low-redshift dominated) totality of redshifts, it was found that the **COSMOS** templates conclusively outperform the **AVEROI_NEW** SEDs in most metrics and configurations. At high redshifts, however, the latter correctly retrieve more true high-redshift galaxies. To ensure the DES+VIDEO catalogue is as versatile as possible as a resource for the astronomical community, photo-zs were added for both collections of

templates, using the best high-redshift configurations summarised earlier in this section. Notably, for the **COSMOS** templates this configuration is the same as the best-performing setup in the general (low-redshift dominated) regime, and it is recommended that any low-redshift studies use the photo-zs from these templates. However, the high-redshift search in Chapter 4 of this thesis has made use of the **AVEROI_NEW** redshifts instead.

The task of separating out the nonsensical photo-zs of stellar sources was explored in Section 3.7.1, which described the process of identifying a star-galaxy separation method suited to this thesis. The strategy that was proposed is an SED-based method where galaxies are identified based on the goodness-of-fit of the best-fitting galaxy template compared to the best-fitting stellar template. An advantage of this approach is that it works on unresolved galaxies (common at e.g. high redshifts), unlike most standard methods for survey datasets which often rely on morphological information. When validated on a set of spectroscopically confirmed stars and galaxies, the method in this thesis was found to be highly successful, misclassifying only around 2% of true stars as galaxies. A related important discovery is that near-infrared data is of crucial importance to SED-based separation — without including VIDEO data in the fitting, the stellar contamination rate increased enormously by a factor of 22–35. When taken together with the improvement in photo-z accuracy established before, this result further emphasises the importance of near-IR data to modern survey science.

5.2.1 Suggestions for future work

The scope of the general photometric redshift research in Chapter 3 could be expanded by applying the DES+VIDEO tests to a wider range of codes. Even though **LePHARE** is likely the most suitable code for this thesis (as explained in Section 3.2.1), extending the tests to other codes will yield additional useful results for the field of photometric redshift studies.

Secondly, future work could make the conclusions of the star-galaxy separation validation in Section 3.7.1 more widely applicable. That verification process used a truth catalogue with a far smaller number of stars (compared to galaxies) than is realistic for survey datasets. The stellar contamination ratios found in this thesis therefore cannot be meaningfully compared to those of other studies. To give a more realistic estimate of the real stellar contamination, which would allow direct comparisons between the star-galaxy separation method in

this thesis and others in the literature, future work could apply the method to a representative sample of simulated stars and galaxies. It is important that such simulations accurately replicate the DES+VIDEO imaging by taking account of the variable PSF and image quality, especially for the inhomogeneous DES data.

5.3 Summary of Chapter 4: high-redshift galaxies

Chapter 4 presented the search for the brightest ($m_{AB} < 25.0$) high-redshift Lyman-break galaxies in the DES+VIDEO dataset. This endeavour made use of the photometric redshifts computed in Chapter 3, applying various targeted cuts that resulted in a final sample of 34 unique galaxy candidates at $z \sim 5$ and 36 at $z \sim 6$. Accounting for the magnitude and depth requirements in the selection cuts, the effective search areas over which these two candidate sets have been gathered equal 6.1 deg^2 and 4.9 deg^2 respectively, the largest footprints currently available in datasets of comparable depth.

The $z \sim 6$ candidates were then compared against those from two similar $z \sim 6$ studies in the literature (by Bowler et al. 2015 and Willott et al. 2013). The first such comparison in Section 4.3.1.1 looked at the expected number of galaxies based on these two studies. It was found that the DES+VIDEO sample size is larger than the numbers from previous studies, but several reasons were identified why those expected numbers are likely too low or unreliable. Subsequently, there is no evidence for considerable contamination in the $z \sim 6$ sample.

The sizeable number of bright $z \sim 6$ DES+VIDEO candidates revealed in this thesis thus holds great scientific promise. Prior to this thesis, the known number of such galaxies totalled only 32. Of the 36 DES+VIDEO candidates found by the current work, 33 are new (the 3 others had already been discovered by Willott et al. 2013). Hence, this thesis more than doubles the number of known bright galaxy candidates at $z \sim 6$.

In Section 4.3.1.2, the second comparison with previous research looked at the candidates from the study by Willott et al. (2013), which partly overlaps the DES+VIDEO footprint. It was found that the results from this thesis agree well with theirs. The DES+VIDEO sample contains all three of the Willott et al. (2013) candidates that pass the $z_{AB} < 25.0$ magnitude cut imposed in this thesis, including an object spectroscopically confirmed at $z_{\text{spec}} = 6.068$. Since Willott et al. (2013) selected their candidates via colour cuts rather than photometric redshifts,

this agreement presents an independent corroboration of the selection method in this thesis. It therefore imparts a good degree of confidence in the DES+VIDEO sample.

5.3.1 Suggestions for future work

The most important avenue of future research for this thesis is obtaining spectra for as many of the DES+VIDEO candidates as possible. If the high-redshift status of these objects is confirmed, this would prove directly that the selection method in this thesis is able to isolate real (previously unknown) high-redshift galaxies. Such confirmation would also increase the number of known bright $z \gtrsim 5$ galaxies with a confirmed high-redshift status, and the spectra would enable measurements of the strength of the Lyman- α line. This can be useful for constraining reionisation in ways described in Section 1.3.2.2.

If the effectiveness of the selection method in this thesis is confirmed by spectroscopic observations, the samples of DES+VIDEO high-redshift galaxies can be used to study the many galaxy properties outlined in Section 1.3.2.2. The UV luminosity function at $z = 6$ is potentially the most impactful of these. As explained in Section 1.3.2.2, this quantity is currently not well constrained at the very bright end due to the small number of previously known objects in this regime. Since the 36 $z \sim 6$ DES+VIDEO objects double the number of available $m_{AB} < 25.0$ candidates at this redshift, they would be able to make a considerable contribution to constraining the LF. To compute the DES+VIDEO LF accurately it will be necessary to perform simulations that measure the purity and completeness of the DES+VIDEO candidate selection. Once again, it is important that such simulations take the inhomogeneous nature of the imaging into account properly. Besides the luminosity function, galaxy stellar masses are another major property from Section 1.3.2.2 that the DES+VIDEO sample could help to investigate. Thanks to the aforementioned IRAC SERVS data, many candidates have available mid-IR data that can probe the observed-frame optical emission from lower-mass stars, which may make up the bulk of stellar mass. Lastly, dust attenuation at high redshifts is a further property suggested in Section 1.3.2.2 that can be examined via the DES+VIDEO candidates.

By finding a substantial number of very bright candidate Lyman-break galaxies at $z \sim 5$ and $z \sim 6$, this thesis has therefore laid some of the groundwork for potential important discoveries within the field of high-redshift galaxy evolution.

Bibliography

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018a, Phys. Rev. D, 98, 043526
- Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018b, ApJS, 239, 18
- Abdalla, F. B., Amara, A., Capak, P., et al. 2008, MNRAS, 387, 969
- Abdalla, F. B., Banerji, M., Lahav, O., & Rashkov, V. 2011, MNRAS, 417, 1891
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634
- Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, The Astrophysical Journal Supplement Series, 240, 23
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
- Allen, D. A. 1976, MNRAS, 174, 29P
- Almosallam, I. A., Jarvis, M. J., & Roberts, S. J. 2016, MNRAS, 462, 726
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
- Arnouts, S., & Ilbert, O. 2011, LePHARE: Photometric Analysis for Redshift Estimate, Astrophysics Source Code Library, ascl:1108.009
- Arnouts, S., Walcher, C. J., Le Fèvre, O., et al. 2007, A&A, 476, 137
- Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, MNRAS, 404, 86
- Ball, N. M., Loveday, J., Fukugita, M., et al. 2004, MNRAS, 348, 1038
- Banerji, M., Abdalla, F. B., Lahav, O., & Lin, H. 2008, MNRAS, 386, 1219
- Banerji, M., Jouvel, S., Lin, H., et al. 2015, MNRAS, 446, 2523
- Baum, W. A. 1962, in IAU Symposium, Vol. 15, Problems of Extra-Galactic Research, ed. G. C. McVittie, 390
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729

- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
- Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, The Astrophysical Journal Supplement Series, 149, 289
- Benítez, N. 2000, ApJ, 536, 571
- Benson, A. J., Bower, R. G., Frenk, C. S., et al. 2003, ApJ, 599, 38
- Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, ApJ, 753, 152
- Bertin, E. 2010, SWarp: Resampling and Co-adding FITS Images Together, Astrophysics Source Code Library, ascl:1010.068
- Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
- 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
- Bevington, P. R. 1969, Data reduction and error analysis for the physical sciences
- Blake, C., Collister, A., Bridle, S., & Lahav, O. 2007, MNRAS, 374, 1527
- Bohlin, R. C., Colina, L., & Finley, D. S. 1995, AJ, 110, 1316
- Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144
- Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
- Bonfield, D. G., Sun, Y., Davey, N., et al. 2010, MNRAS, 405, 987
- Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P., & Franx, M. 2006, ApJ, 653, 53
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, ApJ, 737, 90
- 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. 2014, ApJ, 793, 115
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2012, MNRAS, 426, 2772
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, MNRAS, 440, 2810
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, MNRAS, 452, 1817
- Bradley, L. D., Zitrin, A., Coe, D., et al. 2014, ApJ, 792, 76
- Brammer, G. B., Van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
- Brodwin, M., Lilly, S. J., Porciani, C., et al. 2006, ApJS, 162, 20

- Brown, M. J. I., Moustakas, J., Kennicutt, R. C., et al. 2017, *ApJ*, 847, 136
- Brunner, R. J., Connolly, A. J., Szalay, A. S., & Bershadsky, M. A. 1997, *ApJ*, 482, L21
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Burchett, J. N., Tripp, T. M., Wang, Q. D., et al. 2018, *MNRAS*, 475, 2067
- Butchins, S. A. 1981, *A&A*, 97, 407
- Butchins, S. A. 1983, *MNRAS*, 203, 1239
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, *ApJ*, 712, 511
- Carrasco Kind, M., & Brunner, R. J. 2013, *MNRAS*, 432, 1483
- Casertano, S., de Mello, D., Dickinson, M., et al. 2000, *AJ*, 120, 2747
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, *Phys. Rep.*, 541, 45
- Castellano, M., Fontana, A., Paris, D., et al. 2010, *A&A*, 524, A28
- Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, *ApJ*, 542, 464
- Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, *ApJS*, 43, 393
- Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, 328, 1039
- Collister, A. A., & Lahav, O. 2004, *PASP*, 116, 345
- Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, *AJ*, 110, 2655
- Cowie, L. L., & Hu, E. M. 1998, *AJ*, 115, 1319
- Cox, T. J., Di Matteo, T., Hernquist, L., et al. 2006, *ApJ*, 643, 692
- Crocce, M., Cabré, A., & Gaztañaga, E. 2011, *MNRAS*, 414, 329
- Curtis-Lake, E., McLure, R. J., Pearce, H. J., et al. 2012, *MNRAS*, 422, 1425
- Curtis-Lake, E., McLure, R. J., Dunlop, J. S., et al. 2016, *MNRAS*, 457, 440
- Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, 670, 156
- Dahlen, T., Mobasher, B., Dickinson, M., et al. 2010, *ApJ*, 724, 425
- Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, *ApJ*, 775, 93
- Dalton, G. B., Caldwell, M., Ward, A. K., et al. 2006, in Proc. SPIE, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 62690X

- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, MNRAS, 401, 1043
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
- Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJ, 660, L1
- De Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62
- de Simoni, F., Sobreira, F., Carnero, A., et al. 2013, MNRAS, 435, 3017
- Desai, S., Armstrong, R., Mohr, J. J., et al. 2012, ApJ, 757, 83
- Dickinson, M., Stern, D., Giavalisco, M., et al. 2004, ApJ, 600, L99
- Diehl, H. T., Abbott, T. M. C., Annis, J., et al. 2014, in Proc. SPIE, Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V, 91490V
- Dijkstra, M. 2014, PASA, 31, e040
- Djorgovski, S., Spinrad, H., McCarthy, P., & Strauss, M. A. 1985, ApJ, 299, L1
- Djorgovski, S. G., Gal, R. R., Odewahn, S. C., et al. 1998, in Wide Field Surveys in Cosmology, ed. S. Colombi, Y. Mellier, & B. Raban, Vol. 14, 89
- Djorgovski, S. G., Mahabal, A., Drake, A., Graham, M., & Donalek, C. 2013, Sky Surveys, ed. T. D. Oswalt & H. E. Bond, 223
- Drlica-Wagner, A., Sevilla-Noarbe, I., Rykoff, E. S., et al. 2018, ApJS, 235, 33
- Dunlop, J. S. 2013, in Astrophysics and Space Science Library, Vol. 396, The First Galaxies, ed. T. Wiklind, B. Mobasher, & V. Bromm, 223
- Dye, S., Lawrence, A., Read, M. A., et al. 2018, MNRAS, 473, 5113
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
- Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJ, 763, L7
- Elvin-Poole, J., Crocce, M., Ross, A. J., et al. 2018, Phys. Rev. D, 98, 042006
- Emerson, J. P., Sutherland, W. J., McPherson, A. M., et al. 2004, The Messenger, 117, 27
- Epchtein, N., de Batz, B., Capoani, L., et al. 1997, The Messenger, 87, 27
- Fausti Neto, A., da Costa, L. N., Carnero, A., et al. 2018, Astronomy and Computing, 24, 52
- Feldmann, R., Carollo, C. M., Porciani, C., et al. 2006, MNRAS, 372, 565

- Fernández-Soto, A., Lanzetta, K. M., Chen, H.-W., Levine, B., & Yahata, N. 2002, MNRAS, 330, 889
- Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
- Finkelstein, S. L. 2016, PASA, 33, e037
- Finkelstein, S. L., Papovich, C., Giavalisco, M., et al. 2010, ApJ, 719, 1250
- Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, ApJ, 756, 164
- Finkelstein, S. L., Papovich, C., Dickinson, M., et al. 2013, Nature, 502, 524
- Finkelstein, S. L., Song, M., Behroozi, P., et al. 2015a, ApJ, 814, 95
- Finkelstein, S. L., Ryan, Russell E., J., Papovich, C., et al. 2015b, ApJ, 810, 71
- Fitzpatrick, E. L. 1986, AJ, 92, 1068
- Flaughher, B. 2005, International Journal of Modern Physics A, 20, 3121
- Flaughher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
- Fontana, A., D'Odorico, S., Poli, F., et al. 2000, AJ, 120, 2206
- Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385
- Furusawa, H., Kosugi, G., Akiyama, M., et al. 2008, The Astrophysical Journal Supplement Series, 176, 1
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Sci. Rev., 123, 485
- Gawiser, E., Van Dokkum, P. G., Herrera, D., et al. 2006, ApJS, 162, 1
- Gerdes, D. W., Sypniewski, A. J., McKay, T. A., et al. 2010, ApJ, 715, 823
- Giallongo, E., Menci, N., Poli, F., D'Odorico, S., & Fontana, A. 2000, ApJ, 530, L73
- Giavalisco, M., Macchetto, F. D., & Sparks, W. B. 1994, A&A, 288, 103
- Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004a, ApJ, 600, L93
- Giavalisco, M., Dickinson, M., Ferguson, H. C., et al. 2004b, ApJ, 600, L103
- González, V., Bouwens, R., Illingworth, G., et al. 2014, ApJ, 781, 34
- González, V., Labbé, I., Bouwens, R. J., et al. 2011, ApJ, 735, L34
- Grazian, A., Fontana, A., de Santis, C., et al. 2006, A&A, 449, 951
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, The Astrophysical Journal Supplement Series, 197, 35

- Groth, E. J., Kristian, J. A., Lynds, R., et al. 1994, in Bulletin of the American Astronomical Society, Vol. 26, American Astronomical Society Meeting Abstracts, 1403
- Gruppioni, C., Ciliegi, P., Rowan-Robinson, M., et al. 1999, MNRAS, 305, 297
- Guhathakurta, P., Tyson, J. A., & Majewski, S. R. 1990, ApJ, 357, L9
- Gwyn, S. D. J. 2012, AJ, 143, 38
- Gwyn, S. D. J., & Hartwick, F. D. A. 1996, ApJ, 468, L77
- Hamilton, A. J. S., & Tegmark, M. 2004, MNRAS, 349, 115
- Hibon, P., Malhotra, S., Rhoads, J., & Willott, C. 2011, ApJ, 741, 101
- Hibon, P., Cuby, J. G., Willis, J., et al. 2010, A&A, 515, A97
- Hickox, R. C., & Alexander, D. M. 2018, Annual Review of Astronomy and Astrophysics, 56, 625
- High, F. W., Stubbs, C. W., Rest, A., Stalder, B., & Challis, P. 2009, AJ, 138, 110
- Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
- Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31
- Hogg, D. W. 1999, arXiv e-prints, astro
- Huang, K.-H., Ferguson, H. C., Ravindranath, S., & Su, J. 2013, ApJ, 765, 68
- Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
- Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
- Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
- Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, in Proc. SPIE, Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies, ed. P. J. Quinn & A. Bridger, 411–422
- Ivezić, Ž., Lupton, R. H., Schlegel, D., et al. 2004, Astronomische Nachrichten, 325, 583
- Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Serbian Astronomical Journal, 176, 1
- Jannuzzi, B. T., & Dey, A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies, ed. R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner, 111

- Jarvis, M. J., Bonfield, D. G., Bruce, V. A., et al. 2013, MNRAS, 428, 1281
- Jiang, L., Egami, E., Kashikawa, N., et al. 2011, ApJ, 743, 65
- Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, in Proc. SPIE, Vol. 4836, Survey and Other Telescope Technologies and Discoveries, ed. J. A. Tyson & S. Wolff, 154–164
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in Proc. SPIE, Vol. 7733, Ground-based and Airborne Telescopes III, 77330E
- Kakiuchi, K., Ellis, R. S., Laporte, N., et al. 2018, MNRAS, 479, 43
- Kashikawa, N., Shimasaku, K., Yasuda, N., et al. 2004, PASJ, 56, 1011
- Kelly, P. L., von der Linden, A., Applegate, D. E., et al. 2014, MNRAS, 439, 28
- Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189
- Kessler, R., Marriner, J., Childress, M., et al. 2015, AJ, 150, 172
- Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, ApJ, 467, 38
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
- Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, ApJS, 209, 3
- Koo, D. C. 1985, AJ, 90, 418
- Kron, R. G. 1980, ApJS, 43, 305
- Kurtz, M. J., Geller, M. J., Fabricant, D. G., Wyatt, W. F., & Dell'Antonio, I. P. 2007, AJ, 134, 1360
- Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, ApJ, 649, L67
- Labbé, I., Franx, M., Rudnick, G., et al. 2003, AJ, 125, 1107
- Lang, D., Hogg, D. W., & Schlegel, D. J. 2016, AJ, 151, 36
- Lanzetta, K. M., Yahil, A., & Fernández-Soto, A. 1996, Nature, 381, 759
- Lasker, B. M., Sturch, C. R., McLean, B. J., et al. 1990, AJ, 99, 2019
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
- Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, A&A, 439, 845
- Leistedt, B., & Hogg, D. W. 2017, ApJ, 838, 5
- Leitherer, C., Ferguson, H. C., Heckman, T. M., & Lowenthal, J. D. 1995, ApJ, 454, L19
- Loh, E. D., & Spillar, E. J. 1986, ApJ, 303, 154

- Madau, P. 1995, ApJ, 441, 18
- Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
- Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106
- Malhotra, S., Rhoads, J. E., Pirzkal, N., et al. 2005, ApJ, 626, 666
- Mathis, J. S. 1994, ApJ, 422, 176
- Matthee, J., Sobral, D., Santos, S., et al. 2015, MNRAS, 451, 400
- Mauduit, J. C., Lacy, M., Farrah, D., et al. 2012, Publications of the Astronomical Society of the Pacific, 124, 714
- Mawatari, K., Yamada, T., Fazio, G. G., Huang, J.-S., & Ashby, M. L. N. 2016, PASJ, 68, 46
- McCall, M. L., Rybski, P. M., & Shields, G. A. 1985, ApJS, 57, 1
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
- McLure, R. J., Cirasuolo, M., Dunlop, J. S., Foucaud, S., & Almaini, O. 2009, MNRAS, 395, 2196
- McLure, R. J., Dunlop, J. S., Cirasuolo, M., et al. 2010, MNRAS, 403, 960
- McLure, R. J., Cirasuolo, M., Dunlop, J. S., et al. 2006, MNRAS, 372, 357
- McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, The Messenger, 154, 35
- Merlin, E., Fontana, A., Castellano, M., et al. 2018, MNRAS, 473, 2098
- Messier, C. 1781, Catalogue des Nébuleuses et des Amas d'Étoiles (Catalog of Nebulae and Star Clusters), Connaissance des Temps ou des Mouvements Célestes
- Minkowski, R. L., & Abell, G. O. 1963, The National Geographic Society-Palomar Observatory Sky Survey, ed. K. A. Strand (the University of Chicago Press), 481
- Mobasher, B., Capak, P., Scoville, N. Z., et al. 2007, ApJS, 172, 117
- Mohr, J. J., Armstrong, R., Bertin, E., et al. 2012, in Proc. SPIE, Vol. 8451, Software and Cyberinfrastructure for Astronomy II, 84510D
- Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, PASP, 130, 074501
- Mouhcine, M., Lewis, I., Jones, B., et al. 2005, MNRAS, 362, 1143
- Moustakas, J., Kennicutt, Jr., R. C., & Tremonti, C. A. 2006, ApJ, 642, 775
- Muzzin, A., Marchesini, D., Van Dokkum, P. G., et al. 2009, ApJ, 701, 1839

- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, The Astrophysical Journal Supplement Series, 206, 8
- Nagao, T., Taniguchi, Y., Kashikawa, N., et al. 2004, ApJ, 613, L9
- Nakajima, R., Mandelbaum, R., Seljak, U., et al. 2012, MNRAS, 420, 3240
- Nayyeri, H., Mobasher, B., Hemmati, S., et al. 2014, ApJ, 794, 68
- Neugebauer, G., & Leighton, R. B. 1969, Two-micron sky survey. A preliminary catalogue
- Nyland, K., Lacy, M., Sajina, A., et al. 2017, ApJS, 230, 9
- Oesch, P. A., Bouwens, R. J., Carollo, C. M., et al. 2010a, ApJ, 709, L21
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2010b, ApJ, 709, L16
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, ApJ, 773, 75
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, ApJ, 786, 108
- Oesch, P. A., Brammer, G., Van Dokkum, P. G., et al. 2016, ApJ, 819, 129
- Ono, Y., Ouchi, M., Mobasher, B., et al. 2012, ApJ, 744, 83
- Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, ApJ, 611, 660
- Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301
- Ouchi, M., Mobasher, B., Shimasaku, K., et al. 2009, ApJ, 706, 1136
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, ApJ, 723, 869
- Partridge, R. B., & Peebles, P. J. E. 1967, ApJ, 147, 868
- Peacock, J. A., Schneider, P., Efstathiou, G., et al. 2006, ESA-ESO Working Group on "Fundamental Cosmology", Tech. rep.
- Peebles, P. J. E. 1993, Principles of Physical Cosmology
- Pello, R., Miralles, J. M., Le Borgne, J.-F., et al. 1996, A&A, 314, 73
- Pentericci, L., Fontana, A., Vanzella, E., et al. 2011, ApJ, 743, 132
- Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, ApJ, 793, 113
- Pickles, A. J. 1998, PASP, 110, 863
- Pierre, M., Valtchanov, I., Altieri, B., et al. 2004, J. Cosmology Astropart. Phys., 9, 011

- Pogge, R. W., Atwood, B., Brewer, D. F., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7735, Proc. SPIE, 77350A
- Poli, F., Giallongo, E., Fontana, A., et al. 2003, *ApJ*, 593, L1
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, 663, 81
- Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E., & Rocca-Volmerange, B. 1984, *A&A*, 132, 389
- Puschell, J. J., Owen, F. N., & Laing, R. A. 1982, *ApJ*, 257, L57
- Quadri, R., Marchesini, D., Van Dokkum, P., et al. 2007, *AJ*, 134, 1103
- Ramos, B. H. F., Pellegrini, P. S., Benoist, C., et al. 2011, *AJ*, 142, 41
- Rector, T. A., Levay, Z. G., Frattare, L. M., Arcand, K. K., & Watzke, M. 2017, *PASP*, 129, 058007
- Rector, T. A., Levay, Z. G., Frattare, L. M., English, J., & Pu'uohau-Pummill, K. 2007, *AJ*, 133, 598
- Reid, I. N., Brewer, C., Brucato, R. J., et al. 1991, *PASP*, 103, 661
- Rhoads, J. E., & Malhotra, S. 2001, *ApJ*, 563, L5
- Rhoads, J. E., Malhotra, S., Dey, A., et al. 2000, *ApJ*, 545, L85
- Riechers, D. A., Capak, P. L., Carilli, C. L., et al. 2010, *ApJ*, 720, L131
- 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., et al. 2013, *ApJ*, 768, 71
- Rogers, A. B., McLure, R. J., Dunlop, J. S., et al. 2014, *MNRAS*, 440, 3714
- Rudnick, G., Franx, M., Rix, H.-W., et al. 2001, *AJ*, 122, 2205
- Rykoff, E. S., Rozo, E., & Keisler, R. 2015, arXiv e-prints, arXiv:1509.00870
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, 799, 183
- Salvato, M., Ilbert, O., & Hoyle, B. 2019, *Nature Astronomy*, 3, 212
- Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
- Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, *MNRAS*, 445, 1482
- Sarajedini, V. L., Green, R. F., Griffiths, R. E., & Ratnatunga, K. 1999, *ApJS*, 121, 417

- Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, AJ, 113, 1
- Schechter, P. 1976, ApJ, 203, 297
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, The Astrophysical Journal Supplement Series, 172, 1
- Seaton, M. J. 1979, MNRAS, 187, 73P
- Sevilla, I., Armstrong, R., Bertin, E., et al. 2011, arXiv e-prints, arXiv:1109.6741
- Sevilla-Noarbe, I., Hoyle, B., Marchā, M. J., et al. 2018, MNRAS, 481, 5451
- Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, ApJ, 509, 103
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
- Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
- Songaila, A. 2004, AJ, 127, 2598
- Soumagnac, M. T., Abdalla, F. B., Lahav, O., et al. 2015, MNRAS, 450, 666
- Spinrad, H., Stern, D., Bunker, A., et al. 1998, AJ, 116, 2617
- Staniszewski, Z., Ade, P. A. R., Aird, K. A., et al. 2009, ApJ, 701, 32
- Stanway, E. R., Bunker, A. J., & McMahon, R. G. 2003, MNRAS, 342, 439
- Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
- Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, MNRAS, 408, 1628
- Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, ApJ, 728, L2
- Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
- Stecher, T. P. 1965, ApJ, 142, 1683
- Steidel, C. C., & Hamilton, D. 1993, AJ, 105, 2017
- Sutherland, W., Emerson, J., Dalton, G., et al. 2015, A&A, 575, A25
- Swanson, M. E. C., Tegmark, M., Hamilton, A. J. S., & Hill, J. C. 2008, MNRAS, 387, 1391
- Szalay, A. S., Connolly, A. J., & Szokoly, G. P. 1999, AJ, 117, 68
- Tanaka, M. 2015, ApJ, 801, 20
- Thompson, R. I., Illingworth, G., Bouwens, R., et al. 2005, AJ, 130, 1

- Tilvi, V., Rhoads, J. E., Hibon, P., et al. 2010, *ApJ*, 721, 1853
- Toft, S., Soucail, G., & Hjorth, J. 2003, *MNRAS*, 344, 337
- Van der Burg, R. F. J., Hildebrandt, H., & Erben, T. 2010, *A&A*, 523, A74
- Vanzella, E., Cristiani, S., Fontana, A., et al. 2004, *A&A*, 423, 761
- Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, *ApJ*, 695, 1163
- Wadadekar, Y. 2005, *PASP*, 117, 79
- Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, *Ap&SS*, 331, 1
- Walter, F., Decarli, R., Carilli, C., et al. 2012, *Nature*, 486, 233
- Wang, Y., Bahcall, N., & Turner, E. L. 1998, *AJ*, 116, 2081
- Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, *Phys. Rep.*, 530, 87
- Weiner, B. J., Phillips, A. C., Faber, S. M., et al. 2005, *ApJ*, 620, 595
- Weymann, R. J., Stern, D., Bunker, A., et al. 1998, *ApJ*, 505, L95
- Whitaker, K. E., Labb  , I., Van Dokkum, P. G., et al. 2011, *ApJ*, 735, 86
- Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, *MNRAS*, 464, 4176
- Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, *AJ*, 112, 1335
- Willott, C. J., McLure, R. J., Hibon, P., et al. 2013, *AJ*, 145, 4
- Wittman, D. M., Tyson, J. A., Dell'Antonio, I. P., et al. 2002, in *Proc. SPIE*, Vol. 4836, Survey and Other Telescope Technologies and Discoveries, ed. J. A. Tyson & S. Wolff, 73–82
- Wynn-Williams, G. 2016, Surveying the Skies: How Astronomers Map the Universe, *Astronomers' Universe* (Springer International Publishing)
- Xiang, F. Y., Li, A., & Zhong, J. X. 2011, *ApJ*, 733, 91
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, 195, 10
- Yan, H., Dickinson, M., Giavalisco, M., et al. 2006, *ApJ*, 651, 24
- Yee, H. K. C. 1998, ArXiv Astrophysics e-prints, astro-ph/9809347
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
- Yoshida, M., Shimasaku, K., Kashikawa, N., et al. 2006, *ApJ*, 653, 988

Appendix A

Configuration files

A.1 SWarp configuration file

```
# DES+VIDEO configuration file for SWarp 2.38.0
#
#----- Output
-----
IMAGEOUT_NAME DESTILE_VIDEOTILE_VIDEO_FILTER_swarped.fits  #
    Output filename
WEIGHTOUT_NAME DESTILE_VIDEOTILE_VIDEO_FILTER_swarped_weights.fits
    # Output weight-map filename

HEADER_ONLY N  # Only a header as an output file (Y/N)?
HEADER_SUFFIX .head  # Filename extension for additional headers

#----- Input Weights
-----

WEIGHT_TYPE MAP_WEIGHT # "BACKGROUND","MAP_RMS","MAP_VARIANCE"
    # or "MAP_WEIGHT"
WEIGHT_SUFFIX .weight.fits # Suffix to use for weight-maps
WEIGHT_IMAGE VIDEOTILE_FILTER_maxseeing0p90_2015-03-25_conf.fits
    # Weightmap filename if suffix not used
    # (all or for each weight-map)
WEIGHT_THRESH      # Bad pixel weight-threshold

RESCALE_WEIGHTS Y # Rescale input weights/variances (Y/N)?
```

```
#----- Co-addition
-----
COMBINE Y # Combine resampled images (Y/N)?
COMBINE_TYPE WEIGHTED # "MEDIAN", "AVERAGE", "MIN", "MAX"
                      # "WEIGHTED", "CHI2" or "SUM"

#----- Astrometry
-----
CELESTIAL_TYPE EQUATORIAL # "NATIVE", "PIXEL", "EQUATORIAL",
                           # "GALACTIC" or "ECLIPTIC"
PROJECTION_TYPE TAN # Any WCS projection code or "NONE"
PROJECTION_ERR 0.00001 # Maximum projection error (in output
                      # pixels), or 0 for no
                      approximation
CENTER_TYPE MANUAL # "MANUAL", "ALL" or "MOST"
CENTER DESTILE_RA,DESTILE_DEC # Coordinates of the image center
PIXELSCALE_TYPE MANUAL # "MANUAL", "FIT", "MIN", "MAX"
                      # or "MEDIAN"
PIXEL_SCALE 0.263 # Pixel scale
IMAGE_SIZE 10000, 10000 # Image size (0 = AUTOMATIC)

#----- Resampling
-----
RESAMPLE Y # Resample input images (Y/N)?
RESAMPLE_DIR ./ # Directory path for resampled images
RESAMPLE_SUFFIX _im.fits # filename extension for resampled images

RESAMPLING_TYPE LANCZOS3 # "NEAREST", "BILINEAR", "LANCZOS2",
                           # "LANCZOS3" or "LANCZOS4" (1 per
                           axis)
OVERSAMPLING 0 # Oversampling in each dimension
                  # (0 = automatic)
INTERPOLATE Y # Interpolate bad input pixels (Y/N)?
                  # (all or for each image)

FSCALASTRO_TYPE FIXED # "NONE" or "FIXED"
```

```
FSCALE_KEYWORD FLXSCALE # FITS keyword for the multiplicative
                  # factor to apply to each input
                  image
FSCALE_DEFAULT 1.0 # Default FSCALE value if not in header

GAIN_KEYWORD GAIN # FITS keyword for eff. gain (e-/ADU)
GAIN_DEFAULT 0.0 # Default gain if no FITS keyword found
                  # 0 = infinity (all or for each
                  image)

#----- Background subtraction
-----

SUBTRACT_BACK Y # Subtraction sky background (Y/N)?
                  # (all or for each image)

BACK_TYPE AUTO # "AUTO" or "MANUAL"
                  # (all or for each image)
BACK_DEFAULT 0.0 # Default background value in MANUAL
                  # (all or for each image)
BACK_SIZE 256 # Background mesh size (pixels)
                  # (all or for each image)
BACK_FILTERSIZE 3 # Background map filter range (meshes)
                  # (all or for each image)
BACK_FILTTHRESH 0.0 # Threshold above which the background-
                  # map filter operates

#----- Memory management
-----

VMEM_DIR . # Directory path for swap files
VMEM_MAX 3000 # Maximum amount of virtual memory (MB)
MEM_MAX 3000 # Maximum amount of usable RAM (MB)
COMBINE_BUFSIZE 2000 # RAM dedicated to co-addition(MB)

#----- Miscellaneous
-----

DELETE_TMPFILES Y # Delete temporary resampled FITS files
                  # (Y/N)?
```

```

COPY_KEYWORDS OBJECT # List of FITS keywords to propagate
                  # from the input to the output
                  headers
WRITE_FILEINFO N # Write information about each input
                  # file in the output image header
?
VERBOSE_TYPE FULL # "QUIET", "NORMAL" or "FULL"

NTHREADS 1 # 1 single thread

```

A.2 SExtractor configuration file

```

# Default configuration file for SExtractor 2.3b2
# EB 2003-02-07
#
#----- Catalog
-----

CATALOG_NAME test.cat # name of the output catalog
CATALOG_TYPE FITS_1.0 # "NONE", "ASCII_HEAD", "ASCII", "FITS_1.0"
                      # or "FITS_LDAC"

PARAMETERS_NAME sex.param # name of the file containing catalog
                           contents

#----- Extraction
-----

DETECT_TYPE CCD # "CCD" or "PHOTO"
FLAG_IMAGE flag.fits # filename for an input FLAG-image
DETECT_MINAREA 6 # minimum number of pixels above threshold
DETECT_THRESH 1.6 # <sigmas> or <threshold>, <ZP> in mag.arcsec-2
ANALYSIS_THRESH 1.5 # <sigmas> or <threshold>, <ZP> in mag.arcsec-2

FILTER Y # apply filter for detection ("Y" or "N")?
FILTER_NAME gauss_3.0_7x7.conv # name of the file containing the
                           filter

DEBLEND_NTHRESH 32 # Number of deblending sub-thresholds

```

```
DEBLEND_MINCONT 0.001 # Minimum contrast parameter for deblending

CLEAN Y # Clean spurious detections? (Y or N)?
CLEAN_PARAM 1.0 # Cleaning efficiency

MASK_TYPE CORRECT # type of detection MASKing: can be one of
# "NONE", "BLANK" or "CORRECT"

#WEIGHT_IMAGE weight.fits # bad pixel map (1 for good pixels and 0
# for bad)
#WEIGHT_TYPE MAP_WEIGHT # expects bad pixel map input

#----- Photometry
-----

PHOT_APERTURES 7.41 #MAG_APER aperture diameter(s) in pixels
PHOT_AUTOPARAMS 2.5, 3.5 # MAG_AUTO parameters: <Kron_fact>,<
# min_radius>
PHOT_PETROPARAMS 2.0, 3.5 # MAG_PETRO parameters: <Petrosian_fact>,
# <min_radius>
PHOT_AUTOAPERS 0.0,7.41 # <estimation>,<measurement> minimum
# apertures
# for MAG_AUTO and MAG_PETRO

SATUR_LEVEL 50000.0 # level (in ADUs) at which arises saturation
SATUR_KEY SATURATE # keyword for saturation level (in ADUs)

MAG_ZEROPPOINT 30.0 # magnitude zero-point
MAG_GAMMA 4.0 # gamma of emulsion (for photographic scans)
GAIN 0.0 # detector gain in e-/ADU
GAIN_KEY GAIN # keyword for detector gain in e-/ADU
PIXEL_SCALE 0 # size of pixel in arcsec (0=use FITS WCS info)

#-----Model-Fitting Photometry
-----
#PSF_NMAX 1
```

```
#----- Star/Galaxy Separation
-----
SEEING_FWHM 0 # stellar FWHM in arcsec
STARNNW_NAME sex.nnw # Neural-Network_Weight table filename
#----- Background
-----
BACK_SIZE 256 # Background mesh: <size> or <width>,<height>
BACK_FILTERSIZE 3 # Background filter: <size> or <width>,<height>
BACKPHOTO_TYPE GLOBAL # can be "GLOBAL" or "LOCAL"

#----- Check Image
-----
CHECKIMAGE_TYPE NONE # can be one of "NONE", "BACKGROUND",
# "MINIBACKGROUND", "-BACKGROUND", "OBJECTS",
# "-OBJECTS", "SEGMENTATION", "APERTURES",
",
# or "FILTERED"
CHECKIMAGE_NAME checkimg_ap.fits # Filename for the check-image

#----- Memory (change with caution!)
-----
MEMORY_OBJSTACK 2000 # number of objects in stack
MEMORY_PIXSTACK 2000000 # number of pixels in stack
MEMORY_BUFSIZE 1024 # number of lines in buffer

#----- Miscellaneous
-----
VERBOSE_TYPE NORMAL # can be "QUIET", "NORMAL" or "FULL"
#VERBOSE_TYPE QUIET # can be "QUIET", "NORMAL" or "FULL"
NTHREADS 1 # 1 single thread

RESCALE_WEIGHTS N
BACK_TYPE AUTO
```

```
BACK_VALUE 0
INTERP_TYPE VAR_ONLY
```

A.3 LePHARE configuration file

```
#####
# CREATION OF LIBRARIES FROM SEDs List #
# $LEPHAREDIR/source/sedtolib -t (S/Q/G) -c $LEPHAREDIR/config/
    cfhtlsD1-dgb-speonly-averoin.para #
# help : $LEPHAREDIR/source/sedtolib -h (or -help) #
#####
#
#----- STELLAR LIBRARY (ASCII SEDs)
STAR_SED $LEPHAREDIR/sed/STAR/STAR_MOD_ALL.list # STAR list (full
path)
STAR_FSCALE 3.432E-09      # Arbitrary Flux Scale
STAR_LIB LIB_STAR_ALL      # Bin. STAR LIBRARY ->
                            # $LEPHAREWORK/
                            lib_bin
#
#----- QSO LIBRARY (ASCII SEDs)
QSO_SED  $LEPHAREDIR/sed/QSO/QSO_MARA.list # QSO list (full path)
QSO_FSCALE 1      # Arbitrary Flux Scale
QSO_LIB  LIB_QSO_MARA      # Bin. QSO LIBRARY ->
                            # $LEPHAREWORK/
                            lib_bin
#
#----- GALAXY LIBRARY (ASCII or BINARY SEDs)
GAL_SED $LEPHAREDIR/sed/GAL/AVEROI_NEW/AVEROIN_MOD.list # GAL list
(full path)
GAL_FSCALE 1.      # Arbitrary Flux Scale
GAL_LIB  LIB_AVEROIN_HIZ      # Bin. GAL LIBRARY ->
                            # $LEPHAREWORK/
                            lib_bin
#SEL_AGE $LEPHAREDIR/sed/GAL/HYPERZ/AGE_GISSEL_ALL.dat # Age list(
full path)
                            # (def=NONE)
AGE_RANGE 0.,13.e9 # Age Min-Max in yr
#
#####
```

```

# FILTERS #
# $LEPHAREDIR/source/filter -c $LEPHAREDIR/config/cfhtlsD1-dgb.para
#
# help: $LEPHAREDIR/source/filter -h (or -help) #
#####
# Filter number and context
# G R I Z Y J H K
# 1 2 3 4 5 6 7 8 9
# 1 2 4 8 16 32 64 128 256 = 511
#
# FILTER_LIST decam/g_decam_2.pb,decam/r_decam_2.pb,decam/i_decam_2.
pb,decam/z_decam_2.pb,vista/Z_vista_2.pb,vista/Y_vista_2.pb,
vista/J_vista_2.pb,vista/H_vista_2.pb,vista/Ks_vista_2.pb
# (in $LEPHAREDIR/filt/*)

TRANS_TYPE 0    # TRANSMISSION TYPE
                # 0[-def]: Energy, 1: Nb of
                # photons

FILTER_CALIB 0 # 0[-def]: fnu=ctt
                # 1 : nu.fnu=ctt
                # 2 : fnu=nu
                # 3 : fnu=Black Body @ T=10000K
                # 4 : for MIPS (leff with nu fnu=
                #       ctt and flux with BB @ 10000K

FILTER_FILE DES_FULL_VIDEO.filt  # output name of filter's file ->
                                # $LEPHAREWORK/filt/
#
#####
# THEORETICAL MAGNITUDES #
# $LEPHAREDIR/source/mag_star -c $LEPHAREDIR/config/cfhtlsD1-dgb.
para (star only)#
# help: $LEPHAREDIR/source/mag_star -h (or -help) #
# $LEPHAREDIR/source/mag_gal -t (Q or G) -c $LEPHAREDIR/config/
cfhtlsD1-dgb.para #
# (for gal. & QSO) #
# help: $LEPHAREDIR/source/mag_gal -h (or -help) #
#####
# ----- From STELLAR LIBRARY
STAR_LIB_IN LIB_STAR_ALL # Input STELLAR LIBRARY in $LEPHAREWORK/
lib_bin/

```

```

STAR_LIB_OUT STARALL_more_ext # Output STELLAR MAGN -> $LEPHAREWORK
    /lib_mag/
#
#----- From QSO LIBRARY
QSO_LIB_IN LIB_QSO_MARA # Input QSO LIBRARY in $LEPHAREWORK/lib_bin
/
QSO_LIB_OUT QSO_MARA_more_ext # Output QSO MAGN -> $LEPHAREWORK/
    lib_mag/
#
#----- From GALAXY LIBRARY
GAL_LIB_IN LIB_AVEROIN_HIZ # Input GAL LIBRARY in $LEPHAREWORK/
    lib_bin/
GAL_LIB_OUT AVEROIN_more_ext # Output GAL LIBRARY -> $LEPHAREWORK/
    lib_mag/
#
#----- MAG + Z_STEP + EXTINCTION + COSMOLOGY
MAGTYPE AB # Magnitude type (AB or VEGA)
Z_STEP 0.04,9.,0.04 # dz, zmax, dzsup(if zmax>6)
COSMOLOGY 70,0.3,0.7 # H0,om0,lbd0 (if lbd0>0->om0+lbd0=1)

#####
things which must be changed for gal / QSO
models
MOD_EXTINC 38,62 # model range for extinction (averoin)
#MOD_EXTINC 0,0,0,0 # (QSOs)
EXTINC_LAW SMC_prevot.dat # ext. law (in $LEPHAREDIR/ext/*)
#####

EB_V 0.000,0.05,0.100,0.15,0.200,0.300,0.4,0.5,0.7,1.0,1.3,1.6,2.0
    # E(B-V) (<50 values)
EM_LINES YES
# Z_FORM 8,7,6,5,4,3 # Zformation for each SED in GAL_LIB_IN
#
#----- ASCII OUTPUT FILES OPTION
LIB_ASCII NO # Writes output in ASCII
    # in working directory
#
#####
# PHOTOMETRIC REDSHIFTS #
# $LEPHAREDIR/source/zphoto -c $LEPHAREDIR/config/cfhtlsD1-dgb para
#

```

```

# help: $LEPHAREDIR/source/zphota -h (or -help) #
#####
#
#
#
#----- Input Catalog Informations
CAT_IN input.cat # Input catalog (full path)
INP_TYPE F # Input type (F:Flux or M:MAG)
CAT_MAG AB # Input Magnitude (AB or VEGA)
CAT_FMT MEME # MEME: (Mag,Err)i
# MMEE: (Mag)i,(Err)i
CAT_LINES 1,1000000 # MIN and MAX RANGE of ROWS used in input cat [
    def:-99,-99]
CAT_TYPE LONG # Input Format (LONG,SHORT-def)
CAT_OUT output.out # Output catalog (full path)
PARA_OUT $LEPHAREDIR/config/zphot_output.para # Ouput parameter (
    full path)
#
BD_SCALE 0 # Bands used for scaling
# (Sum 2^n; n=0->nbd-1, 0[-def]:all
# bands)
GLB_CONTEXT -1 # Overwrite Context (Sum 2^n; n=0->nbd-1,
# 0 : all bands used,
# -1[-def]: used context per object
# FORB_CONTEXT -1 # context for forbitten bands
ERR_SCALE 0.01,0.01,0.01,0.01,0.01,0.01,0.01,0.01,0.01 # errors per
    band added in quadrature
ERR_FACTOR 1.0 # error scaling factor 1.0 [-def]
#
#----- Theoretical libraries
ZPHOTLIB AVEROIN_more_ext,STARALL_more_ext,QSO_MARA_more_ext #
    Library used for Chi2 (max:3)
ADD_EMLINES YES
#
##### PHOTOMETRIC REDSHIFTS OPTIONS #####
# FIR LIBRARY
FIR_LIB NONE
FIR_LMIN 7.0 # Lambda Min (micron) for FIR analysis
FIR_CONT -1
FIR_SCALE -1

```

```

FIR_FREESCALE YES # ALLOW FOR FREE SCALING
FIR_SUBSTELLAR NO
# PHYSICAL LIBRARY with Stochastic models from BC07
PHYS_LIB NONE
PHYS_CONT -1
PHYS_SCALE -1
PHYS_NMAX 100000
#
#----- Priors
# MASS_SCALE 6.,16. # Lg(Scaling) min,max [0,0-def]
MAG_ABS -10.,-28. # Mabs_min , Mabs_max [0,0-def]
MAG_REF 4 # Reference number for band used by Mag_abs
# ZFORM_MIN 5,5,5,5,5,5,3,1 # Min. Zformation per SED -> Age
# constraint
Z_RANGE 0.,99.99 # Z min-max used for the Galaxy library
EBV_RANGE 0,9 # E(B-V) MIN-MAX RANGE of E(B-V) used
NZ_PRIOR 3,1,3 # I Band for prior on N(z)
#
#----- Fixed Z (need format LONG for input Cat)
ZFIX NO # fixed z and search best model [YES,NO-def]
#
#----- Parabolic interpolation for Zbest
Z_INTERP YES # redshift interpolation [YES,NO-def]
#
#----- Analysis of normalized ML(exp-(0.5*Chi^2)) curve
#----- Secondary peak analysis
DZ_WIN 0.5 # Window search for 2nd peaks [0->5;0.25-def]
MIN_THRES 0.1 # Lower threshold for 2nd peaks[0->1; 0.1-def]
#
#----- Probability (in %) per redshift intervals
# PROB_INTZ 0,0.5,0.5,1.,1.,1.5 # even number
#
##### ABSOLUTE MAGNITUDES COMPUTATION #####
#
MABS_METHOD 2 # 0[-def] : obs->Ref
# 1 : best obs->Ref
# 2 : fixed obs->Ref
# 3 : mag from best SED
# 4 : Zbin
MABS_CONTEXT -1 # CONTEXT for Band used for MABS

```

```

MABS_REF 4 # 0[-def]: filter obs chosen for Mabs :
# ONLY USED IF MABS_METHOD=2
MABS_FILT 1,2,3,4 # Chosen filters per redshift bin (MABS_ZBIN)
# ONLY USED IF MABS_METHOD=4
MABS_ZBIN 0,0.5,1,1.5,2,3,3.5,4 # Redshift bins (even number)
# ONLY USED IF MABS_METHOD=4
##### OUTPUT SPECTRA #####
#
SPEC_OUT NO # spectrum for each object? [YES,NO-def]
CHI2_OUT NO # output file with all values : z,mod,chi2,E(B-V),...
# BE CAREFUL can take a lot of space !!
#####
# OUTPUT PDZ ANALYSIS
PDZ_OUT NONE
PDZ_OUT # pdz output file name [def-NONE]
# add automatically PDZ_OUT[.pdz/.mabsx
#/.mod/.zph]
PDZ_MABS_FILT 4,10 # MABS for REF FILTERS to be extracted
#
#####
# FAST MODE : color-space reduction #####
#
FAST_MODE NO # Fast computation [NO-def]
COL_NUM 3 # Number of colors used [3-def]
COL_SIGMA 3 # Enlarge of the obs. color-errors[3-def]
COL_SEL AND # Combination between used colors [AND/OR-def]
#
#####
# MAGNITUDE SHIFTS applied to libraries #####
#
APPLY_SYSSHIFT 0. # Apply systematic shifts in each band
# used only if number of shifts matches
# with number of filters in the library
#
#####
# ADAPTIVE METHOD using Z spectro sample #####
#
AUTO_ADAPT YES # Adapting method with spectro [NO-def]
ADAPT_BAND 3,1,3 # Reference band, band1, band2 for color
ADAPT_LIM 15,25.0 # Mag limits for spectro in Ref band [18,21.5-
def]
ADAPT_POLY 1 # Number of coef in polynom (max=4) [1-def]
ADAPT METH 1 # Fit as a function of

```

```
# 1 : Color Model [1-def]
# 2 : Redshift
# 3 : Models
ADAPT_CONTEXT -1 # Context for bands used for training
                  # -1[-def] used context per object
ADAPT_ZBIN 0.01,6 # Redshift's interval used for training
                  # [0.001,6-Def]
ADAPT_MODBIN 1,1000 # Model's interval used for training
                  # [1,1000-Def]
ERROR_ADAPT NO # [YES,NO-def]
                  # Add error in quadrature according to
                  # the difference between observed
                  # and predicted apparent magnitudes
```


Appendix B

False-colour image code

This appendix briefly describes the process of generating the false-colour images in Figure 4.7 and 4.8. These pictures have been produced by a custom code created by the author, which combines $25 \text{ pix} \times 25 \text{ pix}$ ($6.6 \text{ arcsec} \times 6.6 \text{ arcsec}$) cutouts of the DES g , r , i and z imaging into a single picture of each object. Firstly, the code colours each cutout with one of four hues (90° apart on the colour-wheel in order to span the full colour space; see Rector et al. 2007 and Rector et al. 2017). The hues are magenta-blue (270°) for the g -band, cyan (180°) for the r -band, green-yellow (90°) for the i -band and red (0°) for the z -band. In order to amplify the source signal compared to the background noise, the code caps the pixel values in each image at -3 counts and +10 counts and scales them via a square function. To brighten the colours of the final false-colour picture, each pixel is converted to a trio of RGB values, and the three RGB values are amplified¹ by a factor of 4. False-colour images are then created by averaging the four g , r , i , z input images, with the average R, G and B values computed by taking the mean of the four input R, G and B values in each pixel. If the colour-brightening amplification increased any of the R, G, or B values in a given pixel over the maximum of 255, the code scales back all three values of the RGB trio

¹This is necessary because of the specifics of combining four colours. Adding together four colours separated by 90° on the colour wheel results in an image where a pixel that is maximally bright in all four colours shows up as grey instead of white. Since this leads to dimmer colours that obfuscate the candidate sources, the pixel values are amplified to increase the brightness.

by the same amount so that the maximum value is 255 in R, G or B. This ensures that the (false) colour of the brightest pixels is preserved.