



**A census of *K*-band galaxies from the UKIDSS
Large Area Survey**

Anthony Smith

Submitted for the degree of Doctor of Philosophy
University of Sussex
September 2008

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Much of the work presented here has also been submitted for publication in the Monthly Notices of the Royal Astronomical Society ([Smith et al., 2008](#)).

Signature:

Anthony Smith

UNIVERSITY OF SUSSEX

ANTHONY SMITH, DOCTOR OF PHILOSOPHY

A CENSUS OF K -BAND GALAXIES FROM THE UKIDSS LARGE AREA SURVEY

SUMMARY

Much can be learned about the formation of galaxies by taking a census of the present-day population. As a first step towards such an endeavour, I present luminosity and surface brightness distributions of 36 659 galaxies with K -band photometry from the UKIRT Infrared Deep Sky Survey (UKIDSS) Large Area Survey (LAS) Data Release 3 and optical photometry from Data Release 5 of the Sloan Digital Sky Survey (SDSS). Various features and limitations of the new UKIDSS data are examined, such as a problem affecting Petrosian magnitudes of extended sources. Selection limits in K - and r -band magnitude, K -band surface brightness and K -band radius are included explicitly in the $1/V_{\text{max}}$ estimation of the space density and luminosity function. The bivariate brightness distribution in K -band absolute magnitude and surface brightness is presented and found to display a clear luminosity–surface brightness correlation that flattens at high luminosity and broadens at low luminosity, consistent with similar analyses at optical wavelengths. Best fitting Schechter function parameters for the K -band luminosity function are found to be $M^* - 5 \log h = -23.17 \pm 0.04$, $\alpha = -0.81 \pm 0.04$ and $\phi^* = (0.0176 \pm 0.0009)h^3 \text{ Mpc}^{-3}$, with the luminosity density in the K -band found to be $j = (6.500 \pm 0.073) \times 10^8 L_{\odot} h \text{ Mpc}^{-3}$. However, I caution that there are various known sources of incompleteness and uncertainty in my results. Using mass-to-light ratios determined from the optical colours I estimate the stellar mass function, finding good agreement with previous results. Possible improvements are discussed that could be implemented when extending this analysis to the full LAS or to future surveys.

Acknowledgements

Voulez-vous qu'on croie du bien de vous? N'en dites pas.

There are so many people I would like to thank for their support during the course of this DPhil project. But supremely I would like to thank God for kindly giving me the opportunity to do this work, and for his constant generosity, showing itself in every area of my life. *Soli Deo gloria.*

Many thanks go to Jon Loveday for his supervision of this project, for his constant willingness to answer questions and offer advice, and for his support throughout the duration of the project. Thanks also to Céline Eminian, Diana Hanbury and Bruno Henriques for many helpful group meetings. It's been a pleasure to work with Nick Cross on many aspects of this project, and I'm grateful for his advice and his numerous rapid-response emails. I would also like to thank Seb Oliver, Kathy Romer and others at Sussex for their advice on this project, and Nigel Hambly and others at the Edinburgh WFAU and Mike Irwin at the Cambridge CASU for their assistance with accessing and understanding the UKIDSS data. Thanks also to Serena Bertone for discussions and for providing model luminosity and stellar mass functions.

A huge thank you to all in the University of Sussex Astronomy Centre, and in related groups, for creating such a friendly and positive atmosphere in which to study and work. I would like to thank everyone by name, but the list would be too long, and I would certainly forget someone. You know who you are.

Thanks also to many friends in the University of Sussex Christian Union over the years, in particular those in the postgraduate group, with whom I have enjoyed many sandwiches.

I'm very grateful for the support and prayers of many friends at Calvary Evangelical Church in Brighton. Again, the list would be too long to mention everyone, but particular thanks to Phil and Maria, Chris and Katy, Steve and Brenda (for accommodating me between two of my many houses!), Christina, Mark and Rachel, and Brian, Jordan and Iain. Thanks also to the other Phil and Chris, for many breakfasts and cups of tea, and to David Brand, for many stimulating discussions. Also to the many people who have tolerated me as a housemate over the past four years: Mark, Matt, Rich, Dave, Ana, Miriam, Maggie, Joseph and Emrys.

My parents, Paul and Jenny, have been a constant source of support and encouragement, as

have the rest of my family.

I have been supported in this work by a grant from PPARC/STFC.

This work is based largely on data obtained as part of the UKIRT Infrared Deep Sky Survey.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, Cambridge University, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

This research has made use of NASA's Astrophysics Data System.

Contents

List of Tables	xi
List of Figures	xii
1 Introduction: why take a census of galaxies in the local Universe?	1
1.1 Galaxies in the universe	1
1.2 Currently favoured model for cosmology	2
1.3 Modelling galaxy formation	3
1.3.1 Dark matter structure formation	3
1.3.2 Star formation and evolution	4
1.3.3 Quasars and AGN	5
1.3.4 Galaxies: monolithic collapse or hierarchical assembly	5
1.3.5 Galaxies: secular evolution	6
1.3.6 Galaxies in clusters	6
1.4 Coherent models of galaxy formation	6
1.4.1 Selection criteria	6
1.4.2 Key observations	7
1.4.3 Overview of galaxy formation models	8
1.4.4 Summary: outstanding questions for models of galaxy formation	10
1.5 Why take a census of galaxies?	11
1.5.1 Components of a census	11
1.5.2 Applications	11
1.5.3 Analysis of census data	12
1.6 Why near-infrared?	13
1.7 Why study the space density of galaxies?	13
1.7.1 <i>K</i> -band luminosity functions	14
1.7.2 Bivariate and multivariate methods	16

1.8	Galaxy surveys: SDSS and UKIDSS	19
1.8.1	SDSS	19
1.8.2	UKIDSS	19
1.9	Overview of thesis	20
2	Data: UKIDSS and SDSS	22
2.1	Design of the imaging surveys	22
2.1.1	Observations with the various filters	22
2.1.2	Jittering and microstepping	23
2.1.3	Overlap regions: primary and secondary observations	24
2.2	Access to the data	24
2.2.1	LAS: WSA	25
2.2.2	SDSS: CAS	25
2.2.3	LAS and SDSS combined	25
2.3	Image processing	25
2.3.1	Seeing	25
2.3.2	Sky subtraction	26
2.3.3	Warning flags	26
2.3.4	Source extraction and deblending	27
2.4	Photometric galaxy properties	27
2.4.1	Galaxy profiles	27
2.4.2	Vega and AB magnitudes	28
2.4.3	Fixed-aperture magnitudes	28
2.4.4	PSF magnitudes	28
2.4.5	Petrosian magnitudes and radii	29
2.4.6	Kron magnitudes and radii	30
2.4.7	Model magnitudes	30
2.4.8	Half-light radii	31
2.4.9	Surface brightness	31
2.5	Spectroscopic galaxy properties	32
2.5.1	Spectroscopic target selection	32
2.5.2	Redshifts and classification	32
2.6	Dust	33
2.6.1	Galactic extinction	33
2.6.2	Inclination dependence	33

3	Data: completeness, contamination and reliability	35
3.1	Motivation	35
3.1.1	Possible sources of incompleteness, contamination and unreliability . . .	35
3.1.2	Modelling the magnitude uncertainties	38
3.2	Repeat observations in the LAS	40
3.2.1	Sources observed more than once in the K -band	40
3.2.2	Effect of deblending in $YJHK$	45
3.3	Comparison with deeper imaging: UKIDSS DXS	47
3.3.1	Completeness	50
3.3.2	Sources in both LAS and DXS	53
3.4	Completeness and reliability of UKIDSS quantities	58
3.4.1	Deblending	58
3.4.2	Apparent magnitudes and number counts	59
3.4.3	Size	60
3.4.4	Surface brightness	60
3.5	Completeness and reliability of SDSS quantities	62
3.5.1	Spectroscopic target selection	62
3.5.2	Apparent magnitudes	62
3.5.3	Surface brightness	62
3.5.4	Redshifts	63
3.6	Summary	63
4	Luminosity functions and beyond: estimating the properties of the underlying galaxy population	67
4.1	Estimating the area covered	67
4.2	Intrinsic galaxy properties	71
4.2.1	K - and evolution-corrections	72
4.2.2	Absolute magnitude	73
4.2.3	Physical size	74
4.2.4	Intrinsic surface brightness	74
4.3	Statistical methods for investigating the underlying population	75
4.3.1	Bayes' theorem	75
4.3.2	Modelling the underlying population	77
4.3.3	The likelihood	78
4.3.4	Model selection	78

4.3.5	Bayesian estimate of galaxy magnitudes	79
4.4	Jackknife error estimation	80
4.5	Functional forms and fits	80
4.5.1	Schechter function	80
4.5.2	Double Schechter function	82
4.5.3	Chołoniewski function	83
4.5.4	Functional fits	84
4.6	Estimating the space density of galaxies	84
4.6.1	Univariate or multivariate?	84
4.6.2	STY method	85
4.6.3	SWML method	86
4.6.4	Normalization for the STY and SWML methods	87
4.6.5	$1/V_{\max}$ method	88
4.6.6	Choice of method and implementation	90
4.7	Testing the method with simulated data	92
5	Space density in luminosity, surface brightness and stellar mass	96
5.1	K -band bivariate brightness distribution	96
5.1.1	Completeness in colour	99
5.1.2	Completeness in size	101
5.1.3	Chołoniewski fit	106
5.2	K -band luminosity function and luminosity density	107
5.3	r -band luminosity function and luminosity density	111
5.4	Large-scale structure	113
5.5	Variation with redshift	114
5.6	Subdividing the sample	115
5.6.1	Subdividing by colour	115
5.6.2	Subdividing by spectral type	118
5.6.3	Subdividing by concentration	118
5.7	Stellar mass function	120
5.8	Summary	123
6	Conclusions	124
6.1	Summary of results and interpretation	124
6.1.1	Interpreting the luminosity function	124

6.1.2	Interpreting the bivariate brightness distribution	125
6.1.3	Interpreting the bimodality	125
6.2	Measuring the evolution of the Schechter function	126
6.3	Possible improvements to UKIDSS data	127
6.4	Possible improvements to methods	129
6.4.1	Selection of the sample	129
6.4.2	Peculiar motions of nearby galaxies	129
6.4.3	Magnitude errors	129
6.4.4	Compensating for incompleteness	131
6.4.5	Compensating for large-scale structure	131
6.5	Future surveys	132
	Bibliography	133

List of Tables

1.1	Sample sizes of K -band galaxy luminosity functions	16
2.1	Post-processing error bits (ppErrBits) implemented in UKIDSS DR2	26
3.1	Number of sources detected in the overlap region of LAS and DXS	53
3.2	Classification for DXS and LAS sources in the overlap region	54
3.3	Limits set on observable quantities to find the space density	64
4.1	Limits on the visibility of galaxies for the $1/V_{\max}$ method	91

List of Figures

2.1	Response curves for <i>ugrizZYJHK</i> filters	23
3.1	Magnitude errors estimated from repeat observations in the LAS	41
3.2	Completeness as a function of magnitude, estimated through repeat observations	45
3.3	Effect of deblending on the largest UKIDSS aperture magnitudes	48
3.4	Effect of deblending on UKIDSS Petrosian magnitudes	49
3.5	Images of the same galaxies in DXS and LAS	51
3.6	Separation between closest matches for LAS and DXS in overlap region	51
3.7	Fraction of DXS sources detected in the LAS	52
3.8	KClassStat for DXS and LAS in the overlap region	55
3.9	Comparison between DXS and LAS galaxy properties in the overlap region	57
3.10	Number counts in the DXS/LAS overlap region	58
3.11	<i>K</i> -band number counts for the full DR3 LAS	59
3.12	<i>K</i> -band Petrosian magnitude and surface brightness, limited by Petrosian radius	61
3.13	<i>r</i> -band galaxy properties	65
3.14	<i>K</i> -band galaxy properties	66
4.1	Sky coverage of the sample	69
4.2	<i>r</i> -band absolute magnitudes of excluded sources	71
4.3	<i>r</i> – <i>K</i> Petrosian colour of galaxies in the sample	72
4.4	<i>K</i> - and evolution-correction	74
4.5	Schechter function fit parameters for the simulated samples	92
4.6	Chołoniewski fit to simulated data	93
4.7	Chołoniewski fit to simulated data, with one-sigma errors subtracted	95
4.8	Chołoniewski fit to simulated data, with one-sigma errors added	95
5.1	Redshift histogram and <i>K</i> -band absolute magnitudes	97
5.2	BBD for the full sample	98

5.3	BBD in $r - K$ colour and K -band absolute magnitude	100
5.4	BBD in $r - K$ colour and r -band absolute magnitude	101
5.5	Limits in absolute r -band magnitude and physical K -band Petrosian radius as a function of redshift	103
5.6	Bivariate space density of the sample in r -band absolute magnitude and K -band physical Petrosian radius	104
5.7	Bivariate space density of the sample in K -band absolute magnitude and K -band physical Petrosian radius	105
5.8	BBD with one-sigma errors subtracted	106
5.9	BBD with one-sigma errors added	107
5.10	K -band luminosity function	108
5.11	2MASS magnitudes compared with UKIDSS	110
5.12	r -band luminosity function	112
5.13	K -band LF for the NGP and SGP regions	113
5.14	K -band luminosity function for galaxies in three different redshift intervals	114
5.15	Histogram of V/V_{\max}	115
5.16	BBD for red galaxies	116
5.17	BBD for blue galaxies	117
5.18	K - and r -band LFs for red and blue galaxies	117
5.19	BBD for early spectral type	118
5.20	BBD for late spectral type	119
5.21	K - and r -band LFs for early and late spectral type	119
5.22	BBD for high-concentration galaxies	120
5.23	BBD for low-concentration galaxies	121
5.24	K - and r -band LFs for high- and low-concentration galaxies	121
5.25	Stellar mass function	122
5.26	SMF for various KCORRECT M/L ratios	123
6.1	LF with Schechter function fits to a different range of absolute magnitudes	127
6.2	LF with modelled magnitude errors	130

Chapter 1

Introduction: why take a census of galaxies in the local Universe?

*La vérité est si obscurcie en ce temps, et le mensonge si établi,
qu'à moins que d'aimer la vérité, on ne saurait la connaître.*

This chapter provides a general introduction to the subject matter of the thesis, outlining the issues to be addressed in subsequent chapters.

1.1 Galaxies in the universe

Despite their apparent irrelevance to people's day-to-day lives, galaxies still arouse a great deal of interest. This comes from an instinctive human desire to understand our own identity in relation to the rest of the cosmos. Where did we come from? Are we significant? Are we alone?

Since the early decades of the 20th Century, it has seemed plausible that the 'global' properties of the observable Universe (or simply 'the Universe') may be described by a simple mathematical model. In this **hot big bang** model, the Universe expanded over billions of years from a tiny volume of space containing matter and radiation with extremely high density and temperature. It has also become increasingly plausible that the galaxies and stars within the Universe formed as a result of tiny fluctuations in this high-density fluid. However, models for galaxy formation are much more complex than models for the Universe as a whole, since they must describe a wide range of small- and large-scale physical processes. Nonetheless, these models appear to be remarkably successful.

The currently favoured models of cosmology and galaxy formation are described in the following sections, followed by a discussion of some outstanding questions in galaxy formation. It is the goal of this thesis to use observations of galaxies to address some of those particular questions.

1.2 Currently favoured model for cosmology

Cosmology is not the focus of this thesis. However, in studying the intrinsic properties of galaxies, it is almost essential to have an underlying cosmological model, in order for it to be possible to estimate the distance to a galaxy from its redshift.

The hot big bang cosmological model was developed in the early 20th Century based on observations of the distance and recession velocity of various galaxies (e.g., [Hubble, 1929](#)), and based on the general theory of relativity by [Einstein \(1915\)](#) and cosmological models built on this theory ([Friedmann, 1922, 1924](#); [Lemaître, 1931](#); [Robertson, 1935](#); [Walker, 1935](#)). Various observations—of the relationship between redshift and distance, for galaxies and supernovae, of the galaxy power spectrum and of the cosmic microwave background (CMB)—have been used to narrow the range of possible cosmological models. It is now generally accepted that the Universe is geometrically flat, having ‘critical’ energy density, $\Omega = 1$, and that this energy density is composed of matter (dark and baryonic), another component known as ‘dark energy’ ([Frieman et al., 2008](#)), which has negative pressure and hence causes the expansion to accelerate (possibly caused by the cosmological constant, Λ), and a tiny contribution from electromagnetic radiation (negligible in this work). This is known as the Λ CDM model, since the model is dominated by the cosmological constant, Λ , or dark energy, and by cold dark matter (CDM). The energy density is expressed as $\Omega = \Omega_M + \Omega_\Lambda$, where $\Omega_M \simeq 0.30 \pm 0.04$ is the matter density and where $\Omega_\Lambda = 1 - \Omega_M$ is the dark energy density. (The value for Ω_M is that found by [Tegmark et al., 2004](#), using WMAP results for the CMB and the SDSS three-dimensional power spectrum.) Note that slight variations are found in the cosmological parameters depending on the data set(s) used ([Spergel et al., 2007](#)). An epoch of exponential expansion in the very early Universe is postulated (**inflation**), which would explain the homogeneity and flatness of the Universe while also giving rise to small inhomogeneities, which would seed the process of structure formation ([Guth, 1981](#); [Albrecht & Steinhardt, 1982](#); [Linde, 1982, 1983](#)).

Various measures of the distance can be derived for this cosmological model (see [Hogg, 1999](#)). The **comoving distance** to a galaxy at redshift z is derived from the Friedmann–Lemaître–Robertson–Walker metric and the Einstein equations to be

$$D_M = \frac{c}{H_0} \int_0^z \frac{dz}{\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}}. \quad (1.1)$$

From this, the **luminosity distance** (the apparent distance of an object of known luminosity, assuming a Euclidean universe) is given by $D_L = (1+z)D_M$ and the **angular diameter distance** (the apparent distance of an object of known size, assuming a Euclidean universe) is given by $D_A = D_M/(1+z)$. A galaxy’s apparent magnitude, m , is related to its absolute magnitude, M ,

by the **distance modulus**, $DM(z) \equiv m - M = 5 \log(D_L/10 \text{ pc})$.

Questions still remain for this standard cosmological model (see, e.g., [Turner & Tyson, 1999](#); [Peebles, 2004, 2005](#)). For the very early Universe there are significant issues to be addressed regarding inflation, quantum gravity, the matter/antimatter asymmetry and the origins of the Universe. For the post-recombination (post-CMB) Universe, the main issues are the nature of ‘dark energy’ and ‘dark matter’, or indeed whether one or both of these effects could be explained by an alternative theory of gravity (with or without massive neutrinos) or by some other means.

For ease of comparison with previous results, in this thesis a flat cosmological model with $\Omega_M = 0.3$ and $\Omega_\Lambda = 1 - \Omega_M = 0.7$ is used, with Hubble constant $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$. Where a precise value is required for the Hubble constant, $h = 0.7$ is chosen, following [Tegmark et al. \(2004\)](#), who found $h \simeq 0.70^{+0.04}_{-0.03}$. The choice of value for Ω_M has little effect on the results: the smallest value for Ω_M given by [Spergel et al. \(2007\)](#) is 0.226 ± 0.031 , which would give a galaxy at $z = 0.3$ an absolute magnitude 0.036 mag brighter than under the model used here.

1.3 Modelling galaxy formation

With its focus on dark energy and dark matter, cosmology today is concerned with the fundamental physics of the Universe, or with what must (necessarily) happen. In contrast, the study of galaxy formation seeks to understand the particular physical processes that have shaped the galaxies in the Universe, not as a direct consequence of the laws of physics, but because of the (contingent) properties of our Universe. Many different things could have happened, if the Universe had been different, but the aim is to reconstruct the sequence of events that *did* happen.

In this section, various physical processes are considered, along with their effects on the galaxy population.

1.3.1 Dark matter structure formation

Over 80 per cent of the matter in the Universe is believed to be cold dark matter, interacting only through gravity ([Tegmark et al., 2004](#); [Spergel et al., 2007](#)). The physics of dark matter is therefore very simple; the only complications are the vast (and unknown) number of dark matter particles, the time-scales involved and the highly non-linear nature of the problem.

Large N -body numerical simulations have been performed to trace the evolution of the dark matter density field, taking initial conditions from observations of the CMB combined with predictions from the Λ CDM model (including an inflationary epoch). At present the most ambitious such simulation is the Virgo Consortium’s Millennium Simulation (or Millennium Run), which ‘follows $N = 2,160^3 \cong 1.0078 \times 10^{10}$ particles from redshift $z = 127$ to the present in a cubic

region $500h^{-1}$ Mpc on a side’ (Springel et al., 2005). The result of the simulation is a complex network of dark matter clusters and filaments. Semi-analytic models (SAMs) are superimposed on this dark matter simulation in order to extract the positions and properties of galaxies.

The parameters of SAMs are generally tuned to give the best possible agreement with some particular observations, although they often give good agreement with other observations. For example, the SAM of Springel et al. (2005) ‘was not tuned to match observations of galaxy clustering’. It is therefore significant that their simulation agrees well with observations of the galaxy two-point correlation function.

Springel et al. (2006) have shown that simulations are able to reproduce not only the observed spatial distribution of galaxies, but also the absorption features in spectra of distant quasars (the Lyman- α ‘forest’) and results based on weak lensing. They state that ‘Remarkably, all three measures are consistent both with each other and with the standard model at the level that quantitative comparison is currently possible.’ This adds further support to the Λ CDM model being largely correct. However, problems remain when considering structure on smaller scales, since simulations do not reproduce the observed sub-structure within haloes or (arguably) the number of satellite galaxies observed (see Springel et al., 2006, and references therein).

With this general success of galaxy formation models in explaining the large scale structure of the Universe, the subject area has moved from asking whether the galaxies we observe *could have* formed within particular cosmological models (with implications for whether those models may be incorrect) to asking *how* galaxies did form within the well-established ‘standard’ cosmological model.

1.3.2 Star formation and evolution

The first stars (Population III) are believed to have formed at $z \approx 20$, and are expected to have been massive, and hence short-lived (Bromm & Larson, 2004; Kashlinsky et al., 2005). Subsequent populations of stars would contain metals that had been ejected from these Population III stars. Star formation occurs where the density of gas is sufficiently high and where the gas is allowed to cool. It would therefore happen first in high-density environments, such as in the centres of the largest dark matter haloes. Observations in recent years have supported this, with decreasing redshift corresponding to (1) a decrease in the size of the largest galaxies undergoing rapid star formation (‘downsizing’; Cowie et al., 1996; Juneau et al., 2005) and (2) a reversal of the environment–star formation rate relation, with most star formation at high redshift taking place in high-density environments (Poggianti et al., 2006; Sheth et al., 2006).

Star formation leaves many observational signatures. For example, blue optical colours, high

ultraviolet or far-infrared luminosity and various spectral features are all indicative of a high star formation rate (Kennicutt, 1998). As well as this instantaneous star formation rate, the star formation histories of individual galaxies may be investigated through spectroscopic observations (Heavens et al., 2004) or through studies of the stellar populations in galaxies (Freeman & Bland-Hawthorn, 2002).

There are various processes that could induce or suppress star formation during a galaxy's lifetime; these are discussed below.

1.3.3 Quasars and AGN

It is believed that quasars and AGN are seen when matter is accreting onto a black hole at the centre of a galaxy. It appears to be the case that every massive galaxy has a supermassive black hole in its core (Richstone et al., 1998), with the mass of the black hole related to the stellar mass (Magorrian et al., 1998) and velocity dispersion (Ferrarese & Merritt, 2000; Gebhardt et al., 2000) of the bulge of the galaxy. These relations present a challenge to models of galaxy formation and evolution.

Quasars are predominantly found at high redshift, $z \gtrsim 2$ (Richstone et al., 1998), and are generally found in massive dark matter haloes (Shen et al., 2007). Along with the emission from Population III stars, they are thought to have played a significant rôle in the reionization of the Universe, most of which took place at $z > 7$ (Loeb & Barkana, 2001; Spergel et al., 2007).

Croton et al. (2006) have described and modelled two distinct processes that could lead to emission from a supermassive black hole. The first is associated with merging of galaxies and their supermassive black holes, resulting in 'quasar mode' feedback and associated with bursts of star formation. The second is the result of continual and quiescent accretion of hot gas onto the supermassive black hole, leading to 'radio mode' feedback, which would heat the surrounding medium, suppressing cooling flows and thereby limiting the size of the galaxy.

1.3.4 Galaxies: monolithic collapse or hierarchical assembly

Eggen et al. (1962) have proposed a model for galaxy formation, in which spheroidal systems formed their stars early and rapidly during a process of monolithic collapse, with a disc forming subsequently due to the angular momentum of the system. In contrast, White & Frenk (1991), building on the ideas of White & Rees (1978), have proposed a model in which galaxies formed by hierarchical assembly. In models of this nature, bulges grow through merging of galaxies while most star formation occurs in galaxy discs.

There is agreement that at least some structures must have formed at high redshift and that at

least some hierarchical assembly (merging) has taken place. But the question remains as to which process is more dominant, in particular for the most massive elliptical galaxies.

1.3.5 Galaxies: secular evolution

As well as the (comparatively) rapid processes of collapse and merging, galaxies also undergo more gradual processes of secular (i.e., internal) evolution. This is described by [Kormendy & Kennicutt \(2004\)](#), who have shown how galaxy discs will evolve dynamically to produce ‘pseudobulges’, which are similar in appearance to classical bulges. However, they argue that it is generally possible to distinguish between the two kinds of bulge, e.g., through their shapes, brightness profiles, star formation rates and the motions and velocity dispersions of their stars (see also [Drory & Fisher, 2007](#)).

1.3.6 Galaxies in clusters

Many processes can influence galaxy evolution in clusters, where galaxies have significant interactions with each other and with the intra-cluster medium. [Driver et al. \(2006\)](#) have listed some of these, for example, ram-pressure stripping ([Gunn & Gott, 1972](#)), whereby galaxies falling into a cluster will be stripped of their gas, thus suppressing star formation and changing the colour, profile, luminosity and observed bulge-to-total fraction of the galaxy ([Driver et al., 2008](#)).

1.4 Coherent models of galaxy formation

Having reviewed the various processes that can and do influence the formation and evolution of galaxies, the task now is to synthesize these into a coherent framework. The challenge is to assess *which* processes have been most significant, and *when* they have been operating in cosmic history.

1.4.1 Selection criteria

Immediately we are faced with the problem of selection criteria: how is one to decide between two models, both of which attempt to describe the properties not only of one galaxy, but of all the galaxies in the Universe?

This is not an easy matter to address. However, it is worth noting that the problem exists, and that judgements about the relative merits of competing models are often largely (though not entirely) subjective, or made on the basis of the ability of the model to reproduce one or two particular observations, such as the local K -band luminosity function, as if these were the only observations that needed to be explained.

1.4.2 Key observations

Models for galaxy formation have tended to seek to explain certain particular observations. Some of these are as follows:

1. The galaxy luminosity function. An attempt has generally been made to simulate the space density of galaxies per volume of space as a function of galaxy luminosity (generally in the K -band) or, alternatively, as a function of galaxy stellar mass. The galaxy luminosity function may be approximated by a Schechter function (see Section 4.5.1), having an exponential cut-off at high luminosity and an almost flat power-law slope at low luminosity (Binggeli et al., 1988; Benson et al., 2003).
2. The bimodality of the galaxy population. There appear (broadly speaking) to be two distinct populations of galaxies: passive, concentrated, red, luminous, elliptical galaxies, and star-forming, less concentrated, blue, faint, spiral galaxies, with the bimodality seen in various parameters (e.g., Driver et al., 2006; Ball et al., 2006). The red population tends to have a narrow range of colours, and is termed the ‘red sequence’ (Bell et al., 2004). The blue population, on the other hand, covers a wider range of colours, and is termed the ‘blue cloud’. This bimodality may reflect the two-component nature of galaxies, being composed of bulges and discs (Driver et al., 2006).
3. Merging of galaxies. There is much evidence for galaxies interacting and merging (e.g., Bell et al., 2006a,b; Bundy et al., 2007; Lin et al., 2008). Models of galaxy formation generally include galaxy-galaxy interactions, which may be used to explain quasar activity, the bimodality, the properties of bulges and discs and the build-up of the red sequence. Mergers are known as wet, mixed or dry depending on whether both, one or neither galaxy contains gas, respectively (Lin et al., 2008).
4. Dependence of galaxy properties on environment. Galaxies are seen to exhibit a relation between morphology and density, whereby the fraction of galaxies with elliptical morphology increases as the density of galaxies increases, and a colour-density relation, whereby the fraction of galaxies with red optical colour increases with higher density. It has been argued that these two relations are independent, reflecting a change with environment in the fraction of spiral galaxies that are red and the fraction of elliptical galaxies that are blue (Bamford et al., 2008). The star formation rate of galaxies has also been found to correlate strongly with environment (Balogh et al., 2004).
5. Properties of the supermassive black hole (see Section 1.3.3). Galaxy formation models

generally include the central supermassive black hole as a component, relating its properties to those of the surrounding galaxy (or galaxy bulge), and including quasar and AGN activity in some galaxies.

1.4.3 Overview of galaxy formation models

In this section I survey a range of coherent (or semi-coherent) galaxy formation models, giving a brief description of each, and highlighting the most significant areas of uncertainty. Features of these models are generally inspired by observations, but simulations are valuable in demonstrating that these components will have the anticipated effect. I therefore consider simulations first of all, to describe those components that appear to be fairly well established, before considering further observation-inspired features of galaxy formation models that may not yet have been incorporated into simulations.

Simulations

Various semi-analytic models have been implemented to incorporate baryons into the dark matter N -body Millennium Simulation (Springel et al., 2005). One ‘family’ of such semi-analytic models has been created by the same group behind the Millennium Run itself (MPA, Munich). The simulations adopt the hierarchical clustering approach of White & Frenk (1991) while the semi-analytic models are based closely on the ‘unified model for the evolution of galaxies and quasars’ of Kauffmann & Haehnelt (2000), with the assumption ‘that supermassive black holes are formed and fuelled during major mergers’. Croton et al. (2006) have described a semi-analytic ‘model for the formation and evolution of galaxies and their central supermassive black holes’ incorporating gas infall and cooling and feedback from supernovae and active galactic nuclei. The supernova feedback would produce galactic winds, which would drive the gas out of small galaxies, thus limiting the star formation and flattening the faint end of the luminosity function. Significantly, the AGN feedback is modelled in two forms, one of which is induced by merging of galaxies. They find that their model ‘can simultaneously explain: (i) the low observed mass drop-out rate in cooling flows; (ii) the exponential cut-off at the bright end of the galaxy luminosity function; and (iii) the fact that the most massive galaxies tend to be bulge-dominated systems in clusters and to contain systematically older stars than lower mass galaxies’. Using a different IMF and a new parametrization for dust attenuation, De Lucia & Blaizot (2007) have refined the model of Croton et al. (2006) and investigated the formation history of brightest cluster galaxies (BCGs). Another modified version of the Croton et al. (2006) model is presented by Bertone et al. (2007), who have incorporated a new treatment of galactic winds and feedback and found good agreement

with observed luminosity and stellar mass functions, including a lower number density of dwarf galaxies than in previous models.

Another suite of semi-analytic models using the Millennium Run has emanated from the Durham group. Again, using the principles of [Kauffmann & Haehnelt \(2000\)](#), [Bower et al. \(2006\)](#) have presented an implementation of this GALFORM galaxy formation model ‘in which feedback due to active galactic nuclei (AGN) is assumed to quench cooling flows in massive haloes’. They find good agreement with observations of the evolution of the luminosity and stellar mass functions and observations of the star formation history of the Universe. Their model ‘predicts a substantial population of massive galaxies out to redshift $z \sim 5$ ’, thus demonstrating that the apparently anti-hierarchical observations of massive galaxies at high redshift (related to downsizing) can fit comfortably into a hierarchical CDM model.

[Hopkins et al. \(2008a,b\)](#) have constructed a model for the merger-driven co-evolution of quasars, supermassive black holes and elliptical galaxies. Rather than constructing a full semi-analytic model, they use a halo occupation model to estimate the merger rate as a function of redshift between galaxies of various masses. The basic assumptions are ‘that major, gas-rich mergers cause quasar activity’ ([Hopkins et al., 2008a](#)) and ‘that star formation is quenched after a gas-rich, spheroid-forming major merger’ ([Hopkins et al., 2008b](#)). They ‘demonstrate that this model naturally reproduces the observed rise and fall of the quasar luminosity density at $z = 0 - 6$, as well as quasar luminosity functions, fractions, host galaxy colors, and clustering as a function of redshift and luminosity’ ([Hopkins et al., 2008a](#)), finding that the data favour this merger-driven model over ‘a secular model in which quasar activity is driven by bars or other disk instabilities’. Many observations related to red ellipticals follow naturally from the model of [Hopkins et al. \(2008b\)](#), such as ‘the turnover in the efficiency of star formation and baryon fractions in galaxies at $\sim L^*$ (without any parameters tuned to this value), as well as the observed mass functions and mass density of red galaxies as a function of redshift, the formation times of early-type galaxies as a function of mass, and the fraction of quenched galaxies as a function of galaxy and halo mass, environment, and redshift’. They find that various observations favour a merger-driven model of the quenching of star formation over models in which quenching is driven by halo mass (e.g., [Croton et al., 2006](#)) or by secular/disc instabilities (e.g., [Bower et al., 2006](#)).

In spite of these successes, the current approach to simulating galaxy formation has been criticized. One area of disagreement with observations is the ‘satellite problem’, which arises from the fact that Λ CDM models predict that galaxies like the Milky Way would be surrounded by a large number of small dark matter haloes, containing dwarf satellite galaxies, which are not observed. Various forms of feedback are invoked as an explanation for this, to prevent galaxies from

forming in these small dark matter haloes. However, [Gilmore et al. \(2007\)](#) have investigated the small-scale structure of dark matter, suggesting that ‘dark matter forms cored mass distributions, with a core scale length of greater than about 100 pc’. They note that numerical simulations of the cosmological evolution of dark matter have thus far been unable to resolve the necessary small scales, and are therefore likely to be very unreliable at simulating small-scale structure. But many of the features of semi-analytic models (such as feedback) are incorporated precisely to deal with the discrepancy on small scales between simulations and observations ([Gilmore, 2008](#)). One possible modification to the underlying Λ CDM model is to consider warm dark matter, which would lead to a higher minimum mass for dark matter haloes than with cold dark matter ([Strigari et al., 2008](#)).

Observations

[Freeman & Bland-Hawthorn \(2002\)](#) have presented a detailed overview of the properties of the Milky Way Galaxy in an attempt to uncover the origins of the Galaxy through study of the components of the Galaxy, the stellar dynamics and the stellar metallicities. Their model for the evolution of the Galaxy has many stars forming very early on in the halo and in the stellar bulge (which developed a supermassive black hole), with many small satellite galaxies being accreted by the Galaxy, leaving their remnants as halo stars and globular clusters, and with the discs (thick and thin) forming at a later stage. Major mergers of galaxies are not discussed; studies of our Galaxy are unlikely to tell us about the formation of large classical bulges or elliptical galaxies.

[Driver et al. \(2006\)](#) have argued that the galaxy bimodality reflects primarily the two components of galaxies (red bulges and blue discs) rather than two distinct classes of galaxies (red and blue). The model they propose for galaxy formation has ‘early bulge formation through initial collapse’, with a formation peak at $z > 3$, and discs forming later through various mechanisms. This is in contrast to the hierarchical model for the build-up of the red sequence, which sees galaxy bulges and elliptical galaxies being assembled throughout cosmic history through major mergers ([Bell et al., 2004, 2006a](#); [Drory & Fisher, 2007](#)).

1.4.4 Summary: outstanding questions for models of galaxy formation

It is clear that one of the major unresolved issues in studies of galaxy formation is the place and effect of galaxy mergers on the formation of the present galaxy population. ‘Major mergers of comparable mass haloes and comparable mass galaxies play an important role in many galaxy formation models. Such mergers are usually invoked to explain the formation of galactic bulges and elliptical galaxies’ ([Cole et al., 2008](#)). The merger rate itself appears to be consistent with

this (Bell et al., 2006a,b; Lin et al., 2008), although whether major mergers are responsible for the mass assembly of spheroidal galaxies is less clear (Bundy et al., 2007).

Another significant area of uncertainty is related to dwarf satellite galaxies, as has been discussed above.

1.5 Why take a census of galaxies?

Having established some of the specific questions facing galaxy formation models, I now turn to consider how observations may be used to give insight into these issues.

When considering the build-up of massive galaxies and the rôle played by satellite galaxies, it is obvious that detailed studies of only a handful of galaxies will not be sufficient. What is needed is a survey, or census, of a large sample of the galaxy population.

1.5.1 Components of a census

In a national census, participants are required to answer many different questions, for example, about their age, birthplace, marital status, children and employment. A census of galaxies is similar. The aim is to obtain not simply one property of each galaxy (for example, its absolute magnitude), but to quantify a large number of different properties. These may include measurements of the flux over a very wide range in wavelength, spectra, including redshifts, size, surface brightness, structure, morphology and various measurements of the environment of the galaxy. These data will come from a combination of several galaxy surveys, taken at different ground-based or space-based observatories.

1.5.2 Applications

Considering the questions from the previous section, what would a census look like in practice?

When investigating dwarf satellite galaxies, the census would need to have deep photometry, in order to find low-surface brightness galaxies. It would ideally cover the whole sky, in order to get a reasonably complete sample of nearby dwarf galaxies. The resolution of the images would need to be high, in order to resolve individual stars in the nearest galaxies. Stellar dynamics could be studied through spectroscopy of these stellar populations, leading to an estimate of the total mass of the galaxies (including dark matter). And the wavelength range would be chosen to detect the various components of the galaxies: stars, dust and neutral hydrogen.

When studying the mass assembly of galaxies, it is important to survey a representative sample at various epochs. In order to do this, a so-called ‘wedding cake’ survey strategy is often chosen,

in which several surveys of various areas and depths are combined. This is because in order to survey a large volume at high redshift, the most efficient strategy is to survey a small area to a great depth, rather than a large area to a shallow depth. The various surveys should be chosen to detect the various types and components of the galaxies, although dwarf galaxies will be visible only at low redshift. The imaging should be of sufficient depth to make it possible to study the morphology and structure of the galaxies, as well as their magnitudes and colours. Spectroscopy, where available, is of great value in giving information about the distance to the galaxy and the properties of the stellar populations that make up the galaxy.

The focus in this thesis is on surveys at low redshift. Some advantages of working at low redshift ($z \sim 0.1$) are as follows:

1. A more complete sample may be studied, including galaxies with low luminosity or low surface brightness. High-redshift surveys are unable to detect very low-surface brightness galaxies, since surface brightness dims as $(1 + z)^4$. Low-surface brightness galaxies tend to be faint galaxies, so the high-redshift surveys are only able to see the more luminous galaxies.
2. Nearby or large galaxies may be studied in more depth, investigating morphology and structure as well as luminosity and colour.
3. The evolution and selection effects that plague high-redshift surveys are less of a problem. For example, numerous sources of bias creep in to surveys as a function of redshift, and it can be difficult to account for all of these. An additional point is that it is possible to learn much about a population's history by taking a census of the present-day population.
4. High-redshift surveys have to rely largely on photometric redshifts, since it is very inefficient and expensive to obtain spectroscopic redshifts for high-redshift galaxies. This introduces yet more uncertainties into the analysis.

Low-redshift surveys are also important in establishing a baseline for comparison with results at higher redshift (e.g., [Cirasuolo et al., 2007](#)).

1.5.3 Analysis of census data

Once the data from a census have been collected, how can they be analysed?

The significant difference between a census of people and census of galaxies is that one can be fairly sure about the completeness of the census of people; only a tiny proportion will not complete the census. However, with galaxies, there is no way of knowing for sure how complete the sample

is. There are many reasons why a galaxy might not be included in the survey: too faint, too small, too large, too bright, obscured by foreground stars or galaxies. It is absolutely vital to take into account the galaxies that cannot be seen when analysing the data.

After correcting for incompleteness, much can be learned by investigating how the different galaxy properties are related to each other and finding any sub-populations that exist. This can give clues for reconstructing the history of galaxy formation and evolution.

1.6 Why near-infrared?

The advantages (in principle) of building such a census on near-infrared (NIR) observations are well known. First, mass-to-light ratios in the near-infrared are largely insensitive to galaxy or stellar type, certainly much less than in the optical (Bell & de Jong, 2001). This means that the near-infrared light is a good tracer of the total stellar mass in a galaxy. Moreover, the range of mass-to-light ratios is much smaller in the near-infrared, so uncertainties in the stellar mass are much smaller. Not only does this mean that a survey limited in near-infrared magnitude will be approximately limited in ‘apparent stellar mass’, but also that morphological measurements in the near-infrared, for example the Sérsic index and the half-light radius, will reflect the distribution of stellar mass within the galaxy, whereas such measures in the optical will be significantly biased by the presence of young stellar populations.

A second advantage is that the K -corrections in the K -band are also relatively independent of galaxy type (Mannucci et al., 2001), leading to smaller uncertainties in the absolute magnitudes.

A third advantage is that dust is much less of a problem in the near-infrared than in the optical. This means that, whereas optical measurements of galaxy properties are affected by dust obscuration, and therefore strongly dependent on the inclination of the galaxy, producing a smoothing of the galaxy luminosity function, this is not such a problem in the near-infrared (Driver et al., 2007b; Maller et al., 2008). This will be discussed further in Section 2.6.

However, the main disadvantage of the near-infrared (for ground-based telescopes) is the sky brightness, which is around $13.5 \text{ mag arcsec}^{-2}$ in K for the data used here (Dye et al., 2006).

1.7 Why study the space density of galaxies?

The space density of galaxies, commonly denoted by ϕ , is the number of galaxies of a certain type, per unit volume. This is an essential component in any census of galaxies. The number of galaxies of a certain type detected in a survey is not a particularly useful statistic, since this is strongly dependent on the luminosity, distance and type of galaxy. However, once the statistics have been

corrected for incompleteness, a number density may be obtained.

The most common measure of the space density is the **luminosity function** (Binggeli et al., 1988), which is the number of galaxies per unit volume as a function of galaxy luminosity, in some particular waveband. Techniques for estimating the luminosity function (LF) are described in Section 4.6; these include the binned $1/V_{\max}$ estimator (Schmidt, 1968; Rowan-Robinson, 1969), the parameterized STY maximum likelihood method (Sandage et al., 1979) and the binned stepwise maximum likelihood method (SWML; Efstathiou et al., 1988). The LF is often described by a Schechter function (Schechter, 1976); see Section 4.5.

1.7.1 *K*-band luminosity functions

Given the close relationship between the stellar mass and the *K*-band light, the *K*-band galaxy luminosity function is a convenient quantity for numerical or semi-analytic models to predict (e.g., Croton et al., 2006; Bower et al., 2006; De Lucia & Blaizot, 2007; Bertone et al., 2007).

There have been several studies of the low-redshift *K*-band luminosity function in recent years.

Mobasher et al. (1993) used a (small) sample from the optically-selected Anglo-Australian Redshift Survey, combined with *J*-, *H*- and *K*-band UKIRT imaging, to find the luminosity function for field galaxies.

Glazebrook et al. (1995) used a *K*-band imaging survey and a subsequent redshift survey to examine galaxy evolution in the near infrared. Their small-area, deep survey enabled them to probe the *K*-band LF to redshift $z > 0.5$, where they found evidence of evolution, with galaxies being more luminous at higher redshift.

Gardner et al. (1997) found the LF using *K*-band imaging from Kitt Peak National Observatory with redshifts for the *K*-band selected sample obtained at the William Herschel Telescope on La Palma.

Szokoly et al. (1998) used a small sample, with redshifts from an optically-selected sample, to find the *K*-band LF. Their sample covers a small area (0.6 deg^2) but is deep, complete to $K = 16.5$.

Loveday (2000) used the optically-selected Stromlo–APM redshift survey, with CTIO *K*-band imaging, to find the *K*-band luminosity function of nearby field galaxies. Low-luminosities were investigated (fainter than $M_K - 5 \log h = -20$), even with a small sample, by selecting galaxies for *K*-band photometry based on their b_J -band absolute magnitude. The combined limits in b_J and *K* were taken into account by estimating a bivariate luminosity function (see below).

With the Two Micron All Sky Survey (2MASS; Jarrett et al., 2000), an all-sky near-infrared survey became publicly available, providing various opportunities to investigate the local *K*-band

luminosity function.

[Kochanek et al. \(2001\)](#) selected galaxies based on their 2MASS magnitudes and found redshifts for these galaxies, most of which had already been measured. With a large sample it is safe to subdivide, which they did according to morphological type (classified by eye), finding that early-type galaxies tend to be brighter.

[Cole et al. \(2001\)](#) used 2MASS imaging and the b_J -band selected 2dF Galaxy Redshift Survey (2dFGRS) to find the LF with a larger sample than previous estimates. The sample size is such that systematic uncertainties dominate over statistical uncertainties, and these were investigated. The local stellar mass function and the star formation history of the Universe were also estimated.

[Huang et al. \(2003\)](#) used the Hawaii+AAO K -band redshift survey, which is K -band selected and has a much smaller area and greater depth than 2MASS-based LF estimates. They found a faint-end slope significantly steeper than found by most other authors.

[Bell et al. \(2003b\)](#) combined 2MASS imaging with SDSS redshifts to find the luminosity and stellar mass functions. They argued that ‘2MASS is biased against low surface brightness galaxies’ and estimate a ‘true’ LF to account for this incompleteness.

[Eke et al. \(2005\)](#) built on the work of [Cole et al. \(2001\)](#) ‘to study the near-infrared light and stellar mass content of the local Universe’, particularly how the stellar mass is distributed according to group size.

[Jones et al. \(2006\)](#) used the largest K -band sample to-date for their 6-degree Field Galaxy Survey (6dFGS) LF, based on 2MASS imaging with a K -band selected redshift survey. This is a shallow survey that covers a very large area (9075 deg² effective area in K). Corrections are applied based on the peculiar velocities of nearby galaxies in order to avoid biased estimates of the distance derived from the redshift. They find little disagreement between results obtained with the SWML and $1/V_{\max}$ methods, as expected for a large sample (the $1/V_{\max}$ method is sensitive to fluctuations in the galaxy number density with redshift). Comparing with previous results, they find general agreement but are able to extend their analysis to a significantly larger range in absolute magnitude ($-17.5 > M_K - 5 \log h > -27.5$).

Table 1.1 shows the sample size of some previous K -band LF estimates.

The principal uncertainty remaining is connected with the low-luminosity end of the luminosity function. Moreover, there has been significant discussion about possible low-surface brightness incompleteness in 2MASS ([Andreon, 2002](#)), which would affect both the low-luminosity end of the luminosity function and the recovered magnitudes for galaxies with low-surface brightness outer regions. [Kirby et al. \(2008\)](#) have found through deep NIR imaging of nearby galaxies that 2MASS would underestimate the flux of low-surface brightness galaxies by around 2.5 mag.

Table 1.1: Sample sizes of K -band galaxy luminosity functions.

Paper	Number of galaxies in sample
Mobasher et al. (1993)	181
Glazebrook et al. (1995)	124
Gardner et al. (1997)	510
Szokoly et al. (1998)	175
Loveday (2000)	345
Kochanek et al. (2001)	3878
Cole et al. (2001)	5683
Huang et al. (2003)	1056
Bell et al. (2003b)	6282
Eke et al. (2005)	15 644
Jones et al. (2006)	60 869
This work	36 659

There is also disagreement about the very bright end, where surveys detect only a small number of galaxies. This will be discussed in Section 5.2.

1.7.2 Bivariate and multivariate methods

There has been a recognition that describing a galaxy by a single number (luminosity) can yield only a limited amount of information. (This is often compounded by expressing the properties of the entire galaxy population by three numbers, viz., the parameters of the Schechter function.) In order to more fully exploit the multiband and multivariate data that are becoming increasingly available, with their complex visibility limits, the traditional techniques for measuring the luminosity function need to be extended to deal with several quantities at once. The generalisation of the LF is to find the space density of galaxies with certain values of a set of properties, for example, luminosity in various bands, surface brightness, environmental density and concentration.

Much work has been done in recent years doing this kind of analysis with two galaxy properties. Most often this is the space density of galaxies as a function of luminosity and surface brightness, known as the **bivariate brightness distribution** (BBD).

[Chołoniewski \(1985\)](#) used a maximum likelihood technique to find the space density according to galaxy luminosity and radius, introducing an analytical form (the **Chołoniewski function**, see Section 4.5.3) to describe such a distribution. A sample with magnitude-dependent selection

effects only was chosen to demonstrate the method.

[Phillipps & Disney \(1986\)](#) estimated the BBD for spiral galaxies in the Virgo cluster. They argued for the importance of surface brightness, both in limiting the visibility of galaxies and as a characteristic in its own right. Considering selection limits in luminosity, surface brightness and size, they found that there are regions of the BBD diagram that may be well populated by galaxies, but that these galaxies would not be visible in that particular survey. These ideas were developed by [Phillipps et al. \(1990\)](#) who considered how the selection effects vary with redshift. If these are neglected, false conclusions may be drawn about the redshift-evolution of the galaxy population.

[Sodré & Lahav \(1993\)](#) used modified forms of the STY and SWML methods of estimating the luminosity function of galaxies to find the space density bivariate in luminosity and galaxy size. For the STY method, they used the Chołoniewski function, but assuming an exponential distribution in diameter and a Gaussian distribution in magnitude, as opposed to a Schechter function in magnitude and a Gaussian distribution in diameter ([Chołoniewski, 1985](#)).

[Impey & Bothun \(1997\)](#) reviewed the properties of low surface brightness galaxies, arguing that ‘the local galaxy luminosity function cannot be derived without correcting for surface brightness selection effects’.

[Petrosian \(1998\)](#) has examined the danger of performing cosmological tests by assuming (incorrectly) that a sample of galaxies is limited merely by magnitude. In reality, the surface brightness profile of a galaxy has a strong effect on whether it can be detected, and on its observed magnitude.

[Driver \(1999\)](#) investigated the *B*-band bivariate brightness distribution (BBD) of galaxies within a volume-limited sample in the Hubble Deep Field. This was done in order to quantify ‘the contribution of normal, dim, and dwarf galaxies to the local luminosity density’. The luminosity density and mass density were found to be dominated by ‘normal’ galaxies, i.e., those on the Hubble tuning-fork diagram. A clear luminosity–surface brightness correlation was also found.

[Loveday \(2000\)](#), although investigating the LF rather than the BBD, used the bivariate form of the SWML method (see [Sodré & Lahav, 1993](#)) to find the bivariate *K*- and *B*-band luminosity function. This was done as a means of estimating the *K*-band luminosity function from data limited in both *K*- and *B*-band apparent magnitude.

[de Jong & Lacey \(2000\)](#) used the $1/V_{\max}$ method to find the space density of spiral galaxies bivariate (1) in effective surface brightness and effective radius, and (2) in absolute magnitude and effective radius. The latter was fit by a Chołoniewski function. They made a connection between the size–luminosity relation and the angular momentum of galaxy discs ‘being proportional to that of the parent halo’.

Blanton et al. (2001) used the SWML method of Sodr  & Lahav (1993) to find the joint distribution of r -band absolute magnitude bivariate with r -band effective surface brightness in the SDSS commissioning data, taking into account limits in both magnitude and surface brightness. From this BBD they argued that their estimate of the luminosity density is reliable, since the contribution of low-surface brightness galaxies to the total luminosity density is likely to be very small. They also found the bivariate space density in luminosity and colour, and in luminosity and concentration.

Again at optical wavelengths (b_J -band), the BBD has been investigated by Cross et al. (2001), using data from the 2dFGRS. The method employed was a bivariate form of the $1/V_{\max}$ estimator, correcting for redshift incompleteness and clustering. They found that the surface brightness limit adopted has a strong effect on the faint-end slope of the measured luminosity function and on the measurement of the total luminosity density.

This work was built on by Cross & Driver (2002), who fit a Chołoniewski function to the BBD data and further investigated the effect of surface brightness limits on the estimated faint-end slope of the luminosity function.

Blanton et al. (2003c) used the $1/V_{\max}$ estimator, accounting for limits in flux, surface brightness and redshift, to find the bivariate space density of SDSS galaxies. With each galaxy weighted by its value of $1/V_{\max}$, they found the bivariate space density in pairs of seven galaxy properties: absolute $^{0.1}i$ -band magnitude, four optical colours, surface brightness and concentration (measured by the S rsic index, see Section 2.4.1).

A bivariate SWML method was used to find the BBD in B -band luminosity and surface brightness in the MGC by Driver et al. (2005). They fit the BBD with a Chołoniewski function and found it to be a poor fit. Driver et al. (2006) developed this analysis by projecting the BBD onto other axes to investigate the joint space density distribution in various pairs of parameters: colour, S rsic index, central surface brightness and physical radius in addition to luminosity and effective surface brightness. Galaxy bimodality is seen consistently, particularly in the colour-S rsic index plane.

Ball et al. (2006) have used the same bivariate SWML method of Blanton et al. (2001) to investigate the space density of SDSS galaxies bivariate in luminosity and each of the following quantities: morphological type, inverse concentration index, S rsic index, absolute effective surface brightness, reference frame colours, physical radius, eClass spectral type, stellar mass and galaxy environment. Their results consistently show a bimodality of the galaxy population.

Cameron & Driver (2007) have investigated the luminosity–size distribution for galaxies in the Hubble Ultra Deep Field. Through comparison with local results from the MGC, they find evidence for an evolution in surface brightness, with a higher mean surface brightness at higher

redshift ($z \sim 0.7$ compared with $z \sim 0.1$).

In summary, the bivariate brightness distribution, and other bivariate distributions, have been studied widely at optical wavelengths and are beginning to show their potential for giving insight to galaxy formation models. At near-infrared wavelengths, which are more suitable for studying the distribution of stellar mass within galaxies, the luminosity function has been estimated by various authors, but the BBD has not yet been found.

1.8 Galaxy surveys: SDSS and UKIDSS

This thesis uses data from SDSS and UKIDSS. The general design of these surveys, and their suitability to the aims of this thesis, will be introduced in this section; relevant technical details will be discussed in subsequent chapters.

1.8.1 SDSS

Like all large-scale astronomical projects, the Sloan Digital Sky Survey (SDSS) was designed to give a high scientific return from a sizeable investment of financial, technical and human resources. A summary of the survey properties has been provided by [York et al. \(2000\)](#). The most significant technological advances that made SDSS possible were the sensitive CCD detectors and computer power to handle large volumes of data. With these, it became possible to conduct a survey of the local Universe that would be significantly deeper than any previous large-area surveys. A combined imaging and spectroscopic survey was designed, with a new purpose-built 2.5 m telescope placed at the Apache Point Observatory, New Mexico, to survey approximately a quarter of the sky, almost entirely in the north Galactic cap, in five optical bands.

All of the galaxies used in this thesis are drawn from the main galaxy sample of Data Release 5 of the SDSS (DR5; [Adelman-McCarthy et al., 2007](#)), from which the optical photometry and spectroscopic redshifts used in the analysis below are obtained. As of DR5, the survey had obtained ‘five-band photometric data for 217 million objects selected over 8000 deg² and 1 048 960 spectra of galaxies, quasars, and stars selected from 5713 deg² of that imaging data’.

1.8.2 UKIDSS

The UKIRT Infrared Deep Sky Survey (UKIDSS), as defined by [Lawrence et al. \(2007\)](#), was designed to exploit the new Wide-Field CAMera (WFCAM; [Casali et al., 2007](#)) on the existing UK InfraRed Telescope (UKIRT) on Mauna Kea, Hawaii. The WFCAM near-infrared photometric system is described by [Hewett et al. \(2006\)](#). From Hawaii the data are sent to Cambridge for

pipeline processing (Irwin et al., in preparation) and then to Edinburgh to be ingested into the WFCAM Science Archive (WSA; [Hambly et al., 2008](#)).

UKIDSS is in fact five surveys, covering a range of depths, areas and targets using various combinations of the *ZYJHK* filters. Three of the surveys are optimized for extragalactic observations: the Large Area Survey (LAS), Deep eXtragalactic Survey (DXS) and the Ultra-Deep Survey (UDS), while the remaining surveys are designed for Galactic targets: the Galactic Plane Survey (GPS) and the Galactic Clusters Survey (GCS).

This thesis will use data from the Large Area Survey, which is intended to cover 4000 square degrees in the northern Galactic cap. Unlike SDSS, UKIDSS is an imaging survey only. This has influenced the design of the LAS so that it lies entirely within the field of SDSS, making optical photometry and redshifts readily available.

Data used in subsequent chapters are taken from UKIDSS Data Release 3 (Warren et al., in preparation), which was released in December 2007, with the LAS containing coverage in *YJHK*, including 1189 square degrees of coverage in *K* to a 5σ depth of 18.2 mag, in both north and south Galactic poles (hereafter NGP and SGP respectively).

1.9 Overview of thesis

The aim of this thesis is to present a simple census of low-redshift *K*-band galaxies by finding the space density (BBD and LF) of a large sample of galaxies. This is a necessary first step towards constructing a more comprehensive census of galaxy properties, using near-infrared data as a central component.

Chapter 2 introduces the data from SDSS and UKIDSS.

Chapter 3 contains an analysis of the completeness and reliability of the recent UKIDSS LAS data.

Chapter 4 lays down the statistical foundations for using observations of galaxies to estimate the intrinsic properties of the galaxy population. In this chapter the methods used are presented, developed and evaluated. Some of the standard methods of quantifying the distribution are introduced, such as the luminosity function and the bivariate brightness distribution.

In Chapter 5 the *K*-band luminosity function and bivariate brightness distribution of galaxies are presented, along with other analyses of the data. To my knowledge, this represents the first such analysis of the UKIDSS LAS data and the first BBD at near-infrared wavelengths. The stellar mass function is also estimated.

Chapter 6 contains the conclusions of the thesis, along with a discussion about possible improvements and extensions to the work presented here.

AB magnitudes are used throughout for SDSS magnitudes and Vega magnitudes for K -band quantities. For reference, AB and Vega magnitudes are related in the r -band by $r_{\text{AB}} = r_{\text{Vega}} + 0.146$ and in the K -band by $K_{\text{AB}} = K_{\text{Vega}} + 1.900$ ([Hewett et al., 2006](#)).

Chapter 2

Data: UKIDSS and SDSS

*Combien les lunettes nous ont-elles découvert d'êtres
qui n'étaient point pour nos philosophes d'auparavant!*

In this chapter various catalogue and derived quantities obtained from SDSS and UKIDSS are described. In the following chapter, the reliability of these quantities and the completeness of the sample will be investigated.

2.1 Design of the imaging surveys

This section contains a description of implementation of the photometric observations in SDSS and UKIDSS, with attention drawn to any relevant points.

2.1.1 Observations with the various filters

Fig. 2.1 shows the transmission curves for the five SDSS filters (*ugriz*) and the five UKIDSS filters (*ZYJHK*).

The UKIDSS *Z*-band filter has a very similar wavelength range to the SDSS *z*-band filter, so this filter is not used in the UKIDSS Large Area Survey, which is fully contained within the SDSS field.

The SDSS camera contains 30 CCD chips in an array of six columns and five rows. Each row is covered by one of the five *ugriz* filters. Observations are performed in a drift-scan mode, with all five filters being used at all times, and with the camera tracing out large strips over great circles on the sky. The rows of CCD chips are positioned with a gap between them, so a second scan is made of each strip, offset by 93 per cent of the CCD width, in order to fill in the gaps and produce a **stripe**. This produces a small amount of overlap between the strips (York et al., 2000).

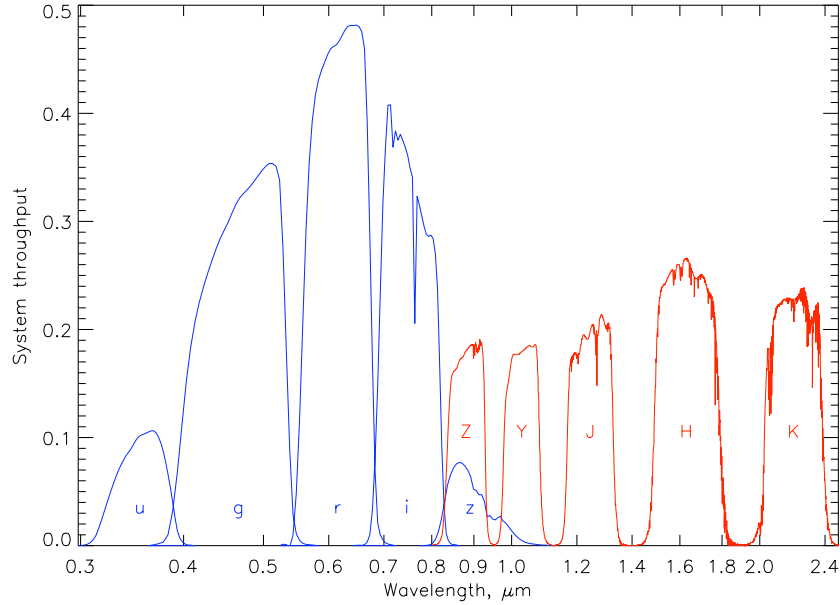


Figure 2.1: Quantum efficiency for SDSS *ugriz* filters (Stoughton et al., 2002, blue curves) and for UKIDSS *ZYJHK* filters (Hewett et al., 2006, red curves). Note that the *Z*-band UKIDSS filter is not used in the LAS. All curves show the response rate for airmass 1.3.

UKIDSS observations are taken using the Wide-Field CAMera (WFCAM; Casali et al., 2007) on UKIRT. Each integration is performed with one of the filters covering the camera’s four detectors. For the Large Area Survey, the filters used are *YJHK*. At any given stage in the survey, the area covered in the four wavelengths will be slightly different. This is particularly true of the *J*-band: the intention is to observe in this passband at two epochs in order to identify proper motions of nearby stars, so a larger area is covered in *J* near the beginning of the survey in order to establish a baseline for these measurements. The four detectors on WFCAM are arranged in a square and are separated by 94 per cent of the width of each detector. In order to cover a contiguous region, the observations are grouped into **tiles** of four telescope pointing positions, producing a small area of overlap between adjacent observations (Lawrence et al., 2007).

In UKIDSS terminology, the image from a single detector is called a **frame**, the set of four frames obtained by a single pointing of the telescope with one of the filters is called a **multiframe**, and a set of individual frames, all covering the same area on the sky but each one with a different filter, is called a **frameset**.

2.1.2 Jittering and microstepping

The total integration time for any single WFCAM pointing in the LAS is either 40 s or 80 s, composed of several short exposures of 5 s, 10 s or 20 s. These individual exposures are offset from

each other in position in order to improve the quality of the data. **Jittering** is when the telescope is offset by a whole number of pixels, ‘in order to minimize the effects of bad pixels and other flat-fielding complications’ (Lawrence et al., 2007). **Microstepping** is when the telescope is offset by $N + 1/2$ or $N + 1/3$ pixels, for 2×2 or 3×3 microstepping respectively. This increases the sampling resolution of the images, which is 0.4 arcsec per pixel without microstepping (Lawrence et al., 2007).

In UKIDSS DR3, there is an almost even split between 40 s and 80 s integrations over all four bands, depending on the atmospheric conditions (using the longer integration time when seeing is poor), and with different integration times corresponding to different regions of the survey area. In the *Y*-band, this consists of 2×20 s or 4×20 s exposures, with no microstepping. In the *J*-band, 2×2 microstepping is always used, with 2×5 s or 2×10 s at each of the four microstepping positions. However, in the *H*- and *K*-bands, there was a change in survey mode after the Early Data Release (EDR). The EDR data in these bands are microstepped (2×2), in the same manner as in the *J*-band. But in subsequent data releases, forming the majority of the sample used in this project, microstepping was not used, and the integrations are composed of 4×10 s or 8×10 s exposures (Dye et al., 2006; Warren et al., 2007).

It follows that, in the *K*-band data used below, some variation in the image quality and depth is to be expected.

The offset between the individual exposures means that a small region at each edge of a co-added frame will not have been covered by every exposure. The image quality will be significantly worse in these regions. Sources detected in these regions are flagged as being near to a frame edge.

2.1.3 Overlap regions: primary and secondary observations

In both SDSS and the LAS, there are small regions of overlap between adjacent strips, stripes or frames. Sources in these overlap regions may be detected more than once, so one of these detections is labelled as the primary observation and the other(s) is (are) labelled as secondary.

2.2 Access to the data

For both UKIDSS and SDSS, with large volumes of data, both images and catalogues, and with incremental data releases, it is not practical for each user to have a local copy of the full data set. The data are therefore made available through online archives. These are described in this section.

In this work, it is generally the catalogues that are used rather than the images, although the images are often used for visual inspection of individual objects. Data from the catalogues are

retrieved using SQL (Structured Query Language) queries through the online WSA¹ (WFCAM Science Archive) for UKIDSS data or through the CAS² (Catalog Archive Server) for SDSS data.

2.2.1 LAS: WSA

Source detections from the individual LAS frames are catalogued in the `lasDetection` table on the WSA. When the frames are combined into framesets, the merged source catalogue is placed in the `lasSource` table. At this stage the sources detected more than once in the overlap regions are labelled as primary or secondary detections.

Many details of the observing times and conditions and of the process of combining different frames and exposures are provided on the WSA in the other tables.

2.2.2 SDSS: CAS

SDSS data are made available via an interface very similar to the WSA (the WSA was modelled on the SDSS CAS). Again, there are various tables containing the photometric data (e.g., `photoObj`) and the spectroscopic data (e.g., `specObj`).

2.2.3 LAS and SDSS combined

The WSA also provides direct access to the SDSS DR5 data, and to a cross-neighbour table, which lists every SDSS source within 10 arcsec of each object in the `lasSource` table. This is used below to create a sample of objects found in both the LAS and SDSS.

2.3 Image processing

UKIDSS and SDSS each have their own pipeline, for extracting lists of sources from the images. In this section some of the properties of the images and some of the issues involved in source extraction are discussed.

2.3.1 Seeing

Photons are refracted as they pass through the earth's atmosphere, so a point source in space will appear as a smeared point when observed through a telescope. This phenomenon is called **seeing** and the transformation from point source to smeared point source is known as the **point spread function**. The seeing is measured as the average FWHM of a point source, estimated from the

¹<http://surveys.roe.ac.uk/wsa/>

²<http://cas.sdss.org/astro/en/>

Table 2.1: Post-processing error bits (ppErrBits) implemented in UKIDSS DR2.

Bit	Decimal	Detection quality issue
4	16	Deblended
6	64	Bad pixel(s) in default aperture
16	65 536	Close to saturated
19	524 288	Possible cross-talk artefact/contamination
22	4 194 304	Lies within a dither offset of the stacked frame boundary

images themselves. LAS observations require estimated seeing to be less than 1.0 arcsec for 40 s integrations (provided the sky is not too bright), or less than 1.4 arcsec for 80 s integrations (Dye et al., 2006).

The seeing is important in that it has an effect on extended sources as well as on point sources: the presence of seeing alters both the size and profile of galaxy images.

For SDSS DR1 data, over 90 per cent of the survey data have seeing better than 1.7 arcsec (Abazajian et al., 2003).

2.3.2 Sky subtraction

The sky brightness for UKIDSS LAS observations in the K -band is around 13.5 mag arcsec⁻². In order to subtract the sky background accurately, variations in the sky brightness across each frame must be taken into account. This is measured using a sky subtraction algorithm that traces variations in the sky brightness with a pixel size of 25.6 arcsec (Dye et al., 2006).

The significance of this is that any objects larger than 25.6 arcsec will interfere with the sky background subtraction, and the flux of the source will be underestimated.

The sky brightness in the r -band is much less of a problem, typically being around 21.0 mag arcsec⁻² (Abazajian et al., 2003).

2.3.3 Warning flags

UKIDSS Data Release 2 introduced quality error bit flags for all sources, given by the ppErrBits column in the detection tables.³ Similar flags are provided for SDSS data. These indicate whether a detection may be of low quality due, for example, to being found close to the edge of a stacked frame, being close to saturated, deblended or having bad pixels in the default aperture. Table 2.1 lists the error bits implemented in UKIDSS DR2.

³<http://surveys.roe.ac.uk/wsa/ppErrBits.html>

2.3.4 Source extraction and deblending

The source extraction algorithm searches the image for peaks above a certain level, which are then identified as the centres of sources. When two or more peaks lie close together, they are assumed to belong to two sources, and the sources must be separated from each other, or **deblended**, with some pixels assigned to one source and some pixels assigned to the other.

Sources that have been deblended are flagged as such in the LAS and in SDSS.

2.4 Photometric galaxy properties

It is very difficult to measure the size of a galaxy. The surface brightness of a galaxy is typically quite high in the centre but becomes very low in the outer regions, with gas and stars continuing out to large distances in the disc or halo. This means it is typically impossible to recover the total flux from a galaxy; the low-surface brightness outer regions will blend in with the sky background, or with neighbouring sources.

Various photometric properties of galaxies may be obtained or derived from the source tables; these are described in this section.

2.4.1 Galaxy profiles

The surface brightness $I(r)$ of an elliptical galaxy or the bulge of a spiral galaxy is approximately described by **de Vaucouleurs' law** (de Vaucouleurs, 1953),

$$I(r) = I_e \exp \left\{ -7.67 \left[\left(\frac{r}{r_e} \right)^{1/4} - 1 \right] \right\} \quad (2.1)$$

where r_e is the **half-light** or **effective radius**, I_e is the surface brightness at radius r_e , and the constant (-7.67) is chosen such that half the total luminosity is emitted within radius r_e . The discs of spiral galaxies are well fit by an **exponential profile**,

$$I(r) = I_0 e^{-r/r_0} \quad (2.2)$$

These may be generalized as two particular examples of **Sérsic (1968)** profiles,

$$I(r) = I_e \exp \left\{ -b_n \left[\left(\frac{r}{r_e} \right)^{1/n} - 1 \right] \right\} \quad (2.3)$$

where n is the **Sérsic index**, with $n = 4$ corresponding to a de Vaucouleurs profile and $n = 1$ to an exponential profile (with $I_0 = I_e e^{b_n}$ and $r_0 = r_e/b_n$). The values for b_n , I_e and r_e are defined as for a de Vaucouleurs profile. A higher value of the Sérsic index indicates a higher concentration.

2.4.2 Vega and AB magnitudes

Throughout this work, Vega magnitudes are used for UKIDSS quantities and AB magnitudes for SDSS quantities. In the AB system, the magnitude is given such that an object with constant flux per unit *frequency* interval has zero colour. (This has an effect on the *K*-corrections, when shifting to a band with different width in frequency.) Conversions between AB and Vega magnitudes for various wavebands are given by [Hewett et al. \(2006\)](#) and [Blanton & Roweis \(2007\)](#).

2.4.3 Fixed-aperture magnitudes

The flux of a galaxy may be measured within a fixed aperture. These are often used for high-redshift surveys, where each galaxy may be assumed to be smaller than the size of the aperture (e.g., [Cirasuolo et al., 2007](#)). They are also useful for measuring galaxy colours, where it is important to have the same aperture in both bands, and for measuring the flux within a spectroscopic fibre.

UKIDSS provides 13 fixed-aperture fluxes, for circular apertures with radius $1/2$, $1/\sqrt{2}$, 1 , $\sqrt{2}$, 2 , $2\sqrt{2}$, 4 , 5 , 6 , 7 , 8 , 10 and 12 arcsec. For the seven smallest apertures, a correction is applied for the effect of seeing on the assumption that the source is a point source.⁴ Including these corrections, the aperture magnitude are given by

$$m = ZP - 2.5 \log \left(\frac{\text{aperFlux}}{\text{expTime}} \right) - \text{aperCor} - \left(\frac{\text{amStart} + \text{amEnd}}{2} - 1 \right) \text{extinction} \quad (2.4)$$

where *ZP* is the zero point in magnitudes for default extinction (airmass 1), *aperFlux* is the flux in ADU, *expTime* is the exposure time in seconds, *aperCor* is the aperture correction, *extinction* is the coefficient for atmospheric extinction, and *amStart* and *amEnd* are the airmass at the start and end of observation respectively.

While it is possible to obtain a similar number of fixed-aperture magnitudes for SDSS galaxies, the only fixed-aperture magnitude used below is the fiber magnitude, which measures the flux within an aperture of 3 arcsec diameter, corresponding to the size of the spectroscopic fibres, with no correction applied for seeing ([Stoughton et al., 2002](#)).

2.4.4 PSF magnitudes

For SDSS point sources, the PSF magnitude, determined by fitting a PSF model to the object, provides the optimal measure of the flux ([Stoughton et al., 2002](#)). For galaxies, this magnitude may be used to estimate the colour at the centre of the galaxy ([Driver et al., 2006](#)).

⁴http://surveys.roe.ac.uk/wsa/www/gloss_a.html#multiframe-detector_apercor1

2.4.5 Petrosian magnitudes and radii

[Petrosian \(1976\)](#) apparent magnitudes are used in this thesis. The same particular form of the Petrosian magnitude is used in both SDSS and UKIDSS and is described here.

The **Petrosian radius** is defined as the radius at which the ratio of the local surface brightness (azimuthally averaged) to the mean surface brightness within that radius is equal to 0.2. The flux of the galaxy is then integrated within an aperture with radius twice the Petrosian radius to give the **Petrosian flux** ([Blanton et al., 2001](#); [Dye et al., 2006](#)).

The advantage of the Petrosian magnitude is that the fraction of the total galaxy light recovered depends on the shape of the galaxy surface brightness profile but not on the amplitude. This is in contrast with isophotal magnitudes, where the galaxy flux is measured within a certain isophote. When these magnitudes are used, if the amplitude of the surface brightness profile is diminished, for example, by cosmological dimming or by Galactic extinction, then the fraction of the galaxy flux recovered by the isophotal magnitude also diminishes. This is discussed in more depth by [Blanton et al. \(2001\)](#).

The fraction of the galaxy light recovered by the Petrosian magnitude varies monotonically with Sérsic index, with a higher deficit for higher concentration ([Graham et al., 2005](#)). For an exponential profile this fraction is 99 per cent (+0.01 mag), falling to 82 per cent (+0.22 mag) for a de Vaucouleurs profile ([Blanton et al., 2001](#)).

There is an additional effect due to seeing. A galaxy with a small angular size will have an apparent profile that approaches the PSF, so the Petrosian magnitude will pick up a different fraction of the galaxy light than it would in the absence of seeing ([Blanton et al., 2001](#)).

For UKIDSS Petrosian magnitudes, it is important to note that there is a limit on the aperture size. This is imposed due to the relatively small pixel size of the sky subtraction algorithm and is set to a maximum diameter of 24 arcsec. The Petrosian radius is half the aperture radius, so the maximum allowed size of the Petrosian radius is 6 arcsec. (To be more precise, the limit on Petrosian radius is set to 15 pixels for no microstepping or 30 pixels where microstepping has been used. The precise pixel size in these cases is always a tiny bit over 0.4 or 0.2 arcsec, respectively, so the Petrosian radius is limited to just over 6 arcsec.) Galaxies which would otherwise have a larger Petrosian radius are assigned a Petrosian radius of 6 arcsec, and therefore the ‘Petrosian radius’ and the ‘Petrosian flux’ are underestimates of the true values.

In what follows, corrections are not applied to compensate for the effect of seeing on the Petrosian magnitudes, or for the different fraction of the galaxy’s flux recovered by the Petrosian magnitudes for different galaxy types.

2.4.6 Kron magnitudes and radii

Although not used to generate the results of this thesis, [Kron \(1980\)](#) magnitudes have been widely used for galaxies. These are provided in the LAS detection table as an alternative to the Petrosian magnitude ([Dye et al., 2006](#)), following the definition used in the SExtractor ([Bertin & Arnouts, 1996](#)) source extraction software. The **Kron radius** is given by

$$r_1 = \frac{\sum rI(r)}{\sum I(r)} \quad (2.5)$$

where the summation is over a two-dimensional aperture ([Graham & Driver, 2005](#)). The flux of the galaxy is then integrated over an aperture with radius kr_1 to give the **Kron flux**, with $k = 2$ for UKIDSS and $k = 2.5$ for 2MASS. As with the Petrosian radius, the UKIDSS Kron radius is limited to 6 arcsec maximum.

According to [Graham & Driver \(2005\)](#), with $k = 2$ the Kron flux is expected to recover 90.6% of the flux of an exponential galaxy or 87.0% for a de Vaucouleurs profile, corresponding to magnitude errors of 0.10 and 0.15 mag respectively. However, they note that this is the case only if the summation in Equation (2.5) is performed to an infinite radius; if the integration ceases at only a few times the effective radius, then up to half of the galaxy flux may be missed for galaxies with high concentration (see also [Andreon, 2002](#)).

2.4.7 Model magnitudes

There is no direct and reliable way to measure directly the total flux of a galaxy, so the total flux may be estimated only by extrapolating the measured galaxy profile according to some model.

The SDSS model magnitudes do this using the better of either a de Vaucouleurs or an exponential profile fit to the two-dimensional image, with arbitrary axis ratio and position angle, and including convolution with the PSF ([Stoughton et al., 2002](#)).

[Blanton et al. \(2005a\)](#) provide [Sérsic \(1968\)](#) profile fits to the azimuthally averaged radial profiles of SDSS galaxies, giving a measure of the Sérsic index of the galaxies. The **Sérsic magnitude** is then obtained by integrating the Sérsic profile in Eqn. (2.3) to infinity. However, using artificial galaxy images, they find that for galaxies with a large size or a high Sérsic index, the sizes and fluxes are underestimated by about 10 and 15 per cent respectively. They attribute this bias partly to the effect of seeing but mostly to uncertainty in the local sky level.

Sérsic profile fits and magnitudes have not yet been implemented for UKIDSS ([Dye et al., 2006](#)).

2.4.8 Half-light radii

The scale-size of a galaxy is conventionally measured using the radius enclosing half of the total light, known as the half-light or effective radius and denoted by r_e . In order to correct for inclination, the half-light radius we would like is the semi-major axis of the elliptical aperture, of the same ellipticity and position angle as the galaxy, containing half the total flux of the galaxy.

For the LAS, this is estimated (Cross et al., in preparation) using the Petrosian flux, the 13 circular aperture fluxes, the ellipticity and the seeing, all made available in the WSA (the pipeline does not measure the half-light radius). The total light is estimated by assuming the Petrosian flux recovers 90 per cent of the total flux. Three half-light radii (HLR) are found: a circular HLR, an elliptical HLR and then a final HLR that corrects for the effects of seeing.

In the SDSS, the radius of the circular aperture enclosing half the Petrosian flux is provided in the catalogue, as is the radius of the circular aperture containing 90 per cent of the flux. The disadvantage of this circular half-light radius is that it varies with a galaxy's inclination with respect to the observer.

2.4.9 Surface brightness

Various measures of the surface brightness may be found using different estimates of the galaxy magnitude and radius: for example, the central surface brightness (related to the detectability of a galaxy) and the effective surface brightness (of more physical significance).

In general, the mean surface brightness within a certain aperture is given in mag arcsec^{-2} by

$$\mu = -2.5 \log \frac{f}{\pi r^2} + \dots \quad (2.6)$$

$$= m + 2.5 \log \pi r^2 \quad (2.7)$$

where f is the flux (in appropriate units), m is the magnitude and r is radius in arcsec of the aperture containing the flux.

The mean surface brightness within the Petrosian aperture is given by

$$\mu_{\text{Petro}} = m_{\text{Petro}} + 2.5 \log \pi (2r_{\text{Petro}})^2 = m_{\text{Petro}} + 2.5 \log 4\pi r_{\text{Petro}}^2 \quad (2.8)$$

where r_{Petro} is the Petrosian radius.

The half-light, or effective, surface brightness is estimated from the half-light radius by (Blanton et al., 2001)

$$\mu_e = m + 2.5 \log 2\pi r_e^2 \quad (2.9)$$

where $m + 2.5 \log 2$ corresponds to half the total flux. If the half-light radius used is the semi-major axis, as described above, then the effective surface brightness gives an estimate of the *face-on* effective surface brightness, correcting for inclination relative to the observer.

For an elliptical half-light radius, the mean surface brightness within the half-light ellipse, not correcting for inclination, is given by

$$\mu_{e,\text{ell}} = m + 2.5 \log 2\pi a_e b_e \quad (2.10)$$

$$= m + 2.5 \log 2\pi(1 - e)a_e^2 \quad (2.11)$$

where a_e and b_e are the half-light semi-major and semi-minor axes respectively, and using the UKIDSS definition of ellipticity, $e = 1 - b/a$ (Dye et al., 2006). This measure of the surface brightness is related to the detectability of the object, since it reflects the observed surface brightness of the galaxy rather than the intrinsic (inclination-corrected) surface brightness.

2.5 Spectroscopic galaxy properties

SDSS is a combined photometric and spectroscopic survey, with the imaging survey used to create a list of spectroscopic targets. When the targets have been chosen, they are assigned to spectroscopic fibres for observation with the multi-object spectrograph on the same telescope. The fibres are positioned in holes drilled in plates, and the plates are tiled for high completeness (Blanton et al., 2003a).

Some of the target galaxies are not followed up for spectroscopy, either because there are insufficient fibres to cover all the galaxies in that field, or because the galaxies are separated by less than 55 arcsec: too close for neighbouring fibres to be positioned ('fiber collisions'). 'This causes about 6 per cent of galaxies to be missed, and these will be biased towards regions with a high surface density of galaxies' (Ball et al., 2006).

2.5.1 Spectroscopic target selection

The selection criteria for the SDSS main galaxy sample are described by Strauss et al. (2002). Star-galaxy separation is achieved by comparing the PSF magnitude with the model magnitude; if they are very similar then the source is probably a star. As well as various quality control requirements, there are limits on the fiber magnitude (within a 3 arcsec aperture) of $r > 15.0$, $g > 15.0$ and $i > 14.5$, to avoid crosstalk between adjacent fibres, and a faint limit of 17.77 for the r -band Petrosian magnitude (17.6 for the EDR). Low-surface brightness galaxies are excluded by requiring the Petrosian effective surface brightness to be brighter than $24.5 \text{ mag arcsec}^{-2}$.

2.5.2 Redshifts and classification

The spectroscopic pipeline computes an emission-line redshift (where possible) and an absorption redshift (by cross-correlating with various templates) and returns the redshift with highest con-

confidence as the final redshift. A spectroscopic classification (quasar, high-redshift quasar, galaxy, star, late-type star or unknown) is made based on the cross-correlation with template spectra. Unreliable redshifts may be excluded by setting constraints on the values of z_{Conf} , z_{Status} and z_{Warning} .

2.6 Dust

Dust can have a significant effect on the observed properties of a galaxy, due both to dust extinction within the Milky Way and to dust attenuation within the galaxy itself.

2.6.1 Galactic extinction

Galactic extinction corrections are used throughout, based on the extinction maps of [Schlegel et al. \(1998\)](#), with UKIDSS incorporating the corrections of [Bonifacio et al. \(2000\)](#). K -band extinction corrections for the whole LAS DR3 lie in the range $A_K = 0.0147 \pm 0.0098$ while r -band corrections for the SDSS DR5 main galaxy sample lie in the range $A_r = 0.0898 \pm 0.0642$ (mean and standard deviation).

2.6.2 Inclination dependence

Dust attenuation in galaxies depends on the amount and distribution of the dust, on the rest-frame waveband and on the inclination at which the galaxy is observed. There are two aspects to this dust attenuation: the face-on dust attenuation and the inclination-dependent dust attenuation.

[Driver et al. \(2007b\)](#) have investigated these for the B -band MGC. They have shown that dust has a considerable effect on the shape of the optical luminosity function, and caution that the effect could still be significant in the near-infrared. To correct for this inclination-dependent dust attenuation would ideally require good bulge-to-disc decomposition, which is beyond the scope of this work. I choose here to present results that are not corrected for dust, thus representing the amount of K -band light that escapes from the galaxy, rather than the amount of K -band light emitted by the stars in the galaxy (some of which will be absorbed by dust, and some of which will be re-emitted in the K -band by dust, [Driver et al., 2008](#), fig. 4). However, in estimating the absolute magnitude of a galaxy the implicit assumption has been made that galaxies emit light in an isotropic way, which is not the case: a disc galaxy will appear fainter when viewed edge-on than when viewed face-on, leading to a corresponding under- or over-estimate, respectively, of the galaxy's total (attenuated) luminosity. This will lead to a blurring of the LF for disc galaxies, but less so than at optical wavelengths.

However, dust must be taken into account when estimating the stellar masses. One way to proceed with this is to include dust in the models (e.g., [Kauffmann et al., 2003](#)).

Chapter 3

Data: completeness, contamination and reliability

*Il y a sans doute des lois naturelles,
mais cette belle raison corrompue a tout corrompu.*

Essential to any statistical analysis is a good understanding of the completeness and reliability of the sample. Does the sample include all galaxies within certain limits, or is a significant fraction missing? Are any unwanted objects included, such as stars and imaging artefacts? How reliable are the measurements of the galaxy properties?

The completeness, contamination and reliability of the LAS sample will be the focus of this chapter, with the limits on the SDSS sample included towards the end.

3.1 Motivation

Before investigating the LAS data, the possible sources of incompleteness are discussed in this section, along with methods of modelling the incompleteness.

3.1.1 Possible sources of incompleteness, contamination and unreliability

Completeness

The selection function is the probability of a galaxy being detected, given certain properties, such as apparent magnitude, surface brightness and angular size. There will be a region of this parameter space at which a detection is almost certain, unless there is some random defect in the survey (bad pixels, etc.). Conversely, there will also be a range of this parameter space in which a detection is impossible. Between these two regions is a significant ‘grey area’, in which the probability

of a detection will lie somewhere between zero and unity. We say a sample is **complete** if the limits placed on the observable quantities are such that a detection is almost certain for any object within those limits.

We may group the sources of incompleteness into two categories: systematic incompleteness, for example, where a galaxy is too faint to be detected, and random incompleteness, where the incompleteness is not dependent on the properties of the galaxy.

Systematic sources of incompleteness include:

1. Faint surface brightness. For point sources, a sufficiently high signal-to-noise ratio (S/N) must be achieved in order for the source to be identified. This magnitude limit is dependent on the PSF and hence on the seeing, since poorer seeing will cause the light from the source to be distributed over a larger number of pixels, and thus the S/N on each pixel will be diminished. The faint magnitude limit is thus more accurately related to a limit in peak surface brightness; when one speaks of a faint magnitude limit for extended sources, this may not be accurate for low-surface brightness galaxies. If a galaxy is large compared with the PSF, then the central (peak) surface brightness of the galaxy must exceed a certain threshold in order for the galaxy to be detected.
2. Small angular size. When it is required that a source be classified as a galaxy rather than a star, a small angular size may cause the sample to be incomplete. This is because star-galaxy separation often requires that a source should have a profile significantly different from the PSF, which will not be the case for very compact galaxies.
3. Failed redshifts. If redshifts are required, then the sample will be incomplete if a galaxy is not included in the spectroscopic sample. There may be systematic reasons for this, related to the criteria defining the spectroscopic sample.

These systematic sources of incompleteness will be considered below, and must either be dismissed as insignificant or else included explicitly in the analysis. The precise nature of these constraints will vary from point to point, depending on such factors as seeing, sky brightness, exposure time and image resolution (microstepping). It is conventional to place conservative limits on the completeness of the sample, in order that this variation in the true limits can be ignored.

Random sources of incompleteness include:

1. Lack of coverage. This is the most obvious and trivial source of incompleteness: if a region of the sky has not been surveyed, the galaxies will not be detected, and this is assumed (reasonably, on sufficiently large scales) to be independent of the properties of the galaxies.

2. Foreground (or background) objects. If a galaxy lies behind a bright star or behind another galaxy, this may prevent it from being detected, and may also contaminate the measurement of the foreground object.
3. Bad pixels. A source which would otherwise be a marginal detection may fall below the detection threshold if there are bad pixels on the detector. (The effect of this is reduced in UKIDSS by jittering.)
4. Failed redshifts. In addition to the systematic sources of incompleteness due to failed redshifts, there will be random incompleteness due to redshifts not being available. This may occur if that part of the sky has not been observed for spectroscopy, or if there were insufficient fibres available, or if the redshift could not be determined for some reason.

Assuming these are truly random (i.e., independent of the properties of the galaxy), these need not be included explicitly in the analysis, except that they will have an effect on the normalization of the results. (This would not be the case for a clustering analysis, where random sources of incompleteness do have an effect.)

Contamination

A large sample of galaxies is unlikely to be completely ‘clean’; there may be some non-galaxies included. The sample may be contaminated by:

1. Stars. In order to separate the galaxies from the stars, there is a balance between completeness and reliability of the sample. If all of the galaxies are to be included, then it is likely that some stars will be included as well. Double stars are particularly susceptible to this (Blanton et al., 2005a). Note that this is unlikely to be a problem in this work, since the galaxies in the sample must be confirmed by their spectra as being galaxies.
2. Asteroids, satellites and planes. These tend to appear in images as straight lines, and may be classified as galaxies by their photometry. Once again, this is unlikely to be a problem when the spectra have been obtained.
3. Artefacts. Dye et al. (2006) mention various artefacts that may appear in UKIDSS images. Since the analysis below requires detection in both SDSS and UKIDSS, as well as SDSS spectroscopy, it is highly unlikely that the sample will be contaminated by artefacts.
4. Bad deblends of large galaxies. This may result in two or more ‘galaxies’ being detected, rather than one large galaxy.

These effects need to be kept to a minimum.

Reliability

An unreliable detection occurs when a genuine galaxy is detected, but the measurements of the galaxy properties are very poor. Possible sources of unreliable detections include:

1. High surface brightness. A galaxy is detected, but some pixels on the detector are saturated so a reliable measurement is impossible.
2. Faint apparent magnitude or surface brightness. The overall S/N of the galaxy is low so the errors on the magnitude are high, resulting in an unreliable measurement.
3. Large angular size. This could lead to problems with deblending or confusion with the sky subtraction.
4. Peculiar velocities. Low-redshift galaxies may have a peculiar velocity that makes the redshift a poor indicator of the distance to the galaxy.
5. Poor deblending. Large galaxies may be deblended, leading to an underestimate of the galaxy's flux and size. Even when deblending should be performed, this may not be done accurately, leading to uncertainty in the galaxy properties.
6. Background (or foreground) objects. If a galaxy overlaps with another object, the flux from the second object may be combined with the light from the galaxy, thus giving it too high a flux.
7. Artefacts. Any artefacts in the image that overlap with a genuine galaxy may render its measurement inaccurate.

Often these problems will evade detection or they may be identified through the galaxy images or as galaxies with very unusual colours. When unreliable measurements are identified, they can be included in the analysis, e.g., as lower or upper limits on the galaxy flux, or they may be treated as non-detections and excluded completely.

When a sample is clearly defined, e.g., by setting a strict magnitude limit, then unreliable measurements can make a difference to whether a galaxy is included in the sample. For example, a poor measurement of the galaxy flux can cause a galaxy to cross the faint magnitude limit in one direction or the other.

3.1.2 Modelling the magnitude uncertainties

The general relationship between uncertainty in flux and uncertainty in magnitude is derived here. This will be used below, when fitting a function to the uncertainty in magnitude, and in Chapter 6,

when simulating the effect of magnitude errors on estimation of the luminosity function.

The flux f is defined for this section as follows:¹

$$m = -2.5 \log f \quad (3.1)$$

where m is the magnitude.

The flux is based on the number of counts on a range of pixels on the CCD chip. This is subject to Poisson statistics, with an uncertainty on the ‘true’ counts of \sqrt{N} , where N is the number of counts recorded. In addition, the uncertainty in the flux will be related to the uncertainty in the local sky background level. In this section, it will be assumed that the uncertainty may be approximated by a Gaussian uncertainty in the galaxy flux (rather than in the magnitude).

Assuming the flux is $f \pm \Delta f$, the magnitude then lies between m_- and m_+ where

$$m_- = -2.5 \log(f + \Delta f) \quad (3.2)$$

and

$$m_+ = -2.5 \log(f - \Delta f) \quad (3.3)$$

(Note that a symmetrical uncertainty in flux does not translate to a symmetrical uncertainty in magnitude, due to the logarithm.) The uncertainty could be defined either as $\Delta m = m - m_-$ or as $\Delta m = m_+ - m$, or as $\Delta m = (m_+ - m_-)/2$. These respectively give

$$m - m_- = -2.5 \log \frac{f}{f + \Delta f} \quad (3.4)$$

$$m_+ - m = -2.5 \log \frac{f - \Delta f}{f} \quad (3.5)$$

$$\frac{m_+ - m_-}{2} = -1.25 \log \frac{f - \Delta f}{f + \Delta f} = -2.5 \log \sqrt{\frac{f - \Delta f}{f + \Delta f}} \quad (3.6)$$

The largest value for Δm will be given by the most negative logarithm (all three give negative logarithms). This is the second, since

$$0 < \frac{f - \Delta f}{f} < \sqrt{\frac{f - \Delta f}{f + \Delta f}} < \frac{f}{f + \Delta f} < 1 \quad (3.7)$$

assuming $\Delta f < f/2$. To be conservative, this definition is chosen. Combined with $m = -2.5 \log f$, which gives $f = 10^{-0.4m}$, this gives a value of Δm that satisfies

$$\Delta f = (1 - 10^{-\Delta m/2.5})f \quad (3.8)$$

$$= (1 - 10^{-\Delta m/2.5})10^{-0.4m} \quad (3.9)$$

¹“Maggies” in KCORRECT (Blanton & Roweis, 2007) terminology.

Estimating the Poisson uncertainty from the flux itself, we have $\Delta f \propto \sqrt{f}$ (noting the arbitrary normalization of f), or $\Delta f = \alpha\sqrt{f}$, where α is a constant. (For bright sources, the fractional uncertainty, $\Delta f/f$, tends to zero.)

From the largest value of Δm above, combined with $\Delta f = \alpha\sqrt{f}$, the magnitude error is given by

$$\Delta m \equiv m_+ - m = -2.5 \log(1 - \alpha f^{-1/2}) \quad (3.10)$$

$$= -2.5 \log(1 - \alpha 10^{0.2m}) \quad (3.11)$$

$$= -2.5 \log\left(1 - 10^{(m+5 \log \alpha)/5}\right) \quad (3.12)$$

This breaks down for $\alpha 10^{0.2m} > 1$, or $m > 5 \log(1/\alpha)$. For example, with $\alpha = 10^{-4}$, this definition of the error will break down for $m > 20$. This relation will be used in the following section and in Fig. 3.1.

3.2 Repeat observations in the LAS

In any one of the LAS bands, a significant number of sources will be observed more than once, where adjacent frames overlap. When the frames are seamed together, one of these observations is selected as the primary observation, to enable a seamless sample to be obtained, but the secondary observations are also included in the catalogue. These duplicate observations are examined in Section 3.2.1. The effect of deblending in the four LAS bands, *YJHK*, is then investigated in Section 3.2.2.

3.2.1 Sources observed more than once in the *K*-band

The LAS was searched for sources in the overlap regions with duplicates, using the `lasSource-Neighbours` table in DR2+. Searching for pairs of sources where the first was the primary detection yielded 498 016 pairs with Petrosian magnitudes, no error flags and separation of less than 0.6 arcsec. Requiring the closest secondary detection (thus including each primary object no more than once) reduced the sample to 467 519 pairs, of which 248 305 are consistently classified as galaxies (`kClass = 1`).

These pairs of sources can be very useful for investigating completeness, contamination and reliability of the catalogues. Here they are used to estimate the errors on the Petrosian magnitudes and to estimate the completeness as a function of Petrosian magnitude.

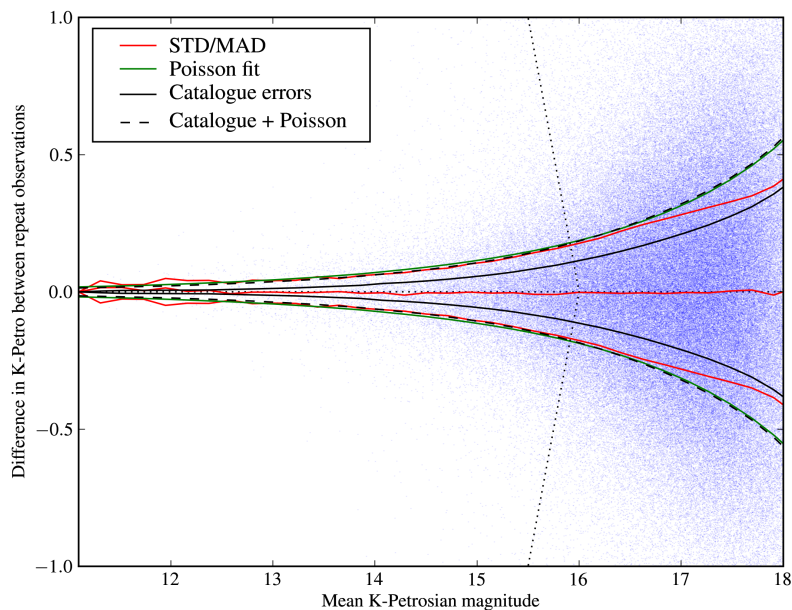


Figure 3.1: Magnitude error estimated from repeat observations of 248 305 galaxies in the LAS. Shown in red are the median and curves showing $\sigma_m = \sigma_{\Delta m}/\sqrt{2}$ where $\sigma_{\Delta m}$ is estimated as $\sigma_{\Delta m} \simeq 1.4826 \times \text{MAD}$, all estimated in 50 independent bins. The smooth green curves show the Poisson-type errors expected for $\alpha = 10^{-4}$ while the black curves show the median catalogue errors (in 50 bins), unaltered (solid line), and when summed in quadrature with Poisson-type errors for $\alpha = 10^{-4.1}$ (dashed line). Points to the left of the black dotted lines correspond to galaxies with both magnitudes brighter than the completeness limit (here set to $K = 16$: see main text).

Magnitude errors

The catalogue provides magnitudes and uncertainties in these magnitudes. Repeat observations make it possible to assess the accuracy of the catalogue errors.

Assuming the observed magnitude, m , is drawn from a normal distribution with mean equal to the true magnitude, m_t , and variance σ^2 , $m \sim N(m_t, \sigma^2)$, then the difference between two observed magnitudes, $y = m_1 - m_2$ will be drawn from a probability distribution,

$$g(y) = \int_{-\infty}^{\infty} f(m_2 + y) f(m_2) dm_2 \quad (3.13)$$

where $f(m)$ is the probability of observing magnitude m . Then $g(y)$ will be a normal distribution with mean 0 and variance $2\sigma^2$. When comparing repeat observations to estimate σ^2 , it is necessary to include this factor of 2.

Figure 3.1 shows the variation in K_{Petro} over repeat observations of the same galaxy as a function of the mean Petrosian magnitude, which is taken as an estimate of the true magnitude of

the galaxy. Also plotted are the median and an estimate of the standard deviation based on the MAD (Median Absolute Deviation), $\text{MAD} = \text{med}_i (|x_i - \text{med}_j (x_j)|)$, which is a robust measure of the variation in the presence of outliers. Poisson-type errors are shown for comparison using Equation (3.11).

Poisson-type errors with $\alpha = 10^{-4}$ give a good fit (by eye) to the observed magnitude errors out to $K \simeq 16.5$. (This corresponds to magnitude errors of around 0.19 mag at $K = 16$.) This reflects the effective Poisson errors due to detector counts, sky subtraction, galaxy shape, surface brightness profile, etc. At fainter magnitudes, incompleteness renders this method inappropriate, as it cannot be assumed that faint galaxies will be detected in two separate observations.

Note that the catalogue magnitude errors are approximately accurate when added in quadrature to Poisson-type errors with $\alpha = 10^{-4.1}$. This may be used to estimate the true magnitude uncertainty from the value given in the catalogues. The correction ranges from a factor 3.5 increase on the catalogue errors at $K \simeq 12$ to a factor 1.5 increase at $K \lesssim 16$.

Completeness

It is possible to use the overlap regions between LAS frames to estimate the completeness. Given a detection of a galaxy with magnitude K in one frame, what is the probability that this galaxy will also be detected in the overlapping frame?

The K -band regions of the LAS DR2+ comprise 9724 frames. The boundaries of these frames were obtained from the currentAstrometry table in order to investigate the geometry of the survey area. The frames were found to cover a total area of 526.1 square degrees, 87.0 square degrees of which lie in the 34 011 overlap regions between frames, so the actual area of the survey is 439.1 square degrees.

Assuming the number of sources in existence with magnitude between m and $m + dm$, per unit solid angle, is given by $\rho(m) dm$. Then, assuming that whether or not a source is detected depends only on its magnitude, the differential number counts detected over a solid angle A_1 will be $n_1(m) = A_1 \rho(m) p(m)$, where $p(m)$ is the probability that a source of magnitude m will be detected. Note that for A_1 a non-seamless sample is used, treating the frames as distinct.

For sources that have been observed twice, the number counts will be

$$n_2(m) = A_2 \rho(m) (p(m))^2 \quad (3.14)$$

where A_2 is the total area of the overlap regions. Note that a fraction (~ 7 per cent) of the overlap area has been imaged more than twice. In these regions, the probability of a repeat observation (twice or more than twice) is increased.

Duplicate sources in the LAS are flagged as either primary ($\text{priOrSec} = \text{frameSetID}$) or secondary ($\text{priOrSec} > 0$ and $\text{priOrSec} \neq \text{frameSetID}$), whereas non-duplicate sources are not flagged at all ($\text{priOrSec} = 0$). It is therefore possible to select those sources that have been observed more than once in repeated observations by requiring $\text{priOrSec} > 0$. The sources for $n_1(m)$ are selected with no constraint on priOrSec , while the sources for $n_2(m)$ require a primary detection ($\text{priOrSec} = \text{frameSetID}$), indicating that a repeat detection has been made. One count is made for all objects, while another is made requiring classification as a galaxy ($\text{kClass} = 1$).

The completeness is given by $p(m)$, which may be found by

$$p(m) = \frac{\rho(m)(p(m))^2}{\rho(m)p(m)} = \frac{n_2(m)/A_2}{n_1(m)/A_1} \quad (3.15)$$

This is subject to Poisson errors, and (setting $q \equiv n_2/n_1$) has variance

$$\text{Var}(p) = \left(\frac{A_1}{A_2}\right)^2 \text{Var}\left(\frac{n_2}{n_1}\right) \quad (3.16)$$

$$= \left(\frac{A_1}{A_2}\right)^2 \left(\left(\frac{\partial q}{\partial \bar{n}_2}\right)^2 \text{Var}(n_2) + 2 \frac{\partial q}{\partial \bar{n}_2} \frac{\partial q}{\partial \bar{n}_1} \text{cov}(n_2, n_1) + \left(\frac{\partial q}{\partial \bar{n}_1}\right)^2 \text{Var}(n_1) \right) \quad (3.17)$$

$$= \left(\frac{A_1}{A_2}\right)^2 \left(\frac{1}{\bar{n}_1^2} \text{Var}(n_2) - 2 \frac{1}{\bar{n}_1} \frac{\bar{n}_2}{\bar{n}_1^2} \text{cov}(n_2, n_1) + \frac{\bar{n}_2^2}{\bar{n}_1^4} \text{Var}(n_1) \right) \quad (3.18)$$

Given that the objects in the overlap regions are also contained in the whole sample, there is a non-zero covariance between n_1 and n_2 . Writing $n_1 = n + n_2$ and noting that n_2 and n are almost independent, the covariance is given by

$$\text{cov}(n_2, n_1) = \overline{n_2 n_1} - \bar{n}_2 \cdot \bar{n}_1 \quad (3.19)$$

$$= \overline{n_2(n + n_2)} - \bar{n}_2 \cdot \overline{(n + n_2)} \quad (3.20)$$

$$= \overline{n_2 n} + \overline{n_2^2} - \bar{n}_2 \cdot \bar{n} - \bar{n}_2^2 \quad (3.21)$$

$$= \text{cov}(n_2, n) + \text{Var}(n_2) \quad (3.22)$$

$$\simeq \text{Var}(n_2) \quad (3.23)$$

Taking $\text{Var}(n_2) = n_2$ and $\text{Var}(n_1) = n_1$, and estimating \bar{n}_2 as n_2 and \bar{n}_1 as n_1 , Equation (3.18) becomes

$$\text{Var}(p) = \left(\frac{A_1}{A_2}\right)^2 \left(\frac{1}{n_1^2} n_2 - 2 \frac{1}{n_1} \frac{n_2}{n_1^2} n_2 + \frac{n_2^2}{n_1^4} n_1 \right) \quad (3.24)$$

$$= \left(\frac{A_1}{A_2}\right)^2 \frac{n_2 (1 - n_2/n_1)}{n_1} \quad (3.25)$$

There is an additional complication in that sources cannot be detected arbitrarily close to the edge of a frame. In the analysis here, sources flagged as being close to the edge of the frame have been removed (as have sources flagged as deblended). This occurs when the source lies within a

jitter plus a microstep plus twice the default aperture radius ($2 \times 2 \text{ arcsec} = 4 \text{ arcsec}$) of the frame boundary.²

Consider the overlap region between two adjacent frames. These typically have a width of approximately 40 arcsec. One side of the region will be close to the edge of one frame, while the other side will be close to the edge of the other frame. The image quality is worse at the edge of a frame, so at either side of the overlap region, the probability of a repeat detection of a source is less than it is in the middle. The typical padding on a frame is around 10 arcseconds, so only around half of the overlap region is free from this effect.

One would ideally like to identify those regions not close to the edge of either frame. However, this is impossible using the priOrSec value alone, and it would be highly non-trivial to select repeat observations explicitly by position.

As an imperfect solution, each frame is given a padding of 4 arcsec. With this 4 arcsec padding, the total area (A_1) is 516.05 square degrees, with 75.81 square degrees (A_2) of overlap in 33 965 regions. This corresponds to a non-overlapping survey area of 440.24 square degrees. The padding on each frame will always be larger than this, so this simplification has the effect that A_1 will be slightly overestimated. The effect on A_2 is less easy to judge. This is because only the primary detection is required to be a good distance from the edge of the frame. If the repeat detection occurs at either side of the overlap region, then its primary detection will not be flagged as close to the edge of the frame, since the detection that is close to the frame edge will not be chosen as the primary. So A_2 may be significantly underestimated, although the probability of a repeat detection will not be uniform across the area.

It was noted above that an assumption has been made that the overlap regions are imaged twice, and not more than twice. However, some overlap regions are imaged three, four, five or six times. With the sample from the beginning of this section (slightly different from the sample used here), it is possible to quantify this effect. When investigating the magnitude uncertainties, it was found that of 498 016 repeat observations of primary sources, there were 467 519 distinct primary sources. In total, there are 54 618 pairs consisting of a primary detection and two or more secondary detections. These pairs consist of 24 121 distinct primary detections, leaving 30 497 extra pairings (~ 7 per cent of 467 519), which were discarded from the earlier analysis. Of the 24 121 distinct primary detections with two or more counterparts, 18 079 (75%) have two counterparts, 5 803 (24%) have three, 144 (0.6%) have four, and 95 (0.4%) have five. This will have an effect on the normalization of the completeness results, but the effect on the general shape is likely to be very small.

²http://surveys.roe.ac.uk/wsa/ppErrBits.html#Source_image_close_to_frame_boun

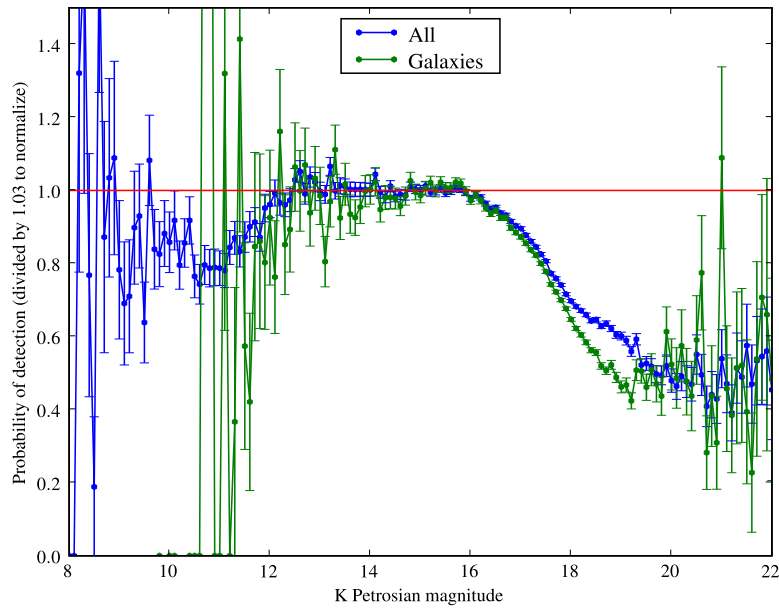


Figure 3.2: Completeness as a function of magnitude, estimated through repeat observations in overlap regions in the LAS, for all objects (blue), and for those objects classified as galaxies (green; $k\text{Class} = 1$). The normalization over the whole range has been adjusted to give approximately unity for $14 < K_{\text{Petro}} < 15.5$.

Figure 3.2 shows the completeness as a function of magnitude, found through Equation (3.15). It can be seen that objects fainter than $K = 16$ are incompletely sampled in the LAS. It appears that objects brighter than $K \simeq 12$ are also incompletely sampled, possibly due to deblending, mis-classification, or large angular size resulting in a failure to match between repeat observations. Also a factor at bright magnitudes would be artefacts, which would not show up in repeat observations, and which would give the impression that the completeness is worse than it actually is. The curves have been normalized down to reach unity (on average) between $14 < K_{\text{Petro}} < 15.5$ to compensate for various weaknesses of the method. However, despite these weaknesses, there is clear evidence for incompleteness at $K > 16$.

3.2.2 Effect of deblending in *YJHK*

During the course of this work it became apparent that a significant number of galaxies had abnormally bright Petrosian magnitudes. With the introduction of the `ppErrBits` quality control flags in UKIDSS DR2 (see Section 2.3.3), it was found that this effect was caused by galaxies which had been flagged as being deblended. In this section the effect of deblending on the apparent magnitudes is investigated for sources observed in the four LAS bands.

Deblending takes place when the source extraction algorithm identifies two or more sources that overlap in the image. Two sources are considered to be overlapping if the lowest isophotes relative to the sky encompass both of the sources. In this case, the flux from the area of overlap must be apportioned appropriately to each of the sources. The method used for UKIDSS deblending is based on that of [Irwin \(1985\)](#) ([Lawrence et al., 2007](#)).

Several factors may affect whether two neighbouring sources are deblended.

1. With shallow imaging it is not possible to probe faint isophotes, so two neighbouring sources are less likely to require deblending, since the faintest observed isophotes are less likely to overlap. This means that deblending will occur more often for deeper imaging.
2. A higher image resolution (using microstepping in UKIDSS) would also affect whether two sources need deblending, as the higher resolution may prevent the isophotes from overlapping.
3. Poorer imaging may prevent deblending from taking place if one of the sources is difficult to detect. In this case, the second source will not be detected at all and its light will be included in the flux from the first source (source confusion). The flux from the first source will therefore be slightly overestimated.

Given two observations of the same source, only one of which is flagged as deblended, it follows that one would expect the deblended observation to return a flux that is similar to, or slightly fainter than, the non-deblended observation.

In the four LAS bands, deblending is performed independently; for example, a galaxy with a close neighbour may be deblended in Y but not deblended in H . This fact may be used to test whether the deblending algorithm is working correctly. If there is a problem with the algorithm, this is likely to show up in the colours of galaxies that have been deblended in one band but not in the other.

The `lasYJHKsource` table was queried for all sources that are detected in all four bands and that have `ppErrBits` in each band set either to 0 (no quality control issues) or 16 (deblended). Approximately 5 per cent of sources are flagged as deblended in one or more bands. Four magnitudes were selected: `AperMag3` (2 arcsec diameter), `AperMag7` (8 arcsec), `AperMag13` (24 arcsec) and `PetroMag` (Petrosian), although only the results for `AperMag13` and `PetroMag` are shown here.

For each pair of bands (six colours, $Y - J$, $Y - H$, etc.), the sources were divided into four groups:

1. Not flagged as deblended in either band,

2. Flagged as deblended in the redder band only,
3. Flagged as deblended in the bluer band only,
4. Flagged as deblended in both bands.

A colour-magnitude diagram was then created for all colours and for the four different magnitudes.

Fig. 3.3 shows the colour-magnitude diagrams for the 24 arcsec diameter fixed-aperture magnitudes and the six colours. From the effects discussed above, the expectation is that the sources deblended in the redder band only should be slightly bluer on average, since some will have too much flux in the bluer band due to source confusion (blue contours) and that the sources deblended in the bluer band only should be slightly redder on average (red contours). Sources deblended in both bands should lie in a similar region to the non-deblended sources. These trends are seen for all six colours, suggesting no problems with the deblending algorithm.

Fig. 3.4 shows the six colour-magnitude diagrams for the Petrosian magnitudes. The same trend is to be expected as for the fixed-aperture magnitudes. However, the opposite trend is seen for the Y -, H - and K -bands. For example, if a source is deblended in Y but not deblended in H , its Petrosian magnitude is found to be much brighter (rather than fainter) in Y , and *vice versa*. Moreover, the brightest Petrosian magnitudes in any band are those of sources flagged as deblended.

It seems conclusive that the deblending algorithm is not working correctly for the Y -, H - and K -band Petrosian magnitudes. The deblending algorithm appears to be making the sources significantly brighter.

The J -band appears to be unaffected by this problem. This could be because the J -band data are microstepped so a different algorithm is being used. In the other bands, microstepping is never employed for the Y -band, but in the EDR the H - and K -band data were microstepped. If this is the reason why the J -band data are unaffected, then the Y -band should be more strongly affected than the H - and K -bands. This appears to be the case, from the $Y - H$ and $Y - K$ panels of Fig. 3.4, where the red contours (deblended in Y only) are offset more strongly than the blue contours (deblended in H or K only).

3.3 Comparison with deeper imaging: UKIDSS DXS

In UKIDSS DR2 there is a small area of just over a square degree with coverage in both the LAS and the DXS (Deep Extragalactic Survey) VIMOS 4 field.³ This makes it possible to compare

³This analysis was performed before DR3 was available and at a time when the proposed K -band magnitude limit was $K < 15$. The overlap region is the same between DR2 and DR3, but the overlap region in DR4 is more than three

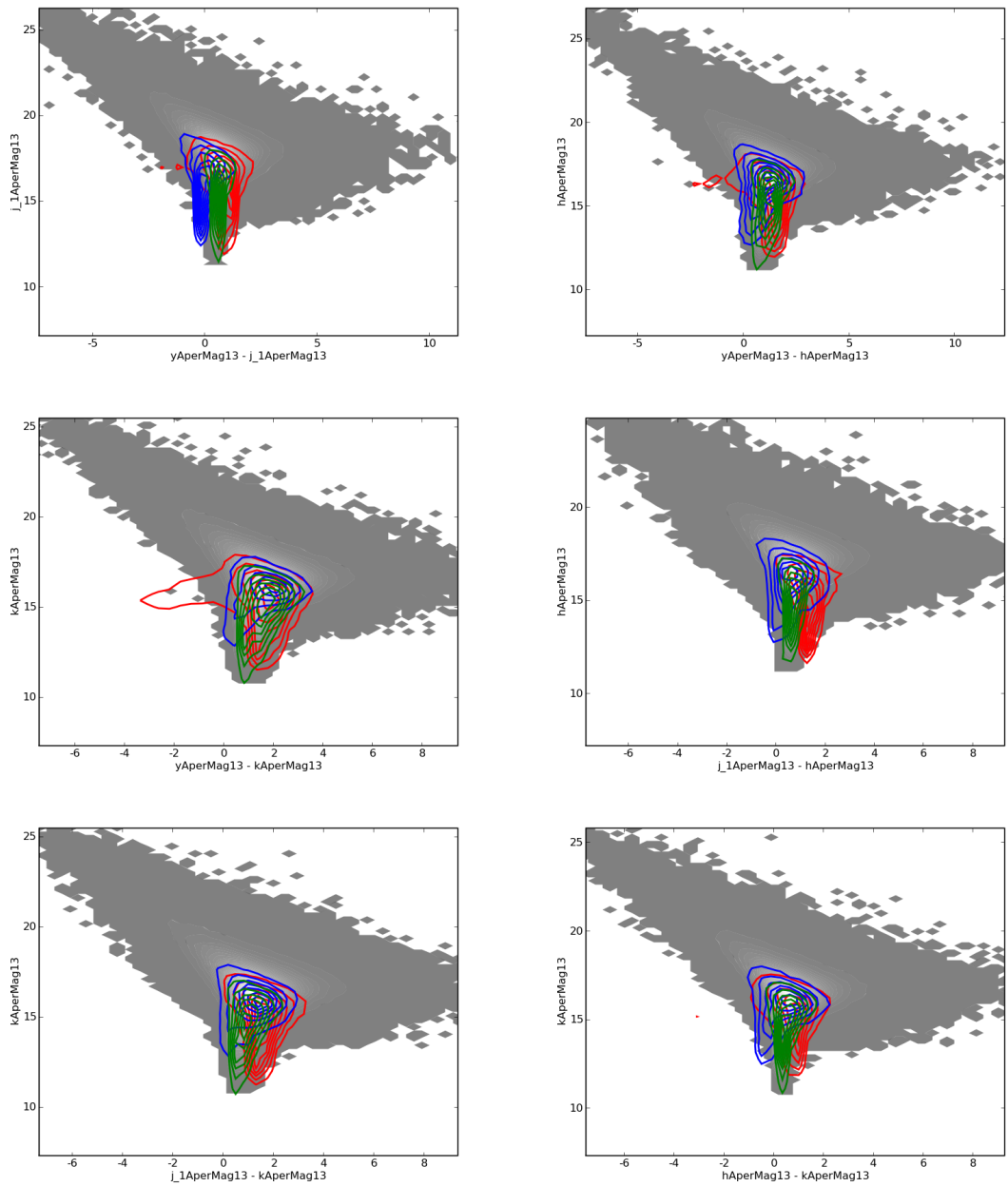


Figure 3.3: Colour-magnitude diagrams illustrating the effect of deblending on the 24 arcsec diameter fixed-aperture magnitudes. The shaded regions show the distribution of sources that are not flagged as debled in either band, with a darker shade of grey indicating a lower density of sources. Blue contours show the density of sources that are debled in the redder band only, red contours correspond to sources debled in the blue band only, while the green contours are for sources debled in both bands. If the deblending algorithm works, the blue contours should be slightly bluer than usual, the red contours should be slightly redder than usual, and the green contours should have similar colours to the shaded regions.

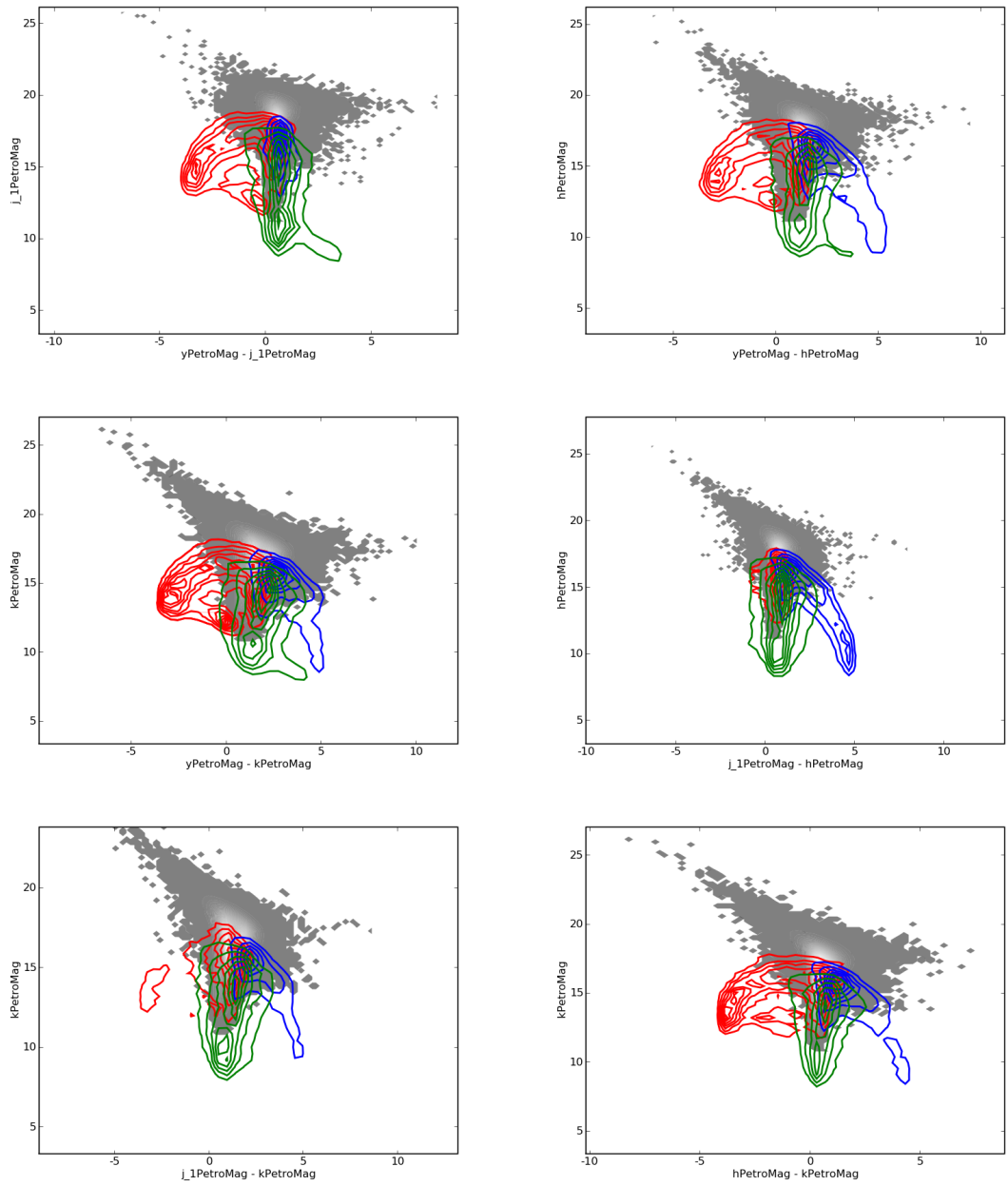


Figure 3.4: As Fig. 3.3 but for Petrosian magnitudes.

the LAS photometry with deeper K -band imaging, without needing to be concerned about many systematic effects (same camera, same telescope, same filter, same pipeline). In the LAS, this region is imaged in K only; in the DXS it is covered in J and K .⁴ This region is used to investigate the completeness of the LAS sample and the accuracy of the photometric measurements.

The database was queried for all sources in this region⁵ in the LAS or the DXS with K -band detections and K -band Petrosian magnitudes brighter than 15.5 mag (or no Petrosian magnitude), requiring a seamless sample. This yielded 3020 LAS sources and 9222 DXS sources. The two tables were then cross-matched within a radius of 2 arcsec giving 2414 matched sources. Deblending corrections have not been applied to the magnitudes in this sample.

Figure 3.5 shows a comparison of the quality of imaging, while the distances between the LAS source and the closest match in the DXS are shown in Fig. 3.6.

Table 3.1 shows the resulting numbers of sources, in the LAS, DXS or in the cross-matched sample, with various criteria applied to the samples. There is a higher density of sources in the DXS and fainter isophotes are probed, so deblending will be invoked more often than in the LAS. This may explain the sharp drop in the number of DXS sources found when $\text{ppErrBits} = 0$ is required.

3.3.1 Completeness

Considering the 327 sources detected in the DXS as galaxies ($\text{kClass} = 1$) with Petrosian magnitudes and no quality control flags, there are 132 with $K_{\text{Petro,DXS}} < 15$. The sample is restricted in this way because there may be sources with $15 < K_{\text{Petro,DXS}} < 15.5$ but $K_{\text{Petro,LAS}} > 15.5$, which would be excluded from this sample, and hence would bias the estimate of completeness. The sample is further limited by requiring Petrosian radius not to be clipped at 6 arcsec (30 pixels in the DXS), reducing the sample to 130.

Figure 3.7 shows the fraction of DXS galaxies detected in the LAS, as a function of DXS Petrosian magnitude. There is one source not detected at all, and a further seven sources detected in the LAS, but either having error flags set or not classified as galaxies.

From the DXS image for the one source not detected in the LAS, it was clearly a source

times as large.

⁴The DXS imaging for deep stacks in DR2 was not as good as it should have been; it was inadvertently processed using an older version of the CASU extractor. According to the WSA (<http://surveys.roe.ac.uk/wsa/releasehistory.html>, release note for 11 June 2007), ‘This version overestimates the background noise, so the faintest objects are not extracted. The magnitude limit is approximately 0.1 magnitudes too bright.’ However, this makes little difference to the analysis here, since it is still much deeper than the LAS.

⁵Precise boundaries are $334.9464237^\circ < \text{RA} < 335.7010139^\circ$ and $-0.1686847^\circ < \text{dec} < +1.4005635^\circ$, excluding the region with $\text{RA} < 334.9992479^\circ$ and $\text{dec} > 0.5138042^\circ$.

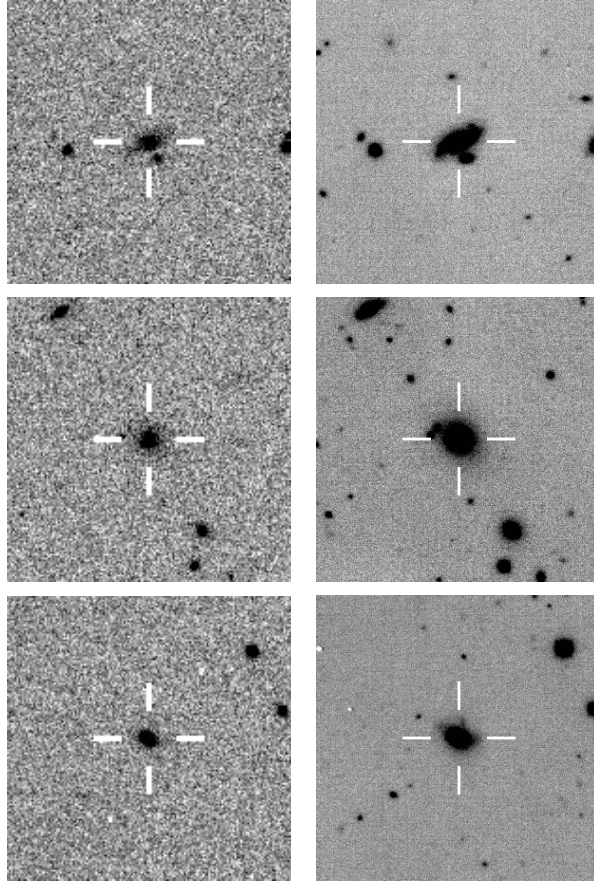


Figure 3.5: K -band images of galaxies seen both in the LAS (left-hand panels) and in the DXS (right-hand panels). The galaxies are drawn from the main LF sample, requiring matches to SDSS spectroscopic galaxies. The cross-hairs show the location of the source in the LAS.

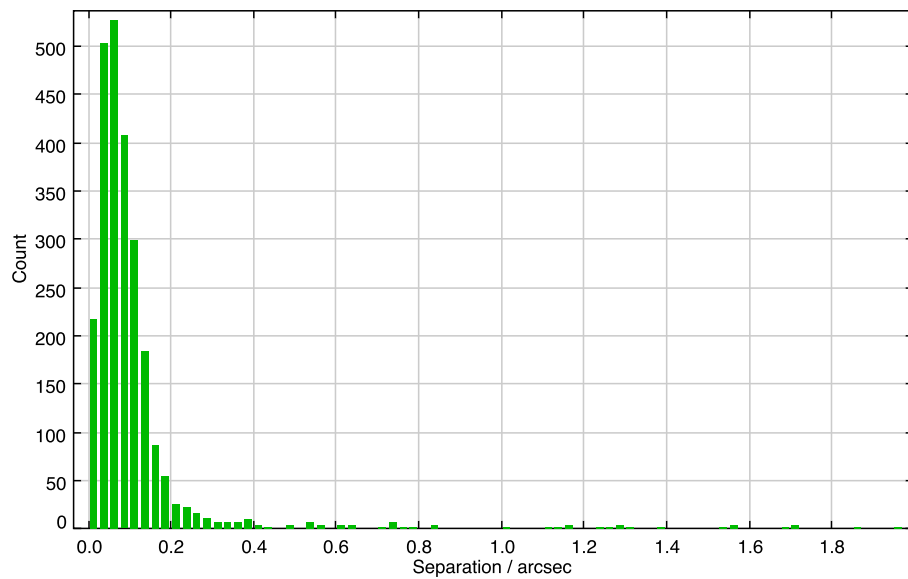


Figure 3.6: Separation between the closest matches between the LAS and the DXS in the overlap region.

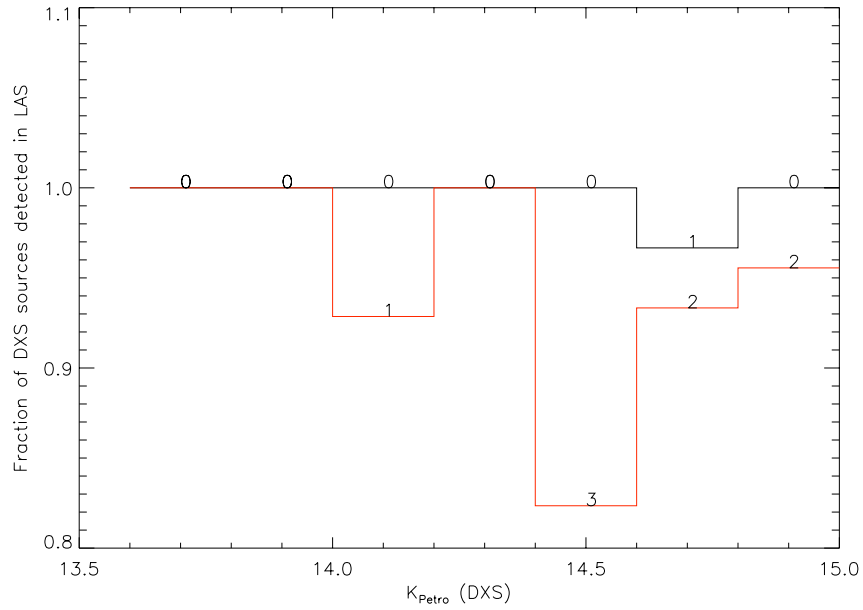


Figure 3.7: Completeness of the LAS, showing the fraction of DXS sources detected in the LAS as a function of $K_{\text{Petro,DXS}}$. The 130 DXS sources are chosen to have no error flags, $K_{\text{Petro,DXS}} < 15$, classified as galaxies and not having their Petrosian radius clipped (at 30 pixels). (One source has $K_{\text{Petro,DXS}}$ brighter than 13 mag and is not shown here; this source is also detected in the LAS.) The black histogram shows the fraction of sources with any detection (with $K_{\text{Petro}} < 15.5$) in the LAS while the red histogram shows the fraction of sources detected in the LAS with no error bits, Petrosian magnitudes and classified as galaxies ($k\text{Class} = 1$). The numbers on the histograms show the number of sources not detected in the LAS.

Table 3.1: Sources detected within the overlap region of LAS and DXS, as described in the text. Galaxies are identified using $kClass = 1$. Note that when stricter conditions are applied to the sample (e.g., galaxy), these conditions are not applied when a non-detection is indicated. For example, ‘LAS only’ for the row marked ‘+ galaxy’ means that there is no corresponding source in the DXS, whether classified as a galaxy or not. ‘TOTAL’ is given by $LAS + DXS - Both$.

	LAS	DXS	Both	LAS only	DXS only	TOTAL
All	3020	9222	2414	606	6808	9828
+ galaxy	1088	4668	604	425	4047	5152
+ $K_{Petro} > 0$	2684	3539	2228	398	1262	3995
+ galaxy + $K_{Petro} > 0$	824	1407	479	258	877	1752
no ppErrBits	2247	2704	1468	217	1202	3483
+ galaxy	610	536	274	109	250	872
+ $K_{Petro} > 0$	2222	1596	1468	192	94	2350
+ galaxy + $K_{Petro} > 0$	603	327	274	102	41	656
no ppErrBits, $K_{Aper3} > 0$	2247	2572	1468	217	1070	3351
+ galaxy	610	536	274	109	250	872

(galaxy), so it was not immediately obvious why it was not in the LAS. Looking at LAS images for that location, there was a source, but an error in the seaming process meant that the source was not included in the seamless catalogue. The DR2 LAS was subsequently re-seamed to rectify this.⁶

The remaining seven are as follows: four classified as stars in the LAS, two flagged as lying within a dither offset of the stacked frame boundary and one flagged as deblended. The four LAS stars are all quite possibly stars, from the DXS images, and the others do not reflect genuine incompleteness in the LAS (over 25 per cent of LAS sources in the overlap region have one or more error flags set).

Although the matched sample is very small, no evidence for incompleteness to $K_{Petro} < 15$ has been found.

3.3.2 Sources in both LAS and DXS

In this section, the 2228 sources in both LAS and DXS with $0 < K_{Petro} < 15.5$ will be considered. Requiring further that there are no error flags set ($kppErrBits = 0$), the sample is reduced to 1468

⁶<http://surveys.roe.ac.uk/wsa/releasehistory.html>, release note for 3 July 2007

Table 3.2: kClass value for 1468 galaxies in both LAS and DXS with $0 < K_{\text{Petro}} < 15.5$ and no error flags. The two surveys differ in their classification for 45 sources (3 per cent).

kClass	LAS		
	Star (-1)	Noise (0)	Galaxy (1)
DXS			
Star (-1)	1149	8	27
Noise (0)	0	0	1
Galaxy (1)	9	0	274

objects (the same sample is obtained by requiring $K_{\text{Kron}} > 0$ or $K_{\text{aperMag3}} > 0$).

Classification

UKIDSS sources are classified in each band separately and then in the merged table, once sources detected in the individual bands have been matched. For each individual band, there is a continuous classification (classStat) and a discrete classification (class), with each source labelled as either a star, galaxy, probable star, probable galaxy, noise or saturated. For the merged tables, the same is true, with a continuous classification (mergedClassStat) and a discrete classification (mergedClass). However, in the LAS the overlap region is imaged in K only, so the K -band and merged classifications are identical. The DXS is imaged in J and K in this region, so in general the kClassStat will not be equal to mergedClassStat. This is the case for the 6927 DXS sources which also have an J -band detection; however, for the 2295 DXS sources with K -band detections only, the kClassStat and mergedClassStat are equal. For consistency of results, only the K -band classifications will be considered.

The discrete classification generally follows the continuous classification, but may be overridden by other photometric or morphological factors, for example, the ellipticity or whether the object is near saturation.

Fig. 3.8 shows the continuous LAS and DXS kClassStat for the 1468 objects and demonstrates that the K -band classification in the LAS shows no evidence of being less reliable than the equivalent in the DXS.

Table 3.2 shows how the LAS and DXS discrete classifications compare with each other. There is general agreement between the two surveys, except that of the 302 sources (out of 1468) with kClass set to galaxy in the LAS, 28 (9 per cent) are classified otherwise in the DXS K -band classification (27 are classified as stars and 1 as noise).

From a visual inspection of DXS images of the 28 sources which may have been misclassified

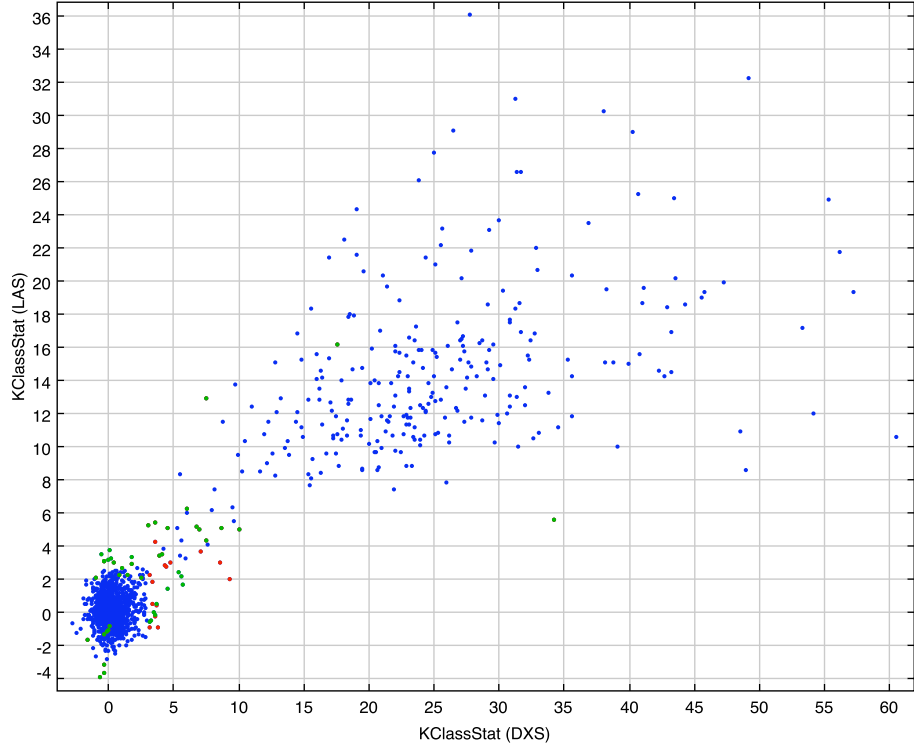


Figure 3.8: K -band continuous star/galaxy classification for the DXS and LAS in the overlap region, showing the 1468 sources with $0 < K_{\text{Petro}} < 15.5$ and no error flags. The objects are classified in the discrete classification as stars if $-2 \leq \text{kClassStat} \leq 3$ and as galaxies if $\text{kClassStat} \geq 3$, except for 23 sources in the LAS and 24 sources in the DXS (45 in total) where the classification does not follow this rule, due to other factors. These objects are coloured red in the plot. A different, but substantially similar, set of 45 objects is classified differently in the two tables. These are coloured green in the plot, and are drawn above the red dots, where an object would fall into both of these groups. The separations for these objects are less than 0.25 arcsec, making it unlikely that a mismatch has caused a disagreement over the classification. Overall, for 97 per cent of the objects, the discrete classification is the same in both LAS and DXS.

by the LAS, it is clear that the majority are indeed stars, often in regions with a very high density of sources or overlapping with other background sources. The conclusion is that a sample of galaxies selected using LAS kClass alone is likely to be contaminated with between 5 and 10 per cent being stars.

Further investigation could be made into the completeness of the LAS galaxy sample, since 9 sources are classified as galaxies in the DXS but as stars in the LAS. However, the incompleteness appears to be small, and since the kClass is not used to define the sample used later in this thesis, this is not important at this stage.

Photometry

Given the sample of 283 galaxies (kClass = 1) in the DXS (see Table 3.2), and excluding two sources which have maximum radii (15 pixels in LAS or 30 pixels in DXS; two sources in LAS, one of which has maximum radius in DXS as well), this leaves 281 galaxies for further analysis.

Figure 3.9 shows a comparison of various quantities between the two surveys. Assuming, for the time being, that the DXS quantities are accurate, it follows that the LAS Kron quantities are very unreliable, perhaps because of difficulties in determining the Kron radius. However, the Petrosian quantities are much better, with a mean offset between LAS and DXS (for $K_{LAS} < 15$) of 0.0120 mag, and a standard deviation of 0.1407 mag. Based on the median absolute deviation (MAD), which is more robust against outliers, the standard deviation is estimated to be 0.0676 mag (median magnitude 14.67). This is somewhat smaller than the uncertainty estimated from repeat observations (0.0975 mag at magnitude 14.67). However, the sample size used for the comparison with DXS is small (125 galaxies in this particular sub-sample) so the results are broadly consistent between the two methods.

Deblending

It was found in Section 3.2.2 that sources deblended in the LAS had their Petrosian fluxes over-estimated. If the DXS is unaffected by these problems then it would be possible to estimate the magnitudes of deblended LAS galaxies.

From Table 3.1 it can be seen that deblending is a significant factor. 6518 (9222 – 2704) DXS sources have at least one quality control bit flagged, and 5794 (57 per cent of the DXS sample) are flagged as deblended. There are 479 objects with $K_{Petro} > 0$ in both surveys and flagged as galaxies in both surveys, of which 274 are free from post-processing error bits in both surveys. Considering the 205 sources (479 – 274) with $K_{Petro} > 0$ in both surveys and flagged as galaxies in both surveys, but flagged with post-processing error bits in at least one of the surveys, there are

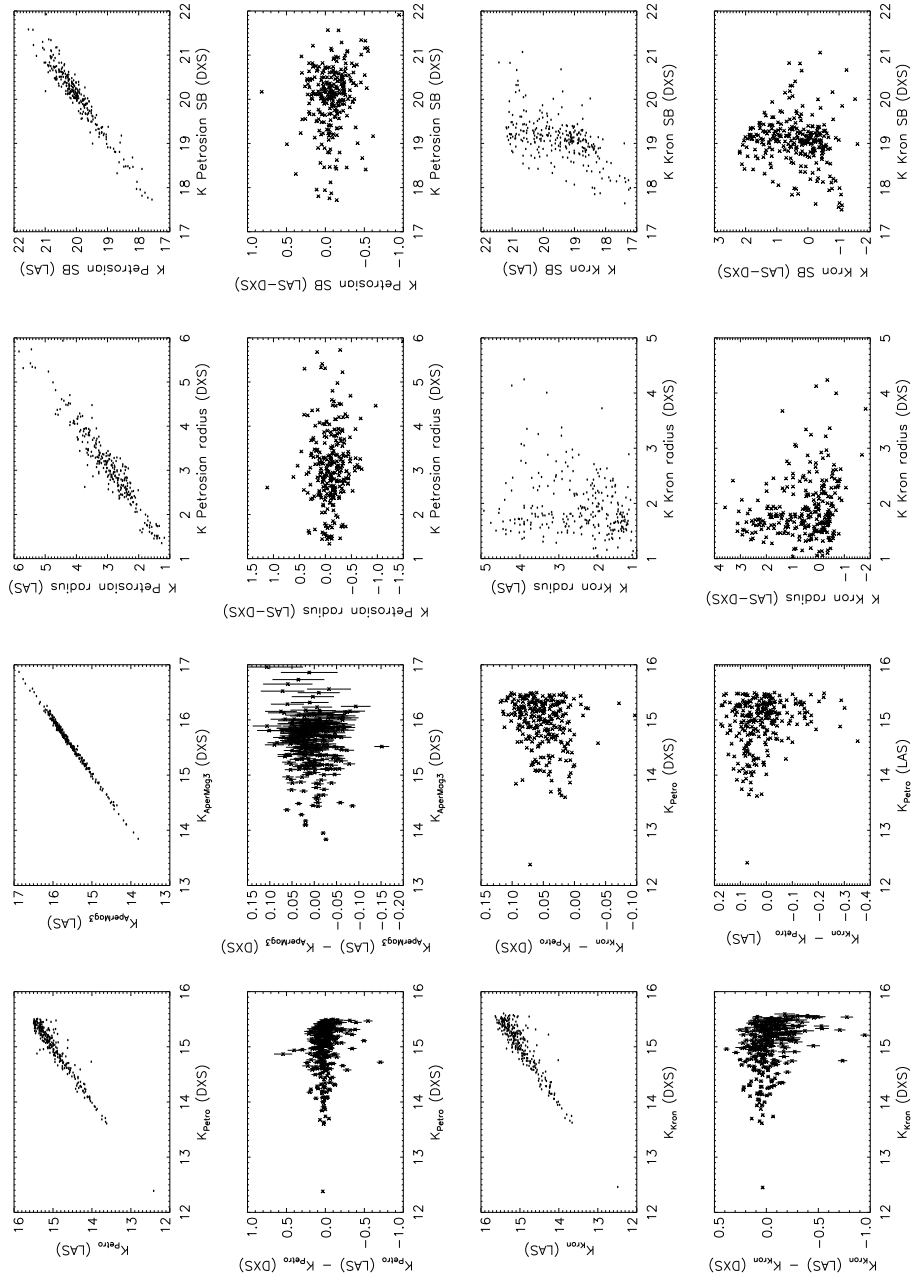


Figure 3.9: Comparison of properties of 281 DXS and LAS matched sources, classified as galaxies in the DXS, in the overlap region. Scatter plots are shown as points and residual plots as crosses, with error bars when available for DXS–LAS comparisons. SB is the mean surface brightness within the Kron or Petrosian aperture, given by $\mu = m + 2.5 \log(\pi(2r)^2)$, where m is the magnitude (Kron or Petrosian) and r is the Kron or Petrosian radius.

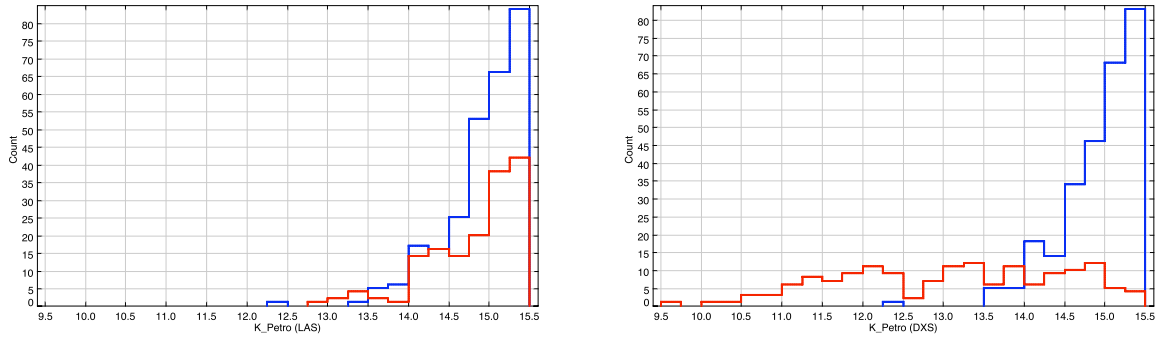


Figure 3.10: Number counts as a function of K -band Petrosian magnitude for LAS (left) and DXS (right) for sources flagged as deblended in the DXS but not in the LAS (red histogram) compared with sources not flagged as deblended in either (blue histogram).

163 with the deblending flag set in DXS but not set in LAS. 9 of these 163 have other flags set in one or both surveys, so 154 are considered further.

Figure 3.10 shows the Petrosian magnitudes for these 154 objects flagged as deblended in the DXS and in the LAS and for 274 sources not flagged as deblended in either survey. When the deblending algorithm is invoked in the DXS, the galaxy is likely to appear much brighter, as can be seen from the red histograms. This is the same problem as was found for deblending in the LAS.

Ideally, one would like to find a deep sample that has been deblended well and compare this with the LAS, which has been deblended badly. It could then be seen what fraction of galaxies (within a certain range in ‘true’ apparent magnitude) has been deblended badly.

3.4 Completeness and reliability of UKIDSS quantities

In this and the following sections, limits on the sample are estimated, within which the sample is assumed to be of uniform completeness, and within which the quantities are assumed to be sufficiently reliable.

3.4.1 Deblending

A significant fraction of the LAS sample is affected by deblending. This will affect the completeness of the sample, but, given the resultant uncertainty in the Petrosian magnitude and radius, it is not possible to quantify how this completeness varies as a function of ‘true’ apparent magnitude. It is assumed that deblending has an effect on the overall completeness, independent of the galaxy’s intrinsic properties. The possible effect of this assumption on the final results will be examined in Chapter 4.

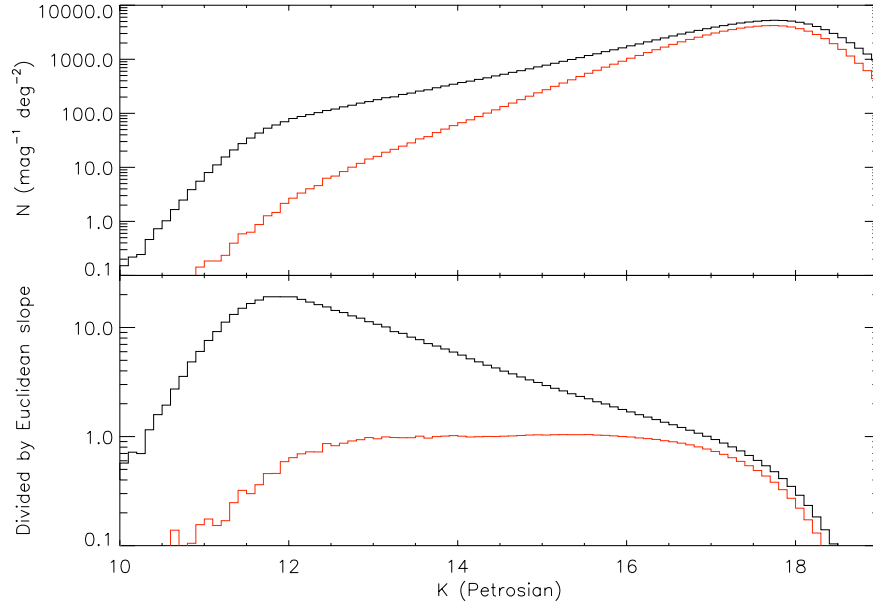


Figure 3.11: K -band number counts for the full DR3 LAS sample, requiring a seamless sample, with no quality control error bits flagged. The black (upper) lines show the number counts for the whole sample (15 079 199 objects), while the red (lower) lines are for sources classified as galaxies ($kClass = 1$; 9 962 258 objects). The upper panel shows the number of sources as a function of de-reddened Petrosian magnitude, while the lower panel shows the same data divided by a Euclidean slope with arbitrary normalization.

3.4.2 Apparent magnitudes and number counts

In Section 3.2.1 a completeness limit for the LAS of $K < 16$ was estimated from repeat observations, with magnitude errors of around 0.19 mag at $K = 16$, also estimated from repeat observations.

Tests have been run on the pipeline by CASU⁷ to estimate the quality of the pipeline outputs. In terms of completeness, they find a magnitude limit of $K = 18.4$ for 50 per cent completeness for stars, the limits for galaxies being approximately 1 magnitude brighter.

Fig. 3.11 shows the number counts as a function of K -band Petrosian magnitude for the whole UKIDSS LAS DR3. Incompleteness in the galaxy number counts becomes apparent fainter than $K = 16$, as evinced by a departure from the Euclidean slope, so the sample is limited to $K < 16$.

In order to estimate the volume probed, for the luminosity function estimate, it is important to consider both the faint and bright magnitude limits (Johnston et al., 2007). In this work, there are other constraints on the low-redshift visibility of the galaxies, for example, the limit in large

⁷<http://casu.ast.cam.ac.uk/documents/wfcam/simulations/index/view>

radius. No bright limit has been applied in the K -band; adding a limit as strong as $K > 13$ has a barely noticeable effect.

3.4.3 Size

As discussed in Section 2.4.5, there is a limit on the Petrosian radius of 6 arcsec. This means that any source with a true Petrosian radius greater than 6 arcsec will have its radius clipped at this value, and the quantity labelled as the Petrosian magnitude will in fact be an underestimate of the Petrosian flux.

To avoid complication, I have set a large radius limit, requiring the Petrosian radius to be less than 6 arcsec. It is worth noting that a significant fraction of the sample (6683 out of 43 939) have their Petrosian radii clipped at 6 arcsec. Including this limit explicitly has a negligible effect on the results, but it is likely to be significant when investigating the very low-redshift and low-luminosity galaxy population.

No small radius limit has been set, as the problem of misclassifying galaxies as stars is thought to be negligible in SDSS (Blanton et al., 2001), and no independent classification is required in the K -band.

3.4.4 Surface brightness

Tests on the UKIDSS source extraction (Cross et al., in preparation) suggest that, for a de Vaucouleurs profile, galaxies with effective surface brightness fainter than $19.5 \text{ mag arcsec}^{-2}$ are likely to have their fluxes and sizes underestimated. For well-defined sample limits, a cut in surface brightness should be imposed at that value. However, when investigating the space density of galaxies with high surface brightness, this limit can be safely ignored, since for the vast majority of the sample, the faint magnitude limit in r provides a stronger constraint on the visibility of the galaxy. So in order to include at least some low-surface brightness galaxies in the analysis, a limit of $\mu_{e,K} < 21 \text{ mag arcsec}^{-2}$ is chosen.

The combined limits in K -band Petrosian magnitude and Petrosian radius will impose a limit on the effective surface brightness for each galaxy. For Sérsic (1968) indices n between 1 and 4, the Petrosian radius is approximately twice the effective radius (Graham et al., 2005), so the faintest effective surface brightness will be given by

$$\mu_{e,K} \simeq 16 + 2.5 \log 2\pi(6 \text{ arcsec}/2)^2 \quad (3.26)$$

$$= 20.38 \text{ mag arcsec}^{-2} \quad (3.27)$$

using the limits in Petrosian magnitude and radius described above. Fainter than this, there is a

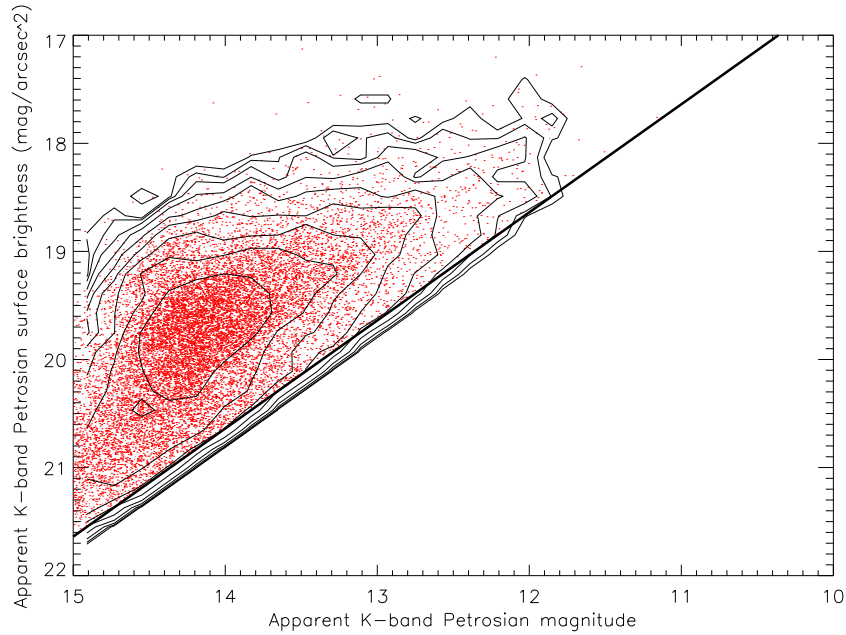


Figure 3.12: K -band Petrosian magnitude and surface brightness, showing galaxies from an earlier version of the LF sample, with a magnitude limit of $K < 15$. Contours show the number density of galaxies. The thick black line shows the limit on Petrosian radius.

sharp decrease in the number counts as a function of surface brightness. One can therefore expect to find significant incompleteness at low surface brightness.

Given the large-radius limit, there is a relation between the faint limit on Petrosian surface brightness (mean surface brightness within the Petrosian aperture) and the bright magnitude limit:

$$\mu_{\text{Petro},K} = K_{\text{Petro}} + 2.5 \log 4\pi r_{\text{Petro}}^2 \quad (3.28)$$

$$= K_{\text{Petro}} + 2.5 \log 4\pi (6'')^2 \quad (3.29)$$

$$= K_{\text{Petro}} + 6.64 \quad (3.30)$$

This is illustrated in Figure 3.12, which clearly shows the importance of including radius in the multivariate luminosity function: if any of the galaxies near the thick black line were significantly closer to us, they would not be included in the sample, so they are close to z_{min} , the minimum redshift at which they could be observed. It can also be seen that the limit in radius provides a much stronger constraint on the sample than the constraint provided by the faint surface brightness limit.

The limits on surface brightness will further be affected by the faint r -band magnitude limit; this will be discussed in Section 5.1.2.

3.5 Completeness and reliability of SDSS quantities

3.5.1 Spectroscopic target selection

The SDSS spectroscopic target selection criteria, discussed in Section 2.5, may introduce some incompleteness into the sample. Some very compact galaxies may be misclassified as stars, since galaxies are defined as objects for which $\Delta_{\text{SG}} \equiv r_{\text{PSF}} - r_{\text{model}} \geq 0.3$ (Strauss et al., 2002). Blanton et al. (2001) use a division of 0.242 rather than 0.3. They argue that extremely few galaxies will be excluded in this way, and their fig. 3 suggests that a cut at 0.3 will have a similarly small effect.

A small incompleteness will be introduced as a result of the SDSS ‘fiber collisions’; the effect of this will be discussed in Section 4.1.

3.5.2 Apparent magnitudes

The faint magnitude limit (after Galactic extinction corrections have been applied) for the majority of the SDSS sample is $r < 17.77$. However, some of the early observations used a brighter limit to define the spectroscopic sample. A significant fraction of the LAS survey area overlaps with these earlier SDSS fields so, following Blanton et al. (2001), we use a faint magnitude limit of 17.6 in r .

Bright limits are applied to the SDSS fiber magnitudes, since the spectroscopic sample requires these to be fainter than 15 in g and fainter than 14.5 in i (Strauss et al., 2002). These, however, have no effect on our results, since the strongest constraint on the low-redshift visibility of each galaxy is generally the large radius limit.

3.5.3 Surface brightness

The SDSS main galaxy sample has a limit of $\mu_{e,r} \leq 24.5 \text{ mag arcsec}^{-2}$ (Strauss et al., 2002). This limit is taken into account here, although it has a negligible effect on our results. Very few galaxies in our sample have $\mu_{e,r} > 23 \text{ mag arcsec}^{-2}$, so we assume the limit in SDSS surface brightness adds no further incompleteness to the sample, once the magnitude limits in r and K and the surface brightness limit in K have been considered.

Blanton et al. (2003b) have investigated the completeness of the SDSS sample at low surface brightness. They find 50 per cent completeness at $\mu_{e,r} = 23.36 \text{ mag arcsec}^{-2}$. Correcting for this incompleteness, the effect on the luminosity function is small and only noticeable at the faint end, where the sample used here is already affected by incompleteness in K -band surface brightness.

3.5.4 Redshifts

In this work, it is assumed that, within the redshift limits, any incompleteness will be independent of galaxy type. This is discussed further in Section 4.1.

The low- and high-redshift limits are chosen to give a large sample with useful redshifts. The redshifts will not be useful if (1) the galaxy has a significant peculiar velocity, meaning that the redshift cannot be used to give a reliable estimate of the distance to the galaxy, or (2) the redshift is large so that the uncertainties in the K - and evolution-corrections cannot be ignored.

Nearby, large galaxies may be accidentally deblended; a low-redshift limit would reduce this effect (Blanton et al., 2001).

The effect of the Earth’s peculiar velocity can be taken into account by converting the observed redshifts to a heliocentric, Galactocentric or Local Group-centred frame. (Heliocentric redshifts are provided by the SDSS pipeline.) In this work, Galactocentric velocity corrections have been applied (Loveday, 2000), which typically change the redshifts such that each galaxy in this sample is 0.005 ± 0.01 mag fainter, with some low-redshift galaxies ($z \gtrsim 0.01$) changed by almost 0.1 mag.

Blanton et al. (2005a) apply corrections to the redshifts to estimate the distances to nearby galaxies, out to $64h^{-1}$ Mpc ($z \simeq 0.021$ or $cz \simeq 6300 \text{ km s}^{-1}$). They convert to a Local Group-centred frame and then apply corrections for each galaxy based on the Willick et al. (1997) model of the local velocity field. Typical corrections to cz are of the order 200–300 km s^{-1} .

Jones et al. (2006) apply a correction to their sample to remove the effect of peculiar motions, which enables them to include galaxies to a low- cz limit of 750 km s^{-1} ($z = 0.0025$).

A low redshift limit of $z > 0.01$ ($cz > 3000 \text{ km s}^{-1}$) is chosen here to limit the effect of peculiar velocities, which are not taken into account. For redshifts less than this, there is a significant uncertainty in the distance to the galaxy because of the uncertainty in the peculiar velocity; at higher redshift the Hubble flow dominates. Note that this limit would need to be relaxed in order to sample galaxies with very low luminosity.

A high redshift limit of $z < 0.3$ is imposed to limit the effect of K - and evolution-corrections. This will be discussed in Section 4.2.1.

3.6 Summary

Table 3.3 shows the various limits on the sample.

In order to understand the possible effect of these limits on the final results, Figs. 3.13 and 3.14 show the observed and intrinsic properties of r -band and K -band galaxies. If the observed and

Table 3.3: Limits set on observed quantities, used to define the sample and to estimate the contribution of each galaxy to the space density.

Quantity	Minimum	Maximum
K Petrosian magnitude	-	16 mag
r Petrosian magnitude	-	17.6 mag
g fiber magnitude	15 mag	-
i fiber magnitude	14.5 mag	-
K Petrosian radius	-	6 arcsec
$\mu_{e,K}$	-	21 mag arcsec ⁻²
$\mu_{e,r}$	-	24.5 mag arcsec ⁻²
z	0.01	0.3

intrinsic properties do not correlate with each other, then placing limits on the observed quantities would not be expected to affect the final results significantly. For K -band galaxies, the low-redshift limit and the faint magnitude limit will restrict the visibility of galaxies with either faint (absolute) surface brightness or faint absolute magnitude. The limit on K -band Petrosian radius will exclude a large number of galaxies, but the effect seems to be independent of the intrinsic galaxy properties.

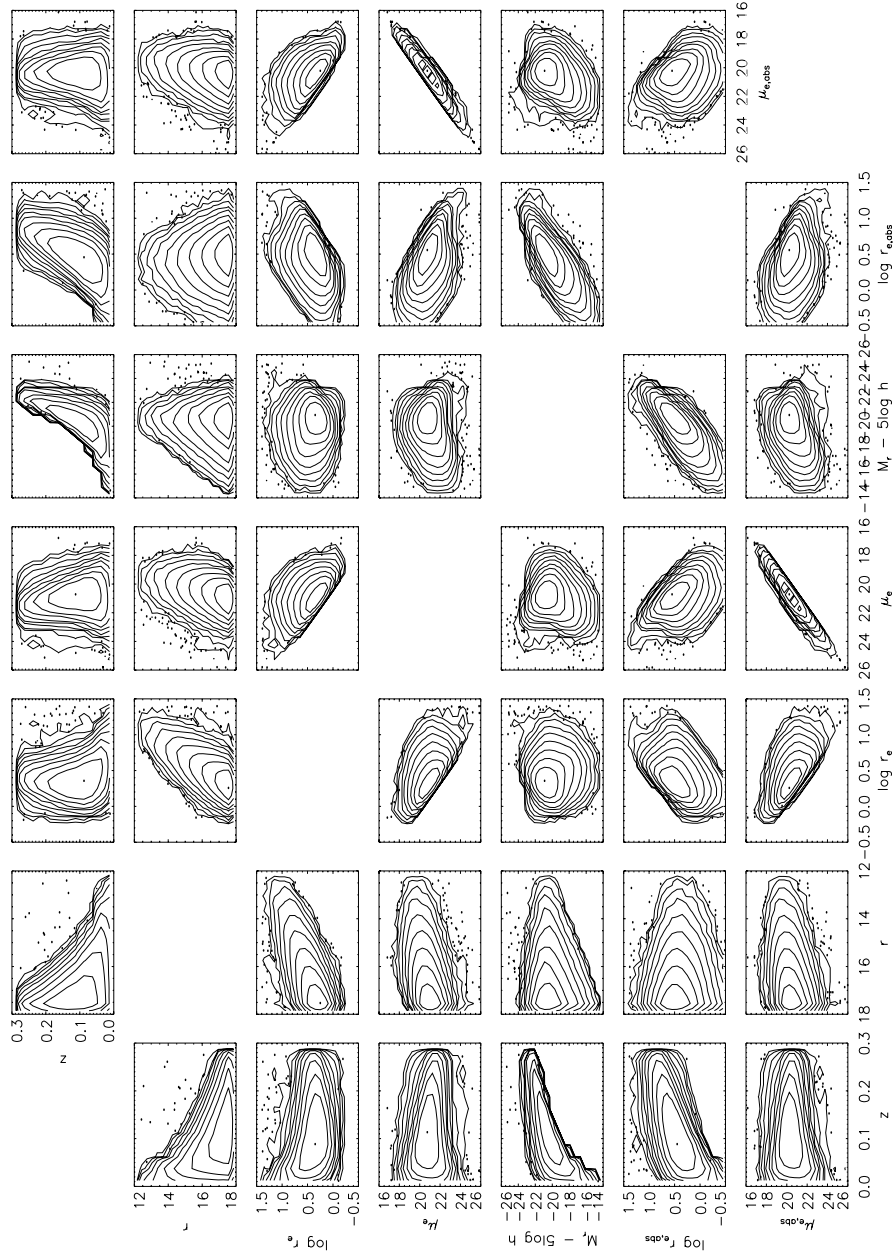


Figure 3.13: r -band galaxy properties, using the 446 197 target galaxies from SDSS DR5 with spectra and good redshifts ($z_{\text{Conf}} > 0.8$), and requiring $r < 17.77$ and $0.01 < z < 0.3$. The contours are spaced logarithmically and show the density of sources, with individual points shown where the density is low. Axes are (respectively, from top to bottom or left to right) redshift, Petrosian magnitude, half-light radius (arcsec), effective surface brightness (mag arcsec^{-2}), absolute Petrosian magnitude, physical half-light radius (h^{-1} kpc) and absolute effective surface brightness (mag arcsec^{-2}).

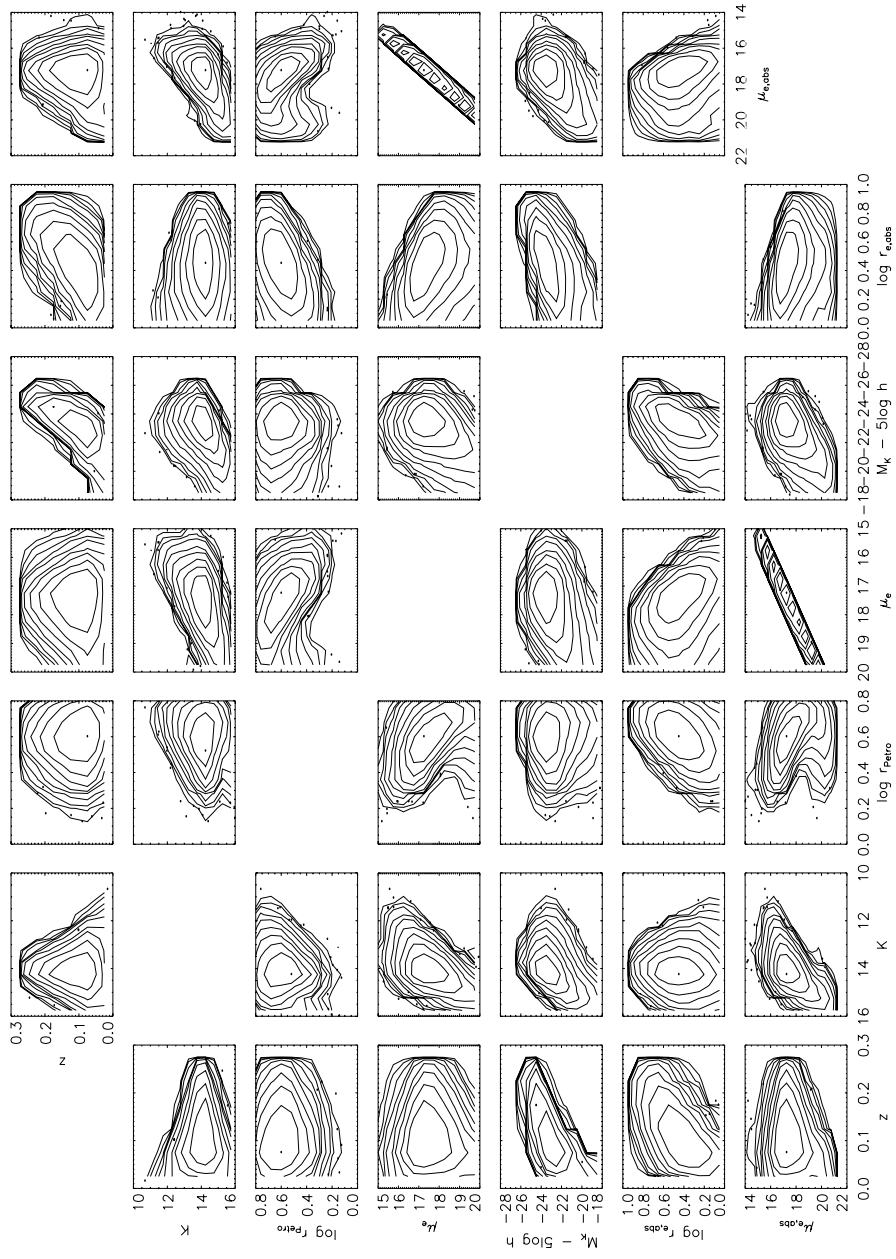


Figure 3.14: As Fig. 3.13 but for 36 659 K -band galaxies (the sample used in Chapter 5). Axes are (respectively, from top to bottom or left to right) redshift, Petrosian magnitude, Petrosian radius (arcsec), effective surface brightness (mag arcsec^{-2}), absolute Petrosian magnitude, physical half-light radius (h^{-1} kpc) and absolute effective surface brightness (mag arcsec^{-2}).

Chapter 4

Luminosity functions and beyond: estimating the properties of the underlying galaxy population

*Deux choses instruisent l'homme de toute sa nature:
l'instinct et l'expérience.*

Robust statistical methods are required to analyse a large sample of galaxies. The data were introduced in Chapters 2 and 3, various methods will be introduced and examined in this chapter, while the data and the methods will be combined in Chapter 5.

Two supplementary issues are addressed first, giving the following outline for the chapter:

1. Estimating the area covered by the sample.
2. Estimating the intrinsic properties of each galaxy from the observed properties.
3. Estimating the properties of the whole galaxy population from the properties of the observed sample.

4.1 Estimating the area covered

The volume, and hence the area, must be well estimated for the normalization of the luminosity function and related quantities. Given the complex geometry of the overlap region between the two surveys, it would not be easy to estimate the area directly. Nor would it necessarily be preferable: if the survey area is small, and the area is used to find the normalization, then this will be subject to fluctuations in the overall density.

One way to estimate the area of the matched UKIDSS–SDSS sample is to use the area of the SDSS sample and then scale it down according to the number of sources in the matched sample. But when doing this it is important to scale the area only by those factors that concern the area and not by those factors that concern the minimum and maximum redshifts for each galaxy (e.g., magnitude limits). This is the method used here: the area is estimated by dividing the number of galaxies in the matched sample by the number density of sources (per square degree) in the SDSS sample.

A substantial fraction of the region of overlap between the two samples was surveyed in the SDSS Early Data Release, so I use the limit of $r < 17.6$ (de-reddened Petrosian magnitude) rather than $r < 17.77$ used in later versions of the SDSS main galaxy selection algorithm. This corresponds to 635 320 target galaxies over the whole area.

The number density is estimated using the total area of the SDSS imaging survey (8000 square degrees) and the number of galaxies targeted for spectroscopy in the SDSS main galaxy sample. But first the size of this sample must be corrected for those objects included that are not galaxies. In SDSS DR5, the number of target galaxies with good spectroscopic redshifts ($z\text{Conf} > 0.8$) and with de-reddened Petrosian magnitudes brighter than $r = 17.6$ is 391 052. Of these, 384 617 are spectroscopically classified as galaxies. This suggests that around $(391\,052 - 384\,617)/391\,052 = 1.6$ per cent of the target galaxies are not in fact galaxies. Taking this into account gives a corrected target sample size of 624 865, giving a source density of 78.11 galaxies per square degree.

The two data sets were matched using the WFCAM Science Archive (WSA).¹ The initial sample was found by selecting all matches labelled as primary detections (i.e., no duplicates in the overlap regions for either survey), with good, nonzero spectroscopic redshifts (SDSS $z\text{Conf} > 0.8$) and having no major quality control issues flagged in K (UKIDSS $k\text{ppErrBits} < 256$), yielding a sample of 108 442 objects. Some (15 pairs) of these were found to be duplicates, where one SDSS object had two spectra, so the object with lower-confidence redshift was removed, leaving 108 427. Of these, only the closest matches within 2 arcsec are used, and the matched SDSS object is required to be classified spectroscopically as a galaxy, leaving a sample of 49 255 galaxies. The sky coverage of this matched sample is shown in Fig. 4.1.

This number will be affected by failed detections, which may introduce some bias into the sample. One type of failed detection is when redshifts have not been obtained, due to lack of coverage, failed redshifts or lack of available fibres to measure the spectra (‘fiber collisions’). These are assumed to introduce no bias into the sample, although it has been noted (Blanton et al., 2003a, 2005b) that the SDSS fiber collisions lead to slight incompleteness at high-density regions,

¹<http://surveys.roe.ac.uk/wsa/>

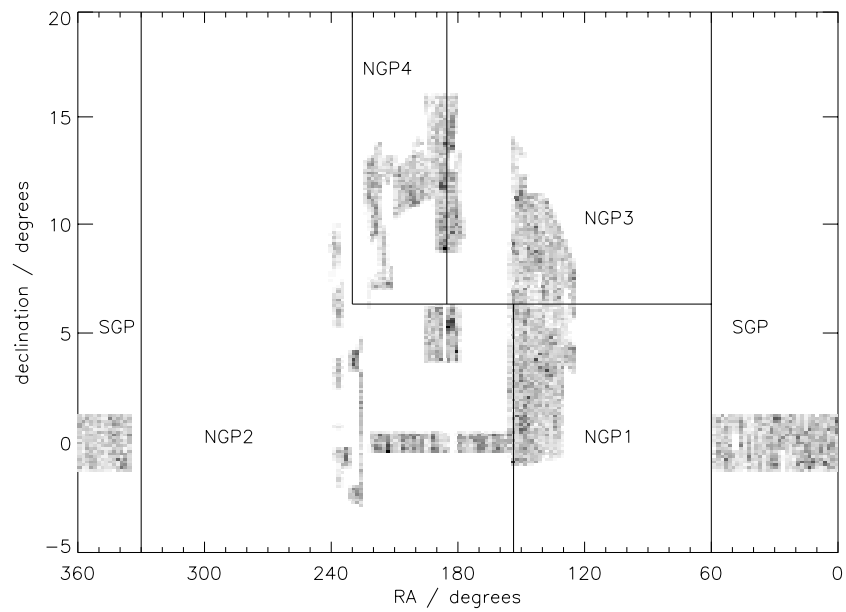


Figure 4.1: Sky coverage of the sample, showing the principal regions used for the jackknife samples (see Section 4.4), each of which is further subdivided into strips in RA (4 in the NGP4 region and 5 in the others) giving 24 jackknife regions in total, each containing approximately the same number of galaxies. A darker shade of grey corresponds to a higher density of sources at that point.

which may introduce a small bias against the type of galaxy found in such environments.

Another kind of failed detection is when there is a problem with the UKIDSS imaging. During the course of this work a problem was discovered relating to the source extraction deblending algorithm. It was found that, for Petrosian magnitudes in Y , H and K , the deblender, when invoked, was making the source significantly brighter, often by around 2 mag. (The J -band data are micro-stepped, unlike Y , H and K , which may explain why this problem is not seen for J -band Petrosian magnitudes; see Section 3.2.2.) Given that Petrosian magnitudes are used in this analysis, I have removed from the sample those sources flagged as deblended. This is achieved by setting the quality error bits flag (introduced in UKIDSS DR2) to zero. Of the 49 255 sources in the matched sample, 46 sources are flagged as having bad pixel(s) in the default aperture and another 4835 (almost 10 per cent) are flagged as deblended, leaving 44 374 in the sample.

Fig. 4.2 shows the r -band absolute magnitude of galaxies in the whole sample, and of sources excluded because of problems with the UKIDSS imaging. It can be seen that the galaxies affected by deblending are preferentially those with a high r -band luminosity, affecting as much as a third of high-luminosity galaxies, and possibly reflecting the observation that the most luminous galaxies are found in dense environments. This suggests that the number density of galaxies with $M_r - 5 \log h < -22$, corresponding very approximately to $M_K - 5 \log h < -25$, could be underestimated by as much as 50 per cent. However, only 793 of the 4835 sources flagged as deblended in K are flagged as deblended by SDSS; this could suggest a problem with the r -band fluxes, which could be overestimated due to under-deblending. It is therefore at least plausible that the luminosity-dependent nature of this cut could be illusory, so I assume that it introduces no bias into my final results. This will be discussed further in Section 5.2.

A small number of the remaining galaxies have very large magnitude errors, greater or much greater than 0.15 mag. So one final cut is to remove sources with large uncertainty in magnitude, in order to restrict the systematic errors in the results. Of the 44 374 galaxies remaining, those with magnitude errors greater than 0.15 mag in r or K are removed (55 in K , 380 in r and 0 in both) leaving 43 939 galaxies. Given the small number affected by this cut, and from Fig. 4.2, it is assumed that any bias induced by this cut will be negligible.

By estimating the area in this way, the assumption is that all SDSS target galaxies would be detected in the LAS, if that part of the sky has been surveyed. If this is not the case, it will have two effects: (1) particular types of galaxies will be underrepresented in the sample (those within the SDSS completeness limits but outside the limits for the LAS) and (2) the overall normalization will be too high, as the area and hence the volume probed will be underestimated.

Fig. 4.3 shows r -band Petrosian magnitude and the $r - K$ Petrosian colour (note that the

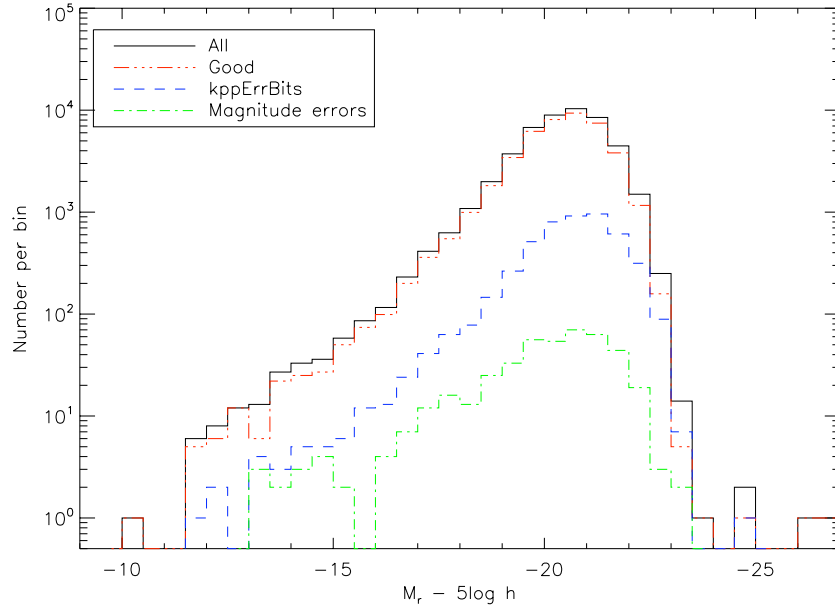


Figure 4.2: r -band absolute magnitude of the sources in the entire sample (‘All’, 49 255), the final matched sample (‘Good’, 43 939), those excluded due to poor K -band imaging, mostly related to deblending (‘kppErrBits’, 4881) and those excluded due to uncertainties in the Petrosian magnitudes (‘Magnitude errors’, 435).

apertures are not the same in r and K so this is not a true colour) for the sources in the matched sample. From the figure it can be seen that (1) there are likely to be very few sources at all lying within the SDSS flux limit but outside the K -band limit, and (2) many of these sources are detected anyway, since there are many sources detected fainter than the (nominal) K -band limit. This suggests that the effect on the overall normalization will be negligible, although the colour-dependent bias will be considered later.

The effective area can now be given as $43\,939/78.11 = 562.54 \text{ deg}^2$.

Calculating the area in this way takes into account any other *random* (not bias-inducing) sources of incompleteness that have not been considered here explicitly.

Note that the final sample will be smaller yet, due to limits imposed on magnitudes, radius and surface brightness, but this is related to the redshift limits (which affect the volume probed) rather than the area covered.

4.2 Intrinsic galaxy properties

The observed properties of galaxies are rarely of direct interest; more relevant are the intrinsic properties, for example, luminosity rather than apparent magnitude. In this section I describe how

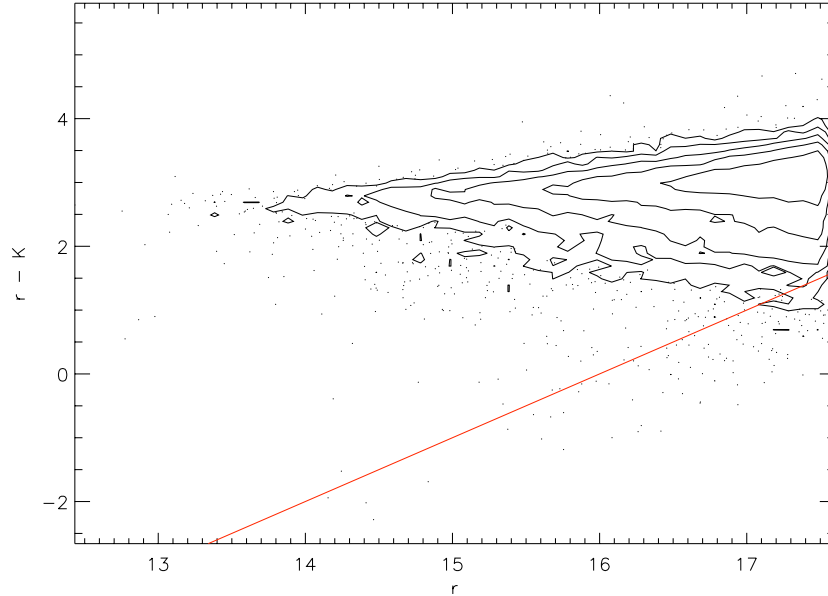


Figure 4.3: r -band Petrosian magnitude and $r - K$ Petrosian colour for 43 939 galaxies in the matched sample, before imposing a cut at $K < 16$. Objects lying below the solid (red) line have $K > 16$ and will not be included in the final analysis. Contours are spaced on a logarithmic scale, with sources shown as points where the density is low.

the intrinsic galaxy properties are derived from the observed properties. Only those properties related to whether a galaxy will be detected are discussed here; other derived quantities, such as the stellar mass, are described elsewhere.

4.2.1 K - and evolution-corrections

Two redshift-dependent corrections are conventionally made for observations of galaxies. First, the **K -correction** (Hogg et al., 2002) accounts for the fact that light emitted at a certain wavelength will be observed at a different wavelength, due to the redshift. So, for example, the observed-frame K -band magnitude will generally be different from the rest-frame (emitted) K -band magnitude. The second correction is the **evolution correction**, which accounts for the fact that the light from distant galaxies was emitted in the past, and that the galaxy will have evolved since that time. The correction is made to estimate the properties the galaxy would have at the present time.

K -corrections (to $z = 0$) are estimated using KCORRECT version 4.1.4 (Blanton & Roweis, 2007). The five optical Petrosian magnitudes ($ugriz$), before the evolution corrections have been applied (see below), are used to fit galaxy templates to each galaxy, from which the K -correction is derived in r and K . WFCAM filter files for KCORRECT have been generated from Hewett et al. (2006). Due to the inconsistent Petrosian apertures between UKIDSS and SDSS bands, good

optical–near-infrared galaxy colours are not available and so I have not succeeded in using the full set of bands (*ugrizYJHK*) for the template fitting.

When covering a significant range in redshift, it is important to include evolution corrections. This is because the bright end of the LF will consist mainly of galaxies at high-redshift, while the faint end will be made up from galaxies at low redshift. A failure to include evolution corrections leads to a distortion in the shape of the luminosity function.

Evolution corrections are applied using a simple $E(z) = Qz$ model, where Q for *ugriz* is taken from Blanton et al. (2003b) and we take $Q = 1$ for K (consistent with stellar population synthesis models, Blanton et al., 2003b, Section 6.2), i.e., $Q = (4.22, 2.04, 1.62, 1.61, 0.76, 1.0)$ for *ugrizK*. Note that these evolution corrections are very simplistic, since different kinds of galaxies evolve in different ways, but the correction is small: at $z = 0.3$, the K -band evolution correction is $Qz = 0.3$ mag.

Fig. 4.4 shows the K - and evolution-corrections in the K -band and r -band, also showing the K -band corrections used by Bell et al. (2003b). Their K -corrections are stronger than those used here, but their value of $Q = 0.8$ is weaker than $Q = 1.0$ used here. These two largely cancel each other out, with the $K(z) - E(z)$ used here being approximately the same as theirs for low redshift.

Others have used a simple analytic form to describe the K -correction. For example, Loveday (2000) used $K(z) = -2.5z$ for his sample of $z < 0.15$ galaxies, while Glazebrook et al. (1995) used a functional fit to various model K -corrections generated from stellar population synthesis models of Bruzual & Charlot (1993), for $z \lesssim 1.5$:

$$K(z) = \frac{-2.58z + 6.67z^2 - 5.73z^3 - 0.42z^4}{1 - 2.36z + 3.82z^2 - 3.53z^3 + 3.35z^4} \quad (4.1)$$

Where K -corrections for a ‘typical’ galaxy are required, these are determined from the observed sample as follows. Galaxy templates in KCORRECT are constructed by a linear combination of five basis templates. KCORRECT returns five coefficients for each galaxy, which give a weighting for each of the basis templates. A ‘typical’ galaxy is constructed from the whole sample by finding the median of each of these five coefficients. This is then used to find the K -correction for a typical galaxy.

4.2.2 Absolute magnitude

Absolute magnitudes are given by

$$M = m - DM(z) - K(z) + E(z) \quad (4.2)$$

where M and m are respectively the absolute and apparent magnitudes (in the same band), $DM(z)$ is the distance modulus, $K(z)$ is the K -correction and $E(z)$ is the evolution-correction.

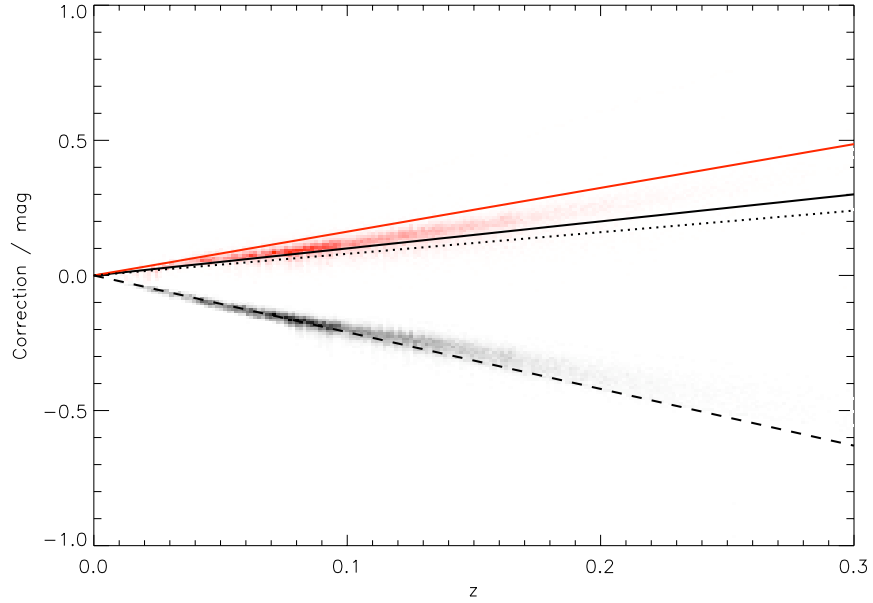


Figure 4.4: K - (shaded regions) and evolution- (solid lines) corrections in the K - (black, lower) and r - (red, upper) bands. Also shown are the mean K -correction (dashed line, $K(z) = -2.1z$) and evolution-correction (dotted line, $E(z) = 0.8z$) from [Bell et al. \(2003b\)](#).

4.2.3 Physical size

K - and evolution-corrections are not applied when estimating the physical radius, which is given in h^{-1} kpc as

$$R = \frac{1000\pi r D_A(z)}{180 \times 3600} \quad (4.3)$$

where r is the angular size in arcsec and $D_A(z)$ is the angular diameter distance in h^{-1} Mpc.

4.2.4 Intrinsic surface brightness

The absolute surface brightness, in mag arcsec^{-2} , is given by

$$\mu_{\text{abs}} = \mu - 10 \log(1 + z) - K(z) + E(z). \quad (4.4)$$

where μ is the apparent surface brightness. Note that in the K -band, for all galaxies in the sample, the absolute surface brightness is not more than 0.5 mag brighter than the apparent surface brightness.

These expressions for M , R_e and $\mu_{e,\text{abs}}$ can be combined, using

$$\mu_e = m + 2.5 \log(2\pi r_e^2) \quad (4.5)$$

along with the equation for the distance modulus (distances measured in h^{-1} Mpc)

$$DM(z) = 5 \log(D_L(z)/10 \text{ pc}) \quad (4.6)$$

$$= 5 \log(100\,000 D_L(z)) \quad (4.7)$$

and the relationship between the luminosity and angular diameter distances

$$D_L(z) = (1+z)^2 D_A(z) \quad (4.8)$$

to obtain (Driver et al., 2005):

$$\mu_{e,\text{abs}} = m + 2.5 \log(2\pi r_e^2) - 10 \log(1+z) - K(z) + E(z) \quad (4.9)$$

$$= M + DM(z) + 2.5 \log \left(2\pi \left(\frac{(180 \times 3600) R_e}{1000\pi D_A(z)} \right)^2 \right) - 10 \log(1+z) \quad (4.10)$$

$$= M + 5 \log \left(\frac{D_L(z)}{10 \text{ pc}} \right) + 2.5 \log(2\pi R_e^2) + 5 \log \left(\frac{648}{\pi D_A(z)} \right) - 10 \log(1+z) \quad (4.11)$$

$$= M + 5 \log(100\,000(1+z)^2 D_A(z)) + 2.5 \log(2\pi R_e^2) + 5 \log \left(\frac{648}{\pi(1+z)^2 D_A(z)} \right) \quad (4.12)$$

$$= M + 2.5 \log(2\pi R_e^2) + 5 \log \left(\frac{64\,800\,000}{\pi} \right) \quad (4.13)$$

$$= M + 2.5 \log(2\pi R_e^2) + 36.57 \quad (4.14)$$

This is the absolute surface brightness in mag arcsec^{-2} ; the absolute surface brightness in mag kpc^{-2} is given by subtracting 36.57.

4.3 Statistical methods for investigating the underlying population

Given some observed data, what can be inferred about the underlying galaxy population? Statistical methods to solve this **inverse problem** are explored in this section.

4.3.1 Bayes' theorem

Given a model for the galaxy population, with (currently unknown) parameters θ , and given a set of observations of the galaxy population, D , the parameters for the model may be found using Bayes' theorem,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \quad (4.15)$$

where $P(D|\theta)$ is the **likelihood**, $P(\theta)$ represents the **prior** knowledge of the model parameters, and $P(D)$ is a normalization constant. $P(\theta|D)$, the **posterior**, returns the probability of any set

of parameters, which makes it possible to find the most likely set of parameters along with the uncertainty in the parameters.

In this case, the data may be thought of as a multivariate histogram, $n(\dots)$, describing the number of galaxies observed per unit solid angle with a particular set of observable properties (e.g., apparent magnitude, angular size, apparent surface brightness, colour and redshift). The actual galaxy population may also be represented as a multivariate histogram, $\phi(\dots)$, giving the number of galaxies per unit volume with a particular set of intrinsic properties (e.g., luminosity, physical size, absolute surface brightness, rest-frame colour and redshift), the **space density** of galaxies. A simple example of this is the luminosity function, $\phi(L)$. The model then has two components: a parametrized description of ϕ and a parametrized description of the **selection function**, $f(\dots)$.

The selection function gives the probability that a galaxy will be observed, given its observable properties. It is generally modelled as a multivariate step function, for example, such that the probability of observing a galaxy with flux brighter than some limit is assumed to be unity, and galaxies fainter than that limit are ignored. The purpose of Chapter 3 was to find the selection function.

Given a model for ϕ and f , a prediction can be made for n . Generally, this will take the following form:

$$n(\dots) d\dots d\Omega = \phi(\dots)f(\dots) d\dots dV \quad (4.16)$$

where $d\Omega$ is the solid angle element and dV is the volume element.

The prior usually does nothing more than describe the range of each parameter under consideration, which must be finite for practical reasons. However, when the parameters are theoretically motivated, there will be some prior knowledge about their values, but generally the parameters are phenomenological, so there is no *a priori* reason to favour one set of values above another.

When the prior is trivial, the problem of finding the most likely set of parameters, θ , becomes a matter of maximizing the likelihood. This may be achieved either by an iterative exploration of the parameter space, to find the set of values that gives the maximum likelihood, or by using a maximum likelihood estimator, to give the parameter values without an iterative process. An example of the former would be the stepwise maximum likelihood (SWML) method for finding the luminosity function, and an example of the latter would be the $1/V_{\max}$ luminosity function estimator, both described in Section 4.6.

It is important to recognize that significant assumptions are made, not only in the choice of prior, but also in the choice of model. For example, by considering only a certain set of galaxy properties, the assumption is made that the selection function is independent of the galaxy properties that have not been included. If this assumption is not correct, it can lead to biased results.

4.3.2 Modelling the underlying population

The model for ϕ can take various forms, as follows.

Binned models

ϕ can be represented as a simple binned multivariate histogram, with a free parameter for each bin. An example of this would be a binned estimate of the luminosity function. This has the advantage that no particular form is assumed for ϕ , allowing the data to give inspiration for theoretical models. However, when considering several galaxy properties simultaneously, either the number of bins becomes very large or a very poor resolution has to be used.

Phenomenological models

Alternatively, a functional form may be assumed for ϕ , such as a Schechter function for the luminosity function (see Section 4.5). The parameters of this function are then tuned to give the maximum likelihood. This has various advantages: the galaxy population may be described by a small and memorable set of parameters, while a multivariate space may be described by a manageable number of parameters. However, the disadvantage is that the true distribution is always more complex than the assumed functional form; deviations from this form can have a significant effect on the results.

If ϕ has a simple analytic form, it is often possible to derive simple analytic predictions for the observed distribution.

Physical models

Perhaps of most scientific interest, and most difficult in practice, is for ϕ to be given a form derived from physical principles. In this case, the model would have physically meaningful parameters, and the most likely set of parameters would give direct information about the physical processes that have shaped the galaxy population.

The form for ϕ may be derived either analytically, e.g., from consideration of the halo model (e.g., [Cooray & Milosavljević, 2005](#)), or from the results of simulations and semi-analytic models (e.g., [Benson et al., 2003](#)). Typically this will involve two stages, with the physical properties of the galaxies estimated first (e.g., stellar mass, star formation history), then the observable properties of the galaxies (e.g., K -band luminosity) derived from stellar population synthesis models (e.g., [Bruzual & Charlot, 2003](#)).

4.3.3 The likelihood

Given a specific model (with certain values for the parameters), and given some observed data, what is the probability of these data being observed if the model is correct? This is the **likelihood**, $P(D|\theta)$.

For binned data, Poisson statistics can be used. Let n_i be the predicted number of galaxies (not necessarily an integer) in bin i . Then the probability of observing x_i galaxies in bin i is

$$P(x_i|\theta) = \frac{n_i^{x_i} e^{-n_i}}{x_i!}. \quad (4.17)$$

So the likelihood is

$$P(D|\theta) = \prod_i P(x_i|\theta) = \prod_i \frac{n_i^{x_i} e^{-n_i}}{x_i!} \quad (4.18)$$

For data that have not been binned, the likelihood is obtained by finding the probability that each individual galaxy is drawn from the expected distribution. This is used in the STY and SWML luminosity function methods (Section 4.6), and may be extended to incorporate uncertainties in the galaxy magnitudes (Blanton et al., 2003b).

4.3.4 Model selection

The Bayesian approach is strongest when two models are being compared, and it makes it possible to say which model is favoured by the data. This approach takes into account the predictive power of each model, so a contrived model will generally be disfavoured when compared with a more parsimonious model.

Bayesian model selection is being used increasingly in cosmology (Liddle et al., 2006) but has not been used widely in statistical studies of galaxies, where the models are more complex than in cosmology. However, it is possible to think of various applications of these techniques to galaxy studies.

When comparing two different functional forms for the underlying distribution, Bayesian model selection could be used. For example, a bimodal distribution could be detected by comparing a single Gaussian with a double-Gaussian model. Or an upturn at the faint end of the luminosity function could be detected by comparing a Schechter function with a double-Schechter function.

With physically-motivated models, this Bayesian approach could be used to compare one semi-analytic model with another. It could be asked whether the data warrant the inclusion of an additional form of feedback in the model, for example.

4.3.5 Bayesian estimate of galaxy magnitudes

In a magnitude limited survey of the depth of SDSS or the UKIDSS LAS, galaxy differential number counts increase steeply towards fainter magnitudes. Therefore, in a magnitude-limited sample, a galaxy drawn at random is much more likely to have a faint magnitude than a bright magnitude. In the presence of magnitude errors, this will have a systematic effect on the observed magnitudes, compared with the true magnitudes.

In order to quantify this effect in the K -band, I use the results of [Huang et al. \(1997\)](#), who have found that the galaxy number counts satisfy $dN \propto 10^{(0.689 \pm 0.013)m} dm$ for magnitudes brighter than $K \sim 17$; fainter than this the number counts are much less steep with magnitude. This knowledge can be used to estimate the true magnitude of the source, m_{true} , given an observed value for the magnitude, m_{obs} , and given an estimate for the uncertainty in the magnitude. Bayes' theorem gives

$$P(m_{\text{true}}|m_{\text{obs}}) = \frac{P(m_{\text{obs}}|m_{\text{true}})P(m_{\text{true}})}{P(m_{\text{obs}})} \quad (4.19)$$

Assuming the magnitude limit m_{lim} is brighter than 17, the prior is given by

$$P(m_{\text{true}}) = \frac{10^{0.689m_{\text{true}}}}{\int_{-\infty}^{m_{\text{lim}}} 10^{0.689m} dm} \quad (4.20)$$

Assuming Gaussian errors in the magnitude, the likelihood is given by

$$P(m_{\text{obs}}|m_{\text{true}}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(m_{\text{obs}} - m_{\text{true}})^2}{2\sigma^2}\right) \quad (4.21)$$

where σ is the uncertainty in the magnitude. Hence

$$P(m_{\text{true}}|m_{\text{obs}}) \propto P(m_{\text{obs}}|m_{\text{true}})P(m_{\text{true}}) \quad (4.22)$$

$$\propto \exp\left(-\frac{(m_{\text{obs}} - m_{\text{true}})^2}{2\sigma^2}\right) 10^{0.689m_{\text{true}}} \quad (4.23)$$

$$= \exp\left(-\frac{(m_{\text{obs}} - m_{\text{true}})^2 - 2\sigma^2 \cdot 0.689m_{\text{true}} \ln 10}{2\sigma^2}\right) \quad (4.24)$$

$$= \exp\left(\begin{aligned} &-\frac{(m_{\text{obs}} + 0.689\sigma^2 \ln 10 - m_{\text{true}})^2}{2\sigma^2} \\ &+ \frac{(0.689\sigma^2 \ln 10)^2 + 2m_{\text{obs}} \cdot 0.689\sigma^2 \ln 10}{2\sigma^2} \end{aligned}\right) \quad (4.25)$$

$$\propto \exp\left(-\frac{(m_{\text{obs}} + 0.689\sigma^2 \ln 10 - m_{\text{true}})^2}{2\sigma^2}\right) \quad (4.26)$$

which is a Gaussian with mean $\langle m_{\text{true}} \rangle = m_{\text{obs}} + 0.689\sigma^2 \ln 10$, suggesting that the true magnitude will generally be fainter than the observed magnitude. This is a small correction for $\sigma \ll 1$ mag; for example, for $\sigma = 0.1$ mag, $m_{\text{true}} - m_{\text{obs}} = 0.0157$ mag, while for $\sigma = 0.14$ mag, the correction is 0.0311 mag.

This correction is valid only for magnitude-limited samples, so it is not applicable to the luminosity function analysis of this thesis, for which the sample has more complex limits.

4.4 Jackknife error estimation

Statistical errors for all quantities (except where indicated) are estimated through the jackknife method (Lupton, 1993). The sample area is divided into 24 regions, each containing approximately the same number of galaxies; the principal subdivisions are shown in Fig. 4.1. The space density is calculated once for the whole sample and then a further 24 times, each time omitting one of the 24 regions. The variance of each value of ϕ is then calculated using:

$$\text{Var}(\phi) = \frac{n-1}{n} \sum_i (\phi_i^{\text{J}} - \overline{\phi^{\text{J}}})^2 \quad (4.27)$$

where ϕ_i^{J} is the i th jackknife resampling of the data and n is the number of jackknife resamplings (24 in this case). A bias correction (Lupton, 1993) is applied to the space density, giving a new value of

$$\phi' = \phi + (n-1)(\phi - \overline{\phi^{\text{J}}}) \quad (4.28)$$

where ϕ is the original estimate for the whole sample.

Uncertainties and bias corrections for the functional fits and integrated quantities (for example, luminosity density) are estimated using the same method.

It would be preferable, but more difficult, to divide the sample into regions of equal area rather than equal numbers of galaxies. Given the way in which the effective area is calculated, these subdivisions cannot say anything about the overall normalization of the space density; for example, whether the south Galactic pole has a lower density of galaxies than the north Galactic pole.

Uncertainties in the magnitudes and other galaxy properties are not included in the analysis. This is likely to have only a small effect on most of the results, but the magnitude errors will lead to a slight overestimation of the space density of galaxies at the bright end of the luminosity function, where the LF is very steep (Jones et al., 2006).

4.5 Functional forms and fits

Various simple functional forms may be used to describe the galaxy population. The most popular of these are described in this section, along with the method used to find the best-fitting parameters.

4.5.1 Schechter function

It has been found that the luminosity function may be approximated by a simple form, the Schechter function (Schechter, 1976), although statistically significant departures from this form have been detected in recent analyses.

Expressed in terms of luminosities, the Schechter function is

$$\phi(L) dL = \phi^* \left(\frac{L}{L^*} \right)^\alpha \exp \left(-\frac{L}{L^*} \right) \frac{dL}{L^*}. \quad (4.29)$$

This has a faint-end power-law slope, α , and an exponential decline at high luminosities. The ‘knee’ of the curve, at which the shape changes from a power-law to an exponential decline, is at $L = L^*$. ϕ^* gives the normalization of the Schechter function.

In terms of magnitudes, using

$$M - M^* = -2.5 \log \frac{L}{L^*} \quad (4.30)$$

and

$$\frac{L}{L^*} = 10^{0.4(M^* - M)} \quad (4.31)$$

$$= e^{0.4(M^* - M) \ln 10} \quad (4.32)$$

the Schechter function is:

$$\phi(M) dM = -\phi(-M) d(-M) \quad (4.33)$$

$$= \phi^* 10^{0.4\alpha(M^* - M)} \exp \left(-10^{0.4(M^* - M)} \right) (-0.4 \ln 10) e^{0.4(M^* - M) \ln 10} d(-M) \quad (4.34)$$

$$= (0.4 \ln 10) \phi^* 10^{0.4(\alpha+1)(M^* - M)} \exp \left(-10^{0.4(M^* - M)} \right) dM. \quad (4.35)$$

There is a special case where $\alpha = -1$, the **flat slope**. Considering the faint end, where $L \ll L^*$, the number density of objects is given by,

$$\phi(L) \propto L^{-1}, \quad (4.36)$$

so the number of objects with luminosity between L_1 and L_2 is given by,

$$\int_{L_1}^{L_2} \phi(L) dL \propto \int_{L_1}^{L_2} L^{-1} dL = \ln \left(\frac{L_2}{L_1} \right) \propto M_1 - M_2, \quad (4.37)$$

where M_1 and M_2 are the absolute magnitudes corresponding to L_1 and L_2 respectively. So for $\alpha = -1$ there is the same number of objects in each bin of equal width in log-luminosity or equal width in magnitudes.

The luminosity density, $\int L\phi(L) dL$, at the faint end in the case $\alpha = -1$ is given by,

$$\int_{L_1}^{L_2} L\phi(L) dL \propto \int_{L_1}^{L_2} LL^{-1} dL = L_2 - L_1. \quad (4.38)$$

So for $\alpha = -1$ there is the same contribution to the total luminosity from each bin of equal width in luminosity.

These two results combined mean that, for a flat faint-end slope, faint galaxies are numerous, but they make a small contribution to the total luminosity density.

Luminosity density

The total number density of galaxies may be calculated (for $\alpha > -1$) by integrating the Schechter function:

$$N = \int_0^{\infty} \phi(L) dL = \phi^* \Gamma(\alpha + 1), \quad (4.39)$$

where Γ is the gamma function, defined for $z > 0$ by

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt. \quad (4.40)$$

For $\alpha > -2$ the total luminosity density may be calculated (Blanton et al., 2001):

$$j = \int_0^{\infty} L \phi(L) dL = L^* \phi^* \Gamma(\alpha + 2). \quad (4.41)$$

This is conventionally given in solar luminosities and may be expressed in terms of the absolute magnitude of the sun, M_{\odot} , as

$$j = \phi^* 10^{0.4(M_{\odot} - M^*)} \Gamma(\alpha + 2) \quad (4.42)$$

Given estimates for the three Schechter function parameters, along with a covariance matrix describing the uncertainties in the parameters, it is possible to estimate the uncertainty in the luminosity density, j .

Given some quantity $g = g(x_i, \dots)$ and given the covariance matrix of the quantities $\{x_i\}$, the variance of g is given by

$$\sigma_g^2 = \sum_{i,j} \frac{\partial g}{\partial x_i} \frac{\partial g}{\partial x_j} \text{Cov}(x_i, x_j) \quad (4.43)$$

From Equation (4.42), the partial derivatives are

$$\frac{\partial j}{\partial M^*} = -(0.4 \ln 10) j \quad (4.44)$$

$$\frac{\partial j}{\partial \alpha} = \psi(\alpha + 2) j \quad (4.45)$$

$$\frac{\partial j}{\partial \phi^*} = \frac{j}{\phi^*} \quad (4.46)$$

where ψ is the digamma function, defined as

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)} \quad (4.47)$$

The uncertainty in j may then be found from Equation (4.43).

4.5.2 Double Schechter function

Some authors have used a Schechter function with two faint-end slopes to describe the luminosity function. Blanton et al. (2005b) and Baldry et al. (2008) express the double Schechter function in

the form

$$\phi(L) dL = \exp\left(-\frac{L}{L^*}\right) \left(\phi_1^* \left(\frac{L}{L^*}\right)^{\alpha_1} + \phi_2^* \left(\frac{L}{L^*}\right)^{\alpha_2} \right) \frac{dL}{L^*}, \quad (4.48)$$

(this is expressed in magnitudes by [Blanton et al., 2005b](#)) while [Loveday \(1997\)](#) uses

$$\phi(L) dL = \phi^* \left(\frac{L}{L^*}\right)^\alpha \exp\left(-\frac{L}{L^*}\right) \left(1 + \left(\frac{L}{L_t}\right)^\beta\right) \frac{dL}{L^*}. \quad (4.49)$$

These are equivalent, as can be seen from the second form,

$$\phi(L) dL = \exp\left(-\frac{L}{L^*}\right) \left(\phi^* \left(\frac{L}{L^*}\right)^\alpha + \phi^* \left(\frac{L}{L^*}\right)^\alpha \left(\frac{L}{L_t}\right)^\beta \right) \frac{dL}{L^*} \quad (4.50)$$

$$= \exp\left(-\frac{L}{L^*}\right) \left(\phi^* \left(\frac{L}{L^*}\right)^\alpha + \phi^* \left(\frac{L^*}{L_t}\right)^\beta \left(\frac{L}{L^*}\right)^{\alpha+\beta} \right) \frac{dL}{L^*} \quad (4.51)$$

which is the same as the first form with $\phi_1^* = \phi^*$, $\alpha_1 = \alpha$, $\phi_2^* = \phi^*(L^*/L_t)^\beta$ and $\alpha_2 = \alpha + \beta$.

A steeper very faint end slope has been noticed by others but not necessarily expressed in a functional form (e.g., [Driver et al., 1994](#)).

Luminosity density

The luminosity density of the double Schechter function is given by

$$j = 10^{0.4(M_\odot - M^*)} (\phi_1^* \Gamma(\alpha_1 + 2) + \phi_2^* \Gamma(\alpha_2 + 2)) \quad (4.52)$$

With the uncertainty in j found as for the Schechter function, with

$$\frac{\partial j}{\partial \alpha_i} = \psi(\alpha_i + 2) j \quad (4.53)$$

$$\frac{\partial j}{\partial \phi_i^*} = \frac{j}{\phi_i^*} \quad (4.54)$$

for $i = 1$ and $i = 2$.

4.5.3 Chołowiecki function

For the bivariate brightness distribution (BBD), the [Chołowiecki \(1985\)](#) function is used, which is a Schechter function in luminosity combined with a Gaussian distribution in surface brightness ([Driver et al., 2005](#)):

$$\begin{aligned} \phi(M, \mu_e) &= \frac{0.4 \ln 10}{\sqrt{2\pi} \sigma_{\mu_e}} \phi^* 10^{0.4(M^* - M)(\alpha + 1)} e^{-10^{0.4(M^* - M)}} \\ &\times \exp\left\{-\frac{1}{2} \left[\frac{\mu_e - \mu_e^* - \beta(M - M^*)}{\sigma_{\mu_e}} \right]^2\right\} \end{aligned} \quad (4.55)$$

where M^* , α and ϕ^* are the usual Schechter function parameters, μ_e^* is the mean surface brightness at M^* , σ_{μ_e} is the standard deviation in surface brightness and β is the slope of the relationship between absolute magnitude and mean surface brightness. Note that integrating the Chołowiecki function over surface brightness gives a Schechter function.

4.5.4 Functional fits

Functional fits for the BBD, LFs and SMF are found using IDL routines written by myself, which make use of the MPFIT routines written by Craig Markwardt.²

The range of points over which the functions are fit is restricted in order to avoid known regions of incompleteness: only data brighter than -20 in $M_K - 5 \log h$, -18 in $M_r - 5 \log h$ and 19 mag arcsec $^{-2}$ in surface brightness, or more massive than 9.5 in $\log(h^{-2}M_\odot)$ for the stellar mass, are used to generate the functional fits.

Empty bins at the bright end are ignored for the Schechter function fits. However, for the Chołoniewski function, it is necessary to include the empty bins, which are assigned a large uncertainty proportional to $1/V_{\max}$ for a hypothetical galaxy at the centre of the bin, as shown in Sections 4.7 and 5.1.

4.6 Estimating the space density of galaxies

In this section I review and evaluate various methods for estimating the space density of galaxies, whether for a univariate or a multivariate distribution.

There are three widely-used methods for estimating the luminosity function (for a comparison of the various methods, see Willmer, 1997; Takeuchi et al., 2000). If the functional form of the luminosity function is known, the best method to use is the STY (Sandage et al., 1979) maximum likelihood method. The other two methods have no prior assumptions about the shape of the luminosity function. These are the SWML method (Efstathiou et al., 1988), a version of the STY method where the functional form is described by a value in each bin, and the $1/V_{\max}$ method (Schmidt, 1968), which is a simple estimator of the value of the space density in each bin. These are described below in more detail.

4.6.1 Univariate or multivariate?

In a survey limited by more than one measured quantity, e.g., two magnitudes, or magnitude and surface brightness, there are two ways to proceed. One is explicitly to take into account all of the known selection effects in a multivariate analysis. The other is to set strong limits in one or more quantities so that the selection effects in the remaining quantities are negligible. This can lead to a univariate analysis, if sufficiently strong limits are set on one quantity (usually the magnitude) such that all other selection effects can be ignored. Either of these approaches ensures that completeness is taken into account and that the contribution of each observed galaxy to the

²<http://cow.physics.wisc.edu/~craigm/idl/fitting.html>

total space density is estimated accurately.

The sample used here is limited in both r - and K -band apparent magnitude, with $r < 17.6$ and $K < 16$. For a univariate analysis limited in K -band magnitude, a stronger K -band flux limit must be set in order for the r -band limit to be negligible. From Fig. 4.3 it can be seen that a large number of galaxies in the sample have $r - K > 3$, so a univariate analysis would require a limit of approximately $K < 14$ in order to avoid introducing a bias against red galaxies. If such a strong limit were introduced, the sample would be much smaller, with all galaxies fainter than $K = 14$ excluded, providing only a small improvement in depth over 2MASS, which has a completeness limit of $K_S = 13.5$ (Jarrett et al., 2000), while covering a much smaller area. This makes it strongly desirable to perform a multivariate analysis, in order to include a larger number of galaxies in the sample.

Moreover, the limit in large K -band Petrosian radius would be negligible only if a sufficiently stringent bright K -band flux limit is introduced, such that none of the galaxies included have a Petrosian radius larger than 6 arcsec. However, Fig. 3.14 makes it clear that large galaxies are found even at very faint K -band apparent magnitude, so a cut of this nature is not possible.

4.6.2 STY method

The STY method (Sandage et al., 1979) takes a functional form for the luminosity function and provides the most direct way to constrain the parameters of that functional form using the data.

Given a parameterized luminosity function, $\psi(M, z)$, that has a constant shape with redshift, i.e., $\psi(M, z) = \rho(z)\phi(M)$, the probability density that a galaxy selected from a magnitude-limited sample at redshift z_i will have absolute magnitude M_i is given by (Willmer, 1997)

$$p_i = p(M_i, z_i) = \frac{\psi(M_i, z_i)}{\int_{-\infty}^{M_{\text{faint}}(z_i)} \psi(M, z_i) dM} = \frac{\phi(M_i)}{\int_{-\infty}^{M_{\text{faint}}(z_i)} \phi(M) dM} \quad (4.56)$$

where $M_{\text{faint}}(z_i)$ is the faintest absolute magnitude at redshift z_i that would be visible in the survey and where $\rho(z_i)$ cancels out of the equation. The probability that the observed sample of N galaxies is drawn from this particular form of ϕ (the likelihood) is defined as

$$\mathcal{L} = \prod_{i=1}^N p_i \quad (4.57)$$

The parameters of the luminosity function are then adjusted to maximize the likelihood.

Note that this assumes the shape of the luminosity function does not vary with redshift, although the normalization is allowed to vary, making the method robust to changes in the density of the sample as a function of redshift (clustering).

Note also that this does not incorporate the uncertainties in the magnitude. These could be taken into account by using

$$p_i = \frac{\int \phi(M) h_i(M) dM}{\int_{-\infty}^{M_{\text{faint}}(z_i)} \phi(M) dM} \quad (4.58)$$

where $h_i(M)$ is the probability that galaxy i has absolute magnitude M ; this would generally take a Gaussian form. Alternatively, if galaxy i has a Gaussian magnitude uncertainty of width Δm_i , then the luminosity function could be convolved with with a Gaussian of width Δm before calculating p_i (cf. [Blanton et al., 2003b](#))

$$\phi'(M) = \phi(M) \otimes G(\Delta m) \quad (4.59)$$

This new form for the luminosity function, $\phi'(M)$, is then used in Equation (4.56).

If $\phi(M)$ is chosen to be a Schechter function, then the values of M^* and α are explored to maximize the likelihood; the value of ϕ^* has to be determined by other means, since the STY method does not give the normalization.

The method can be extended in various ways. More complex survey limits may be incorporated into the upper- and lower-limits of the integral in Equation (4.56). Moreover, the method can be changed from a univariate to a multivariate method by replacing $\phi(M)$ with $\phi(M, \dots)$ and by using a multivariate integral in Equation (4.56). However, a multivariate-STY method would require a multivariate parameterization of the space density, for example, a Chołoniewski function in magnitude and surface brightness.

4.6.3 SWML method

The stepwise maximum likelihood (SWML) method was developed by [Efstathiou et al. \(1988\)](#) (EEP) in order to see deviations of the luminosity function from the Schechter function form used in the STY method. The SWML method is actually a special case of the STY method, in which the chosen functional form is not a Schechter function but a binned luminosity function with a free parameter for the value in each bin.

By separating the luminosity function into N_p bins, equation (4.56) can be written as

$$p_i = \frac{W(M_i - M_k) \phi_k}{\sum_{j=1}^{N_p} \phi_j \Delta M H(M_j - M_{\text{faint}}(z_i))} \quad (4.60)$$

where W and H are two window functions, W assigning the galaxy to the correct bin and H producing a cut-off for faint absolute magnitudes.

The luminosity function is found by maximizing the likelihood with respect to the parameters $\{\phi_k\}$, i.e.,

$$\frac{\partial \ln \mathcal{L}}{\partial \ln \phi_k} = 0 \quad (4.61)$$

which gives

$$\phi_k \Delta M = \frac{\sum_i^N W_{ik}}{\sum_i^N \left(H_{ik} / \sum_j^{N_p} \phi_j \Delta M H_{ij} \right)}. \quad (4.62)$$

This is used to give the values of $\{\phi_k\}$ by iteration. As with the STY method, the normalization must be found independently.

The weighting function, W , is set to be 1 or 0 by [Efstathiou et al. \(1988\)](#), but may be used to correct for incompleteness in a redshift survey ([Driver et al., 2005](#)).

Like with the STY method, the SWML method may be extended to include more complex survey limits (through the window function H) or to a full multivariate form. Uncertainties in the galaxy magnitudes could also be incorporated, as with the STY method.

In principle it would also be possible to implement the SWML method using a hybrid binned and functional form for a bivariate distribution, for example, Gaussian in surface brightness and binned in luminosity.

The SWML method generally gives good results. However, the bins are highly correlated, so any problem that occurs is likely to affect the whole luminosity function. For example, the method is built on the assumption that the luminosity is uncorrelated with position ([Efstathiou et al., 1988](#)), so that the shape of the luminosity function does not vary with redshift (although the normalization is allowed to vary). However, if the survey includes large voids or superclusters, with the most massive galaxies found generally in high-density environments, then the shape, as well as the normalization, of the luminosity function will vary. This will bias the SWML luminosity function in a highly nontrivial manner.

These effects may be exacerbated when the method is extended to several dimensions, since the number of bins is greatly increased.

Uncertainties in the SWML space density could be calculated from the covariance matrix. However, this matrix has many terms, so conventionally the errors are estimated through the second derivatives of the likelihood function at the point of maximum likelihood ([Efstathiou et al., 1988](#); [Blanton et al., 2001](#)).

4.6.4 Normalization for the STY and SWML methods

Various methods have been used to find the normalization of the STY and SWML luminosity functions.

The most widely-used method (e.g., [Blanton et al., 2001](#); [Kochanek et al., 2001](#); [Blanton et al., 2003b](#); [Ball et al., 2006](#)) makes use of the minimum variance estimator of [Davis & Huchra \(1982\)](#). This method assigns to each galaxy a weight according to volume, i.e., a higher weight for galaxies at higher redshift. In contrast, normalizing to the number counts gives each galaxy an equal weight,

so the normalization will be biased towards the normalization at the median redshift of the sample. It is preferable to use this volume-weighted method, although it relies strongly on the model for luminosity evolution; if this model is inadequate then the normalization will be affected (Blanton et al., 2003b).

For their bivariate SWML method, Driver et al. (2005) find the normalization by measuring the number of galaxies in a volume-limited sub-sample.

Jones et al. (2006) normalize the SWML and STY methods ‘by a χ^2 minimization with respect to the equivalent $1/V_{\max}$ distribution’.

4.6.5 $1/V_{\max}$ method

So far the idea has been to find a model for the luminosity function, generate predictions from that model, and use the observations to constrain the parameters of the model. However, it is possible to work directly from the observations by means of an **estimator** for the luminosity function.

The $1/V_{\max}$ estimator (Schmidt, 1968; Rowan-Robinson, 1969) is a maximum likelihood estimator for the binned luminosity function (Wall & Jenkins, 2003, p. 154). The luminosity function is divided into bins in absolute magnitude. The value in each bin, ϕ_i , is given by

$$\phi_i = \sum_j \frac{1}{V_{\max,j}} \quad (4.63)$$

for all galaxies j with absolute magnitude falling within that bin. $V_{\max,j}$ is the maximum volume within which galaxy j would be visible. This is estimated by supposing that a particular galaxy were shifted to a higher or lower redshift. The maximum redshift for that galaxy is the redshift at which the galaxy would be seen at the faint magnitude limit. Similarly, the minimum redshift corresponds to the galaxy being seen at the bright magnitude limit (if such a limit exists). V_{\max} is then the survey volume within these redshift limits.

This description assumes that the survey is limited only by apparent magnitude. However, the method can easily be extended to incorporate any number of limits on the survey, each one of which will provide a constraint on the minimum or maximum redshift at which a particular galaxy would be included in the sample.

In this work, it is necessary to include various limits on the sample. This is because the volume within which a galaxy has $r < r_{\text{lim}}$ may be smaller than the volume within which the galaxy has $K < K_{\text{lim}}$.

The method may also be extended to give a multivariate binned space density, or the space density binned as a function of any galaxy property (e.g., stellar mass). The value in bin i , ϕ_i , is given by Equation (4.63) for all galaxies j lying in that bin. However, it must be safe to assume

that each bin samples the full range of galaxy types that may be found in that bin. For example, in a sample limited in more than one waveband, the lowest-luminosity bins may sample only red galaxies or only blue galaxies, leading to incompleteness in the estimated luminosity function. This will be discussed at length later in this thesis.

The $1/V_{\max}$ method, unlike the SWML method, is liable to give incorrect results in the presence of clustering, where the number density of galaxies varies with redshift. One means of compensating for this is to use a volume-limited sample of bright galaxies to estimate how the number density varies with redshift. A correction is then applied so that each bin of equal volume has the same number density of galaxies (Cross et al., 2001; Baldry et al., 2008). However, as with the SWML method, this assumes that clustering is independent of luminosity.

Occasionally a galaxy will be detected extremely close to the survey limits. In certain cases this means that a galaxy would not have been seen if it had been very slightly further away *or* if it had been very slightly closer to us. This galaxy would have a very small value for V_{\max} , and hence a very large value for $1/V_{\max}$. This could lead to a large overestimate of the value for ϕ in the corresponding bin.

V/V_{\max} test

The V/V_{\max} test (Schmidt, 1968) can be used to identify incompleteness or evolution. If z is the redshift at which a galaxy is observed, z_{\min} and z_{\max} are respectively the minimum and maximum redshifts at which the galaxy could in principle be observed, and $V(z)$ is the survey volume between redshift 0 and z , then

$$V_{\max} = V(z_{\max}) - V(z_{\min}) \quad (4.64)$$

and

$$V = V(z) - V(z_{\min}). \quad (4.65)$$

In the absence of evolution and with no incompleteness, V/V_{\max} should be drawn from a uniform distribution between 0 and 1, while the mean value for the whole sample, $\langle V/V_{\max} \rangle$, is expected to be $0.5 \pm (12n)^{-1/2}$ where n is the number of galaxies in the sample. If the mean differs significantly from 0.5, this could be due to (1) incompleteness in the sample, (2) evolution of the number density of galaxies, or (3) evolution of the luminosity of galaxies, or some combination of these factors.

As an example, if the sample is assumed to be complete to a certain faint magnitude limit, but in fact is complete only to a brighter limit, then all of the values of V_{\max} will be too large. Hence $\langle V/V_{\max} \rangle$ will be less than 0.5.

4.6.6 Choice of method and implementation

Of the three methods, I have chosen to investigate the $1/V_{\max}$ method, since this is the easiest to implement, and since it is the most robust against catastrophic failure. The $1/V_{\max}$ estimator is more robust than SWML against systemic failure, but requires a large survey volume in order to overcome the effects of large-scale structure. This is unlikely to be a problem in this thesis, except for low-luminosity galaxies, since a large volume has been probed.

I investigated the SWML method using the MULTILUM code, written by Jon Loveday and used by Loveday (2000) and Ball et al. (2006). However, owing to the complexity of the four-dimensional parameter space investigated (K - and r -band magnitude, K -band effective surface brightness and K -band Petrosian radius), the SWML method was found to give implausible results. I chose instead to focus on the $1/V_{\max}$ method, using my own implementation in IDL.

The volume within which each galaxy would be visible, V_{\max} , is estimated for each galaxy by considering the tightest constraints on maximum and minimum redshift provided by the limits in redshift, K -, r -, g - and i -band magnitudes, K - and r -band surface brightness and K -band Petrosian radius (see Table 3.3). Each galaxy is then given a weight $w_i = 1/V_{\max,i}$. The space density ϕ of galaxies of a certain type (e.g., binned in absolute magnitude and/or surface brightness) is given by

$$\phi = \sum_i w_i \quad (4.66)$$

for galaxies i of that type.

The constraints on z_{\min} and z_{\max} are found independently for each parameter using the following equations. For absolute magnitude, M :

$$m_{\text{bright}} = M + DM(z_{\min,M}) + K(z_{\min,M}) - E(z_{\min,M}) \quad (4.67)$$

$$m_{\text{faint}} = M + DM(z_{\max,M}) + K(z_{\max,M}) - E(z_{\max,M}) \quad (4.68)$$

where m_{bright} and m_{faint} are the magnitude limits. For absolute surface brightness, μ_{abs} :

$$\mu_{\text{bright}} = \mu_{\text{abs}} + 10 \log(1 + z_{\min,\mu}) + K(z_{\min,\mu}) - E(z_{\min,\mu}) \quad (4.69)$$

$$\mu_{\text{faint}} = \mu_{\text{abs}} + 10 \log(1 + z_{\max,\mu}) + K(z_{\max,\mu}) - E(z_{\max,\mu}) \quad (4.70)$$

where μ_{bright} and μ_{faint} are the limits. And for radius, r (in arcseconds) is related to R (in h^{-1} kpc) by

$$\left(\frac{r_{\max}}{3600}\right) \left(\frac{\pi}{180}\right) 1000 D_A(z_{\min,R}) = R \quad (4.71)$$

$$\left(\frac{r_{\min}}{3600}\right) \left(\frac{\pi}{180}\right) 1000 D_A(z_{\max,R}) = R \quad (4.72)$$

where D_A is the angular diameter distance in h^{-1} Mpc and r_{\max} and r_{\min} are the limits.

Table 4.1: Limitations on redshift for 36 659 galaxies in the $1/V_{\max}$ method. Shown are the number of galaxies limited in minimum or maximum redshift by the corresponding observable parameter. This corresponds to the tightest constraint on z_{\min} and z_{\max} for each galaxy. See Table 3.3 for the limits set on each parameter.

Observable	Number limited in z_{\min}	Number limited in z_{\max}
K Petrosian magnitude	0	193
r Petrosian magnitude	0	36 340
g fiber magnitude	0	0
i fiber magnitude	0	0
K Petrosian radius	36 579	0
$\mu_{e,K}$	0	1
$\mu_{e,r}$	0	0
z	80	125

Note that the effects of observing a different rest-frame band than the K -band are not included in the radius calculation. The Petrosian radius of a galaxy does vary as a function of wavelength, with galaxies being more concentrated in red bands than in blue, but this is a tiny effect considering the size of the variations in the wavelength: the observed K -band, with effective wavelength $2.20 \mu\text{m}$ (Hewett et al., 2006), corresponds to a rest-frame effective wavelength of $1.69 \mu\text{m}$ at redshift 0.3, which is slightly redder than the H -band ($1.63 \mu\text{m}$).

The overall redshift limits for each galaxy are then found from the above redshift limits by

$$z_{\min} = \max(z_{\min,M}, \dots, 0.01) \quad (4.73)$$

$$z_{\max} = \min(z_{\max,M}, \dots, 0.3) \quad (4.74)$$

where 0.01 and 0.3 are the redshift limits on the sample (Table 3.3).

Given these minimum and maximum redshifts, the volume probed for each galaxy is given by

$$V_{\max} = V(z_{\max}) - V(z_{\min}) \quad \text{if } z_{\max} > z_{\min} \quad (4.75)$$

$$= 0 \quad \text{otherwise} \quad (4.76)$$

Galaxies with $V_{\max} = 0$ are removed from the sample; this affected 4 galaxies, reducing the sample size to 36 659.

Table 4.1 shows the relative importance of each of the sample limits in constraining V_{\max} . By far the strongest constraint on the maximum redshift at which a galaxy would be visible is the faint

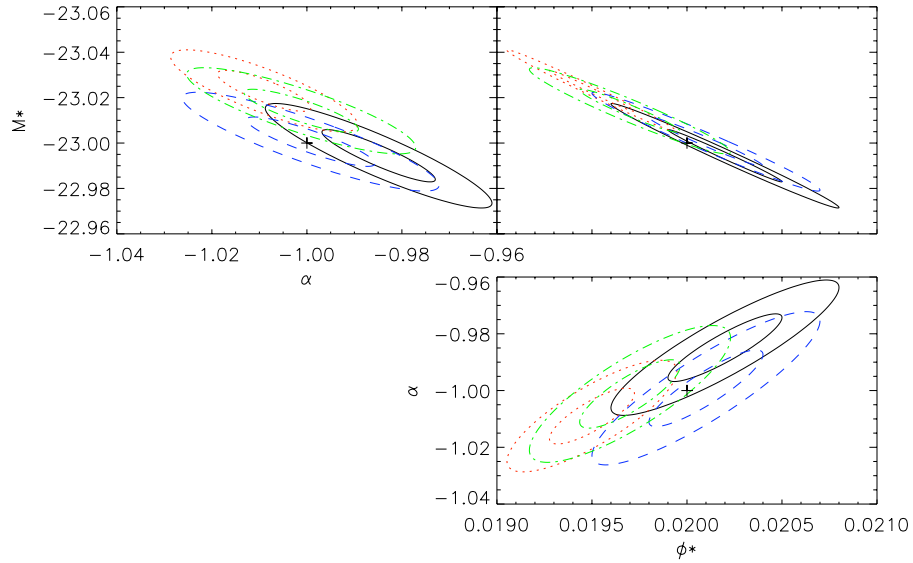


Figure 4.5: Schechter function parameters for the simulated samples. Plus symbols show the input parameters, $M^* - 5 \log h = -23$, $\alpha = -1$ and $\phi^* = 0.02h^3 \text{ Mpc}^{-3}$, while contours show the one- and two-sigma error contours on the recovered parameters for the four different simulated samples. These uncertainties are estimated from 24 jackknife estimations of the Schechter function.

r -band magnitude limit. For the minimum redshift, by far the strongest constraint comes from the limit on the large K -band Petrosian radius.

4.7 Testing the method with simulated data

I have tested the V_{\max} estimator and fitting routines using four simulated samples drawn from a known Chołoniewski function, with Gaussian $r - K$ colours (mean and standard deviation derived from the observed sample, but with no dependence on luminosity), and subject to the same observational limits of the data sample, except with $r < 17.77$ rather than $r < 17.6$. Each sample contains between 42 632 and 42 885 galaxies. The simulated samples were generated using code I have written in Python and IDL.

Fig. 4.5 shows the recovered Schechter function parameters from the simulated samples. There is no obvious systematic bias and the recovered uncertainties, shown by the contours, give a reasonable measure of the closeness of the recovered parameters to the input Schechter function.

The luminosity density, in solar units, may be calculated from the Schechter function as

$$j = \int_0^\infty L \phi(L) dL = \phi^* 10^{0.4(M_\odot - M^*)} \Gamma(\alpha + 2) \quad (4.77)$$

where M_\odot is the absolute magnitude of the sun (taken to be 3.32 in the K -band, [Bell et al., 2003b](#)),

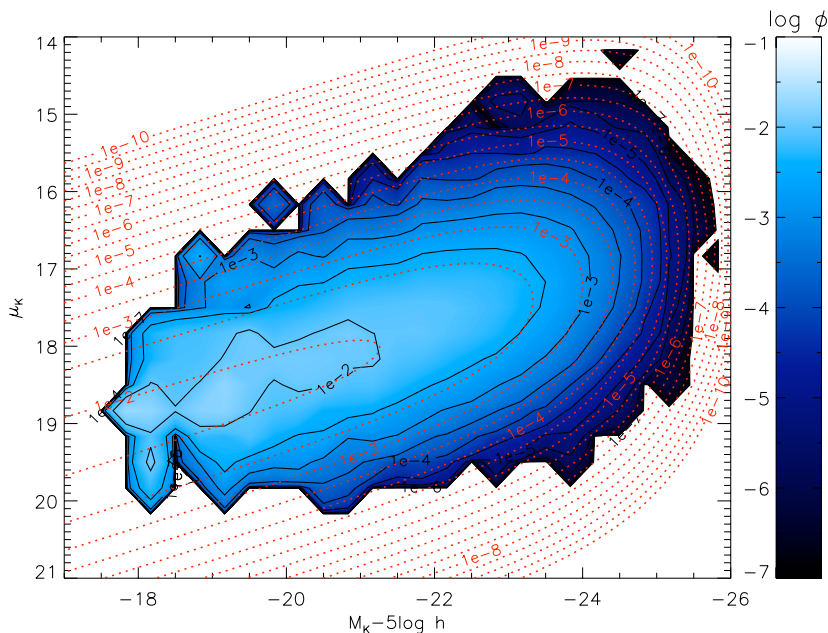


Figure 4.6: Input Chóloniewski function for the simulations (red dotted contours) and the recovered BBD from one of the simulated samples (shaded regions and black solid contours). The space density, ϕ , is in units of $h^3 \text{Mpc}^{-3} \text{mag}^{-1} (\text{mag arcsec}^{-2})^{-1}$. A Chóloniewski function was fit to the recovered BBD (not shown). The input (recovered) Chóloniewski function parameters are $M^* - 5 \log h = -23(-23.03 \pm 0.01) \text{mag}$, $\alpha = -1(-0.96 \pm 0.01)$, $\phi^* = 0.02(0.0224 \pm 0.0003) h^3 \text{Mpc}^{-3}$, $\mu_e^* = 17.5(17.363 \pm 0.005) \text{mag arcsec}^{-2}$, $\sigma_{\mu_e} = 0.6(0.580 \pm 0.003) \text{mag arcsec}^{-2}$ and $\beta = 0.3(0.287 \pm 0.003)$. Very similar results are obtained for the other three simulated samples.

or from the weights of each galaxy by

$$j = \sum_i 10^{0.4(M_\odot - M_i)} w_i. \quad (4.78)$$

For the simulated samples the luminosity density was estimated in these two ways and compared to the luminosity density from the input Schechter function. By summing the galaxy weights, the recovered luminosity density typically underestimated the input luminosity density by around 1 per cent, whereas there was no obvious bias from integration of the recovered Schechter function.

Fig. 4.6 shows the input Chóloniewski function and the recovered BBD for one of the simulated samples. The Chóloniewski function gives a good fit to the recovered BBD, but the recovered BBD is itself biased with respect to the input Chóloniewski function, for all four simulated samples, most noticeably towards higher surface brightness. The BBDs presented in Chapter 5 should therefore be considered only approximately correct.

The Chóloniewski fit requires every bin in the BBD, including the empty bins, to have an

uncertainty. For the bins that have galaxies, the uncertainty is estimated through jackknife resampling, as described in Section 4.4. For the empty bins, as described in Section 4.5.4, the uncertainty is estimated by first finding V_{\max} for a hypothetical galaxy at the centre of the bin, and then multiplying $1/V_{\max}$ by an arbitrary (large) number, chosen to give satisfactory results for the fits. This method assigns a large uncertainty to the empty bins at low luminosity, where a small volume has been sampled, and a small uncertainty to the high-luminosity empty bins, where a large volume has been sampled.

To illustrate the uncertainties in the BBD, Figs. 4.7 and 4.8 respectively show the BBD with errors subtracted and added. However, simply adding or subtracting the one-sigma errors from the value at each bin, then plotting on a contour plot, would not give a meaningful estimate of the uncertainty, since this is strongly dependent on the number of bins. Instead, one-sigma errors for the whole BBD are estimated by scaling all the bins together, in proportion to the size of the uncertainty in each bin, so that the reduced χ^2 gives a one-sigma error. This corresponds to

$$\chi^2 \equiv \sum_i \frac{(X_i - \mu_i)^2}{\sigma_i^2} = \sqrt{2k} \quad (4.79)$$

where k is the number of degrees of freedom (number of bins), and since the χ^2 distribution has variance $2k$. So each bin will be varied so that $(X_i - \mu_i)^2/\sigma_i^2 = \sqrt{2k}/k = \sqrt{2/k}$, where X_i is the upper or lower one-sigma value for bin i , μ_i is the "true" value, and σ_i is the uncertainty in bin i . I.e., $X_i = \mu_i \pm \sigma_i \sqrt{2/k}$; these are the values shown in Figs. 4.7 and 4.8.

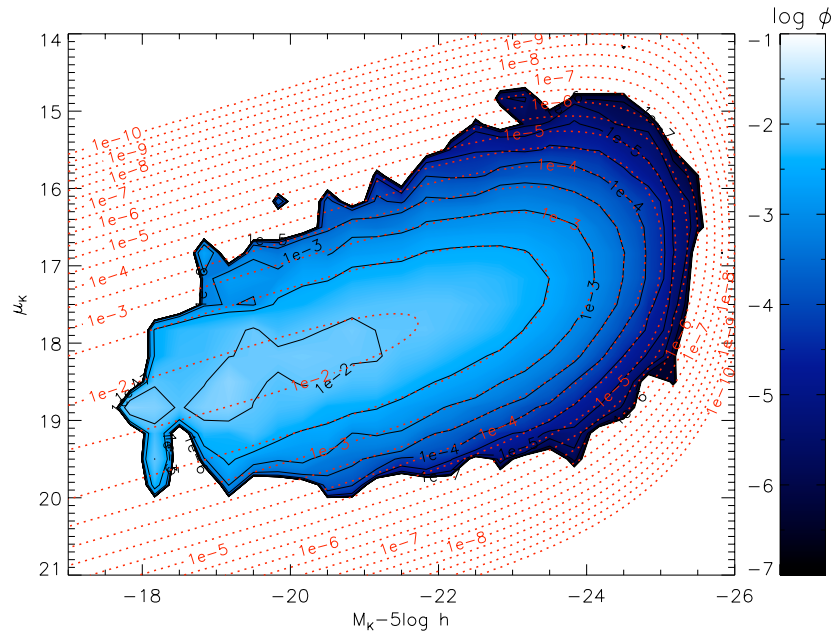


Figure 4.7: Recovered Chóloniewski function for the simulations (red dotted contours) and the recovered BBD from one of the simulated samples, with one-sigma (χ^2) errors subtracted (shaded regions and black solid contours). Recovered Chóloniewski parameters are given in Fig. 4.6.

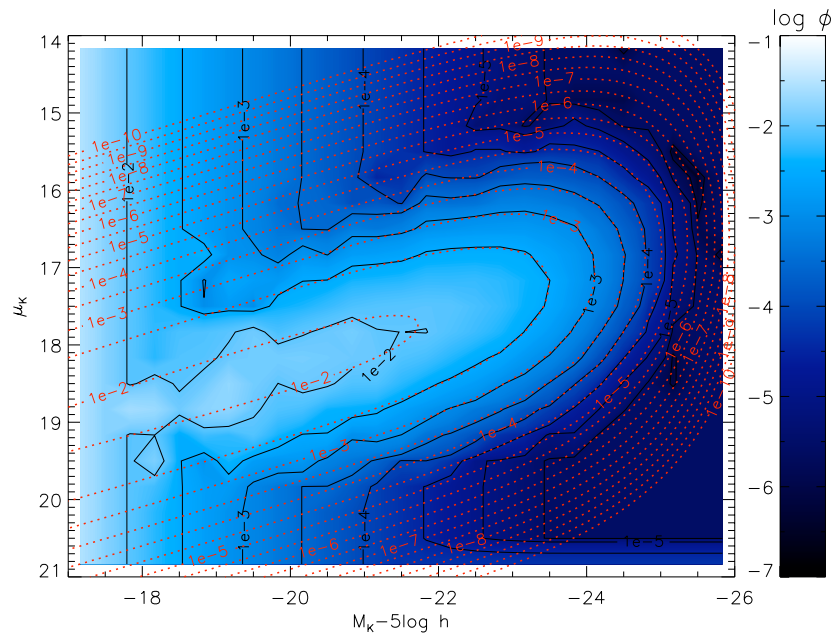


Figure 4.8: As Fig. 4.7 but with one-sigma (χ^2) errors added.

Chapter 5

Space density in luminosity, surface brightness and stellar mass

*Voilà ce que je vois et ce qui me trouble.
Je regarde de toutes parts, et je ne vois partout qu'obscurité.
La nature ne m'offre rien qui ne soit matière de doute et d'inquiétude.*

Given the completeness limits in magnitude, radius, surface brightness and redshift from Chapter 3, the sample of 36 659 galaxies has been defined and will be analysed in this chapter. Fig. 5.1 shows the redshift distribution of this sample, along with the distribution in absolute magnitude, which is calculated according to the method in Chapter 4. The presence of large-scale structure can be seen.

In this chapter, the principal results will be presented and compared with those of the literature, such as the bivariate brightness distribution and luminosity functions for the whole sample and for subdivisions of the data. The reliability of these results will be assessed by investigating the effect on the results of choosing different limits to the sample: in the observable parameters, in sky position and in redshift.

5.1 *K*-band bivariate brightness distribution

Fig. 5.2 shows the bivariate brightness distribution (BBD) in *K*-band absolute magnitude and absolute effective surface brightness, estimated using the $1/V_{\max}$ as described in Equation (4.66). A Chołoniewski function is fit to the BBD: see Section 5.1.3 below.

The value of the space density at any point on the BBD is estimated assuming the visibility of the full range of galaxy types that exist with that absolute magnitude and surface brightness. This may in fact not be the case, given the additional limits in faint *r*-band apparent magnitude and

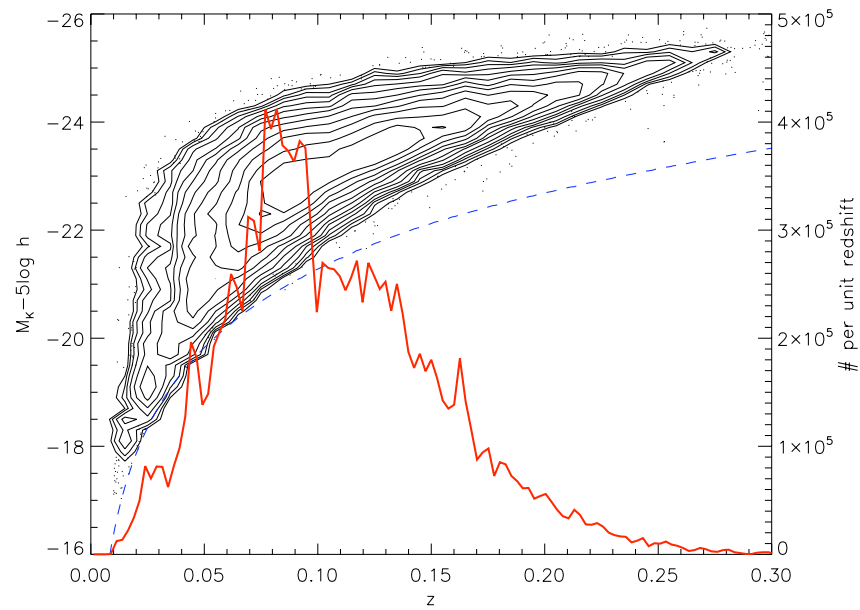


Figure 5.1: Redshift and K -band absolute magnitude distribution of the sample (contours, points and left-hand y -axis) and histogram of redshift distribution (thick red curve, right-hand y -axis). For reference, the absolute magnitude as a function of redshift corresponding to a source at the K -band faint magnitude limit, with typical K - and evolution-corrections and neglecting the r -band limit, is shown by the blue dashed curve. It can be seen that relatively few galaxies are observed near the K -band magnitude limit; this is because of the r -band magnitude limit.

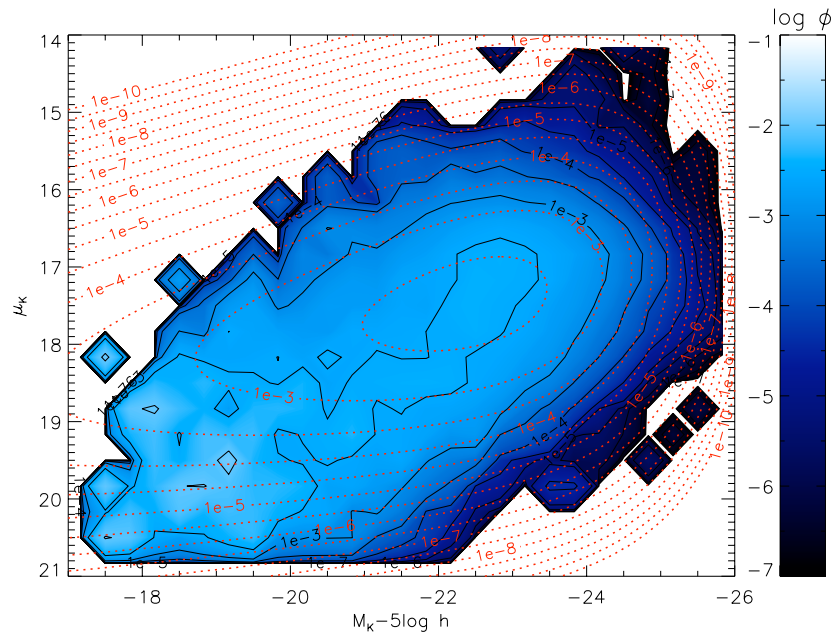


Figure 5.2: BBD for the full sample in K -band absolute magnitude and absolute effective surface brightness. Shaded regions and solid black contours show the space density, ϕ , as in Fig. 4.6. The best-fitting Choloniewski function, estimated using $M_K - 5 \log h < -20$ and $\mu_{e,abs} < 19$, is shown by the red dotted contours. Parameters of the fit are $M^* - 5 \log h = -22.96$ mag, $\alpha = -0.38$, $\phi^* = 0.0203 h^3 \text{ Mpc}^{-3}$, $\mu_{e,abs}^* = 17.37 \text{ mag arcsec}^{-2}$, $\sigma_{\mu_{e,abs}} = 0.643 \text{ mag arcsec}^{-2}$ and $\beta = 0.191$.

large K -band Petrosian radius. These specific issues will be discussed in the following sections. They illustrate a general limitation of having a multiply-limited sample: if a sample is limited in more than two observable parameters, then it is not safe to assume completeness when the sample is binned in only one or two of those parameters.

One way round this problem is to assume a functional form for the parameters to be integrated, and then integrate to infinity. For example, when integrating the BBD to obtain the luminosity function (LF), it could be assumed that in each luminosity bin the BBD is Gaussian in surface brightness. This will compensate for the limited range in surface brightness, but the reliability of the resulting luminosity function will be limited by (1) the assumption that surface brightness is Gaussian at fixed luminosity and (2) the quality of the Gaussian fit in each luminosity bin. A similar method could be used to integrate over size and colour to obtain the BBD, correcting for the limited range in r -band magnitude and K -band Petrosian radius.

Another way to deal with this problem is to check that the space density becomes negligible before the volume becomes too small. For each point in the multi-dimensional parameter space (intrinsic galaxy properties), there will be a certain volume within which galaxies of these properties could be detected in the survey. If this volume is too small, then one cannot give a measurement of ϕ at that point. However, if ϕ has already tailed off to a negligible level before the volume has become too small, then it seems reasonable to ignore contributions from this region of parameter space to the integrated value of ϕ .

5.1.1 Completeness in colour

Figs. 5.3 and 5.4 show the effect of the combined r - and K -band flux limits on the completeness, as a function of K - and r -band absolute magnitude respectively.

From Fig. 4.3 it can be seen that near the r -band flux limit blue galaxies cannot be seen, and that near the K -band flux limit red galaxies cannot be seen. For the red galaxies that would not be seen at near the K -band flux limit, consider galaxies at redshift z with $r = 17.6$ and various $r - K$ colours, neglecting K - and evolution-corrections. These sources will have K -band absolute magnitude

$$M_K = K - DM(z) = r - (r - K) - DM(z). \quad (5.1)$$

Fig. 5.3 shows this relation between M_K and $r - K$ (with $r = 17.6$) for $z = 0.01$ and 0.02 , corresponding respectively to the minimum redshift considered here and the redshift at which the survey volume is just over $10^4 h^{-3} \text{ Mpc}^3$. It can be seen that red galaxies will not be seen at faint K -band luminosity, and some incompleteness is therefore to be anticipated at $M_K - 5 \log h > -19$.

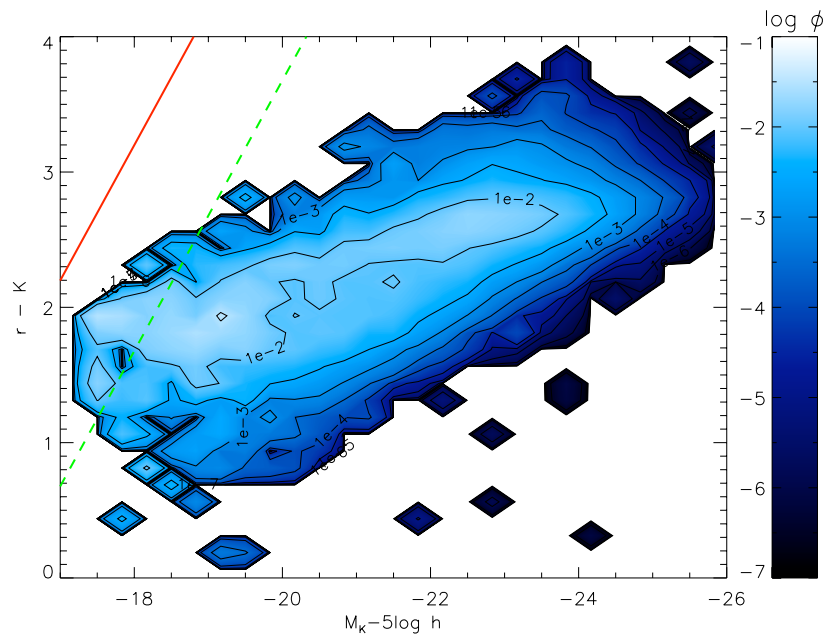


Figure 5.3: Space density of galaxies, in units of $h^3 \text{Mpc}^{-3} \text{mag}^{-2}$, as a function of K -band absolute magnitude and rest-frame $r - K$ Petrosian colour. Note that this is not a true colour since the apertures differ between the bands. The straight lines show the position on the plot of hypothetical galaxies at the r -band flux limit, with various $r - K$ colours, situated at $z = 0.01$ (upper-left line, solid, red) or at $z = 0.02$ (lower-right line, dashed, green). $z = 0.02$ corresponds to a survey volume of $1.06 \times 10^4 h^{-3} \text{Mpc}^3$.

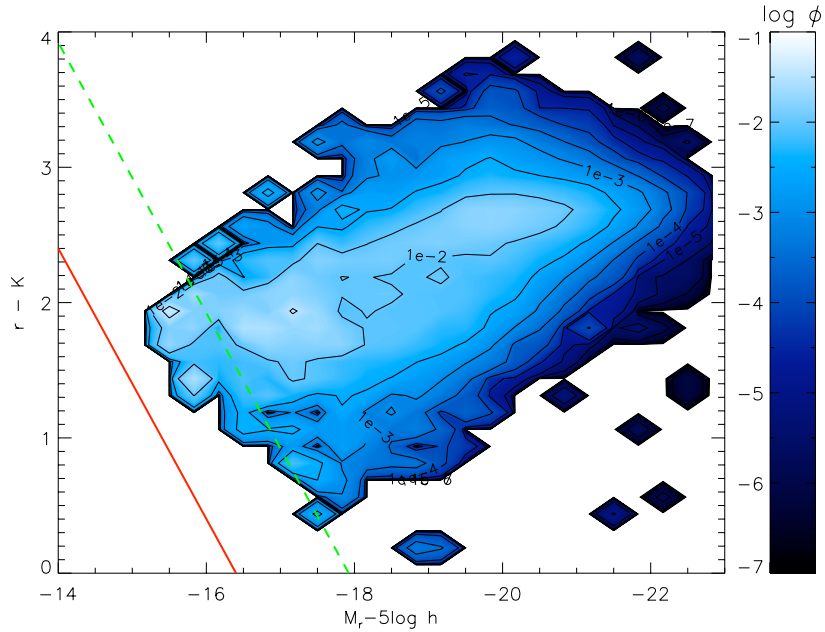


Figure 5.4: As Fig. 5.3 but for r -band absolute magnitude. The straight lines correspond to hypothetical galaxies at the K -band flux limit, with various $r - K$ colours, situated at $z = 0.01$ (lower-left line, solid, red) or at $z = 0.02$ (upper-right line, dashed, green).

A similar relation may be derived for the blue galaxies that would not be seen near the K -band flux limit. Considering a galaxy at redshift z with $K = 16$ (the K -band flux limit), the r -band absolute magnitude will be

$$M_r = r - DM(z) = K + (r - K) - DM(z). \quad (5.2)$$

Fig. 5.4 shows this relation between M_r and $r - K$ (with $K = 16$), from which it may be seen that blue galaxies will not be seen at faint r -band luminosity. Significant incompleteness is to be expected for $M_r - 5 \log h > -17.5$.

The limit in large K -band Petrosian radius also has an effect on the completeness in colour, as described in the following section.

5.1.2 Completeness in size

For the large radius limit, considering a source with a certain K -band magnitude and surface brightness, and assuming the Petrosian radius r_P is twice the effective radius, Equation (2.9) can be written as

$$\mu_{e,K} \simeq K + 2.5 \log 2\pi \left(\frac{r_P}{2} \right)^2. \quad (5.3)$$

For $K = 16$ and $\mu_{e,K} = 19.5 \text{ mag arcsec}^{-2}$, this corresponds to a limit in Petrosian radius of 4.0 arcsec, considerably less than the 6 arcsec limit intrinsic to the data; only for surface brightness

fainter than $20.38 \text{ mag arcsec}^{-2}$ does the 6 arcsec large radius limit dominate.

However, when the Petrosian magnitude limit of $r < 17.6$ is included, for certain $r - K$ colours, the limits in angular size and magnitude can give a stronger constraint on surface brightness. The above equation can be written as

$$\mu_{e,K} \simeq r - (r - K) + 2.5 \log 2\pi \left(\frac{r_P}{2}\right)^2. \quad (5.4)$$

For $r = 17.6 \text{ mag}$ and $r_P = 6 \text{ arcsec}$, this gives $\mu_{e,K} < 19.5 \text{ mag arcsec}^{-2}$ for $(r - K) > 2.48$. For these galaxies, the limit in radius is more significant than the (approximate) limit in K -band effective surface brightness. Figure 4.3 shows that many galaxies in the sample have $r - K > 3.5$ (observed-frame colours), corresponding to a surface brightness limit of $\mu_{e,K} \lesssim 18.5 \text{ mag arcsec}$. At the high-redshift limit of the sample, this would correspond to incompleteness for very red galaxies with absolute K -band effective surface brightness $\mu_{e,K,abs} \lesssim 18 \text{ mag arcsec}$ (see Section 4.2.4). The K -band low-surface brightness limit is therefore colour-dependent.

The effect of these limits may be seen more clearly by considering the two strongest constraints on the visibility of the galaxies: the faint r -band magnitude limit and the large K -band radius limit.

At any given redshift, only galaxies brighter than a certain absolute r -band magnitude are visible, and only galaxies smaller than a certain absolute K -band Petrosian radius are visible. These limits are found from Equations (4.2) and (4.3) by setting $r < 17.6 \text{ mag}$ and $r_P < 6 \text{ arcsec}$ (Petrosian radius). In this section, K -corrections are determined by assuming typical KCORRECT coefficients, as described in Section 4.2.1.

Fig. 5.5 shows the variation with redshift of the faint r -band absolute magnitude limit and the large physical K -band Petrosian radius limit, for the redshift range of the sample.

Fig. 5.6 shows the bivariate space density of the sample in r -band absolute magnitude and K -band physical Petrosian radius. Also shown is the boundary of the region in which galaxies would be visible, considering the limits in r -band apparent magnitude and K -band Petrosian radius. It can be seen that there is a sharp cut-off for large galaxies at medium or high luminosity, while galaxies that are both very luminous and very large (if they exist) would not be visible at all.

Fig. 5.7 shows the bivariate space density of the sample in K -band absolute magnitude and K -band physical Petrosian radius. The limits on visibility of galaxies with various rest-frame $r - K$ band colours are shown. Limits in K -band apparent magnitude are not taken into account, but Table 4.1 shows that this limit makes no difference to the vast majority of the sample.

For blue galaxies, the limit in K -band Petrosian radius makes no difference, except at the high-redshift end, as the limit in K -band effective surface brightness provides a stronger constraint. However, for red galaxies, the radius limit has a strong effect throughout the luminosity range, excluding from the sample those galaxies that have low surface brightness. At low luminosity,

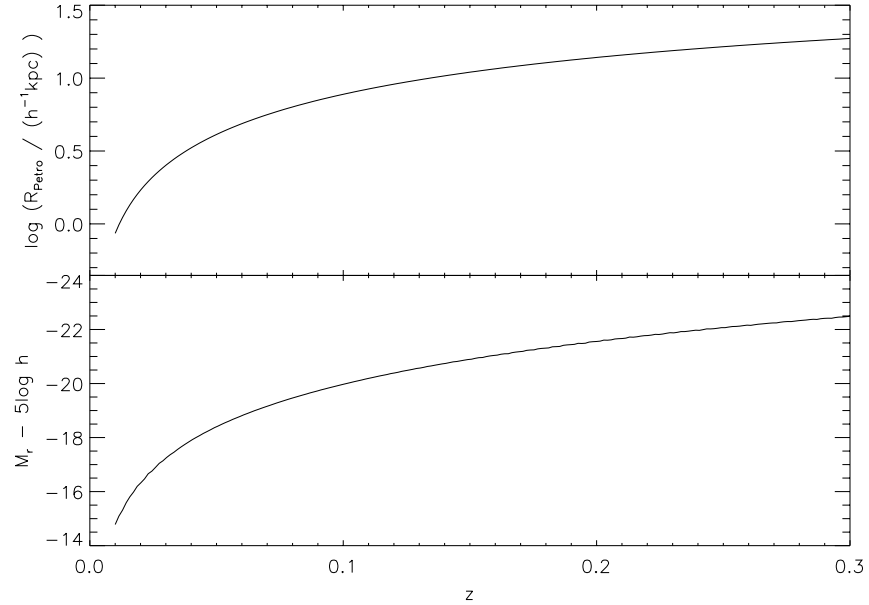


Figure 5.5: Limits in absolute r -band magnitude and physical K -band Petrosian radius as a function of redshift, determined using Petrosian magnitude $r < 17.6$, with typical K -corrections, and Petrosian radius $r_P > 6$ arcsec.

the r -band magnitude limit places a constraint on faint absolute K -band absolute magnitude, as is also seen in Fig. 5.3. At high luminosity, the limit in radius sets a maximum physical size to the galaxy, regardless of the luminosity, in both r and K . This corresponds to the maximum redshift of the sample ($z = 0.3$) and is given by Equation (4.3):

$$R = \frac{1000\pi r D_A(z)}{180 \times 3600} = \frac{1000\pi(6 \text{ arcsec}) D_A(0.3)}{180 \times 3600} = 18.7 h^{-1} \text{ kpc} \quad (5.5)$$

From the proximity of this limit to the shaded region in Fig. 5.7, it seems likely that there are some very large and very luminous galaxies that have been excluded because of this limit.

That the low-surface brightness limit in the K -band is colour-dependent can be seen by comparing the cut-off towards low surface brightness (towards the top-left of the plots) in Figs. 5.6 and 5.7.

It is difficult, if not impossible, accurately to quantify this incompleteness at low surface brightness, which will affect the low-surface brightness regions of the BBD as well as the luminosity function (see next section). However, from Fig. 5.7 it appears that, at medium or high luminosity, the low-surface brightness limit is well clear of the peak in the space density. So it seems reasonable to assume that the vast majority of the space density has been detected, for any given luminosity and for all colours, except at extremely high luminosity, where the high-redshift limit is significant.

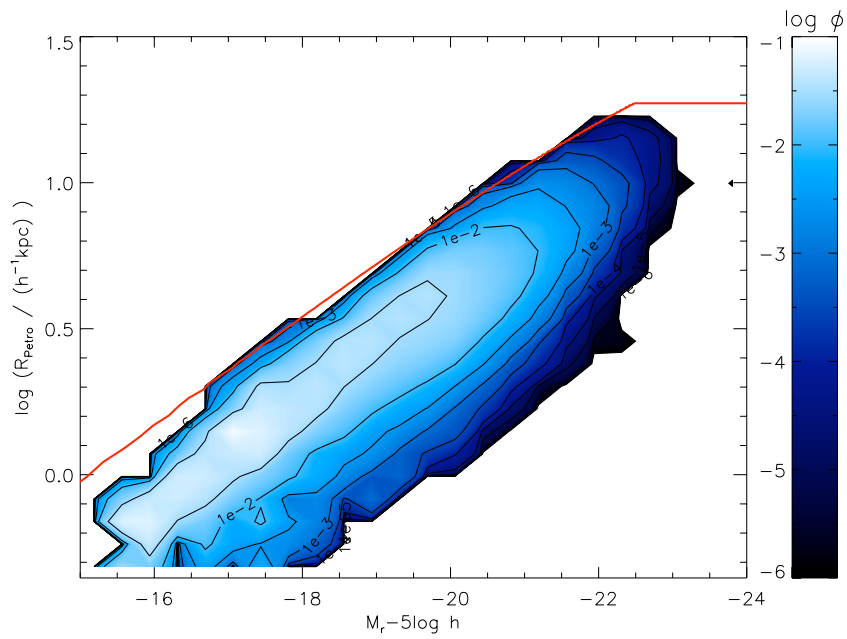


Figure 5.6: Bivariate space density of the sample in r -band absolute magnitude and K -band physical Petrosian radius. The solid red line shows the combined limit from the r -band absolute magnitude and K -band Petrosian radius, which are shown separately in Fig. 5.5. Only galaxies below and to the right of the line would be visible in the survey (the top-left of the plot is the low-surface brightness region). At high luminosity, the limit in radius becomes constant; this corresponds to the high-redshift limit of $z = 0.3$.

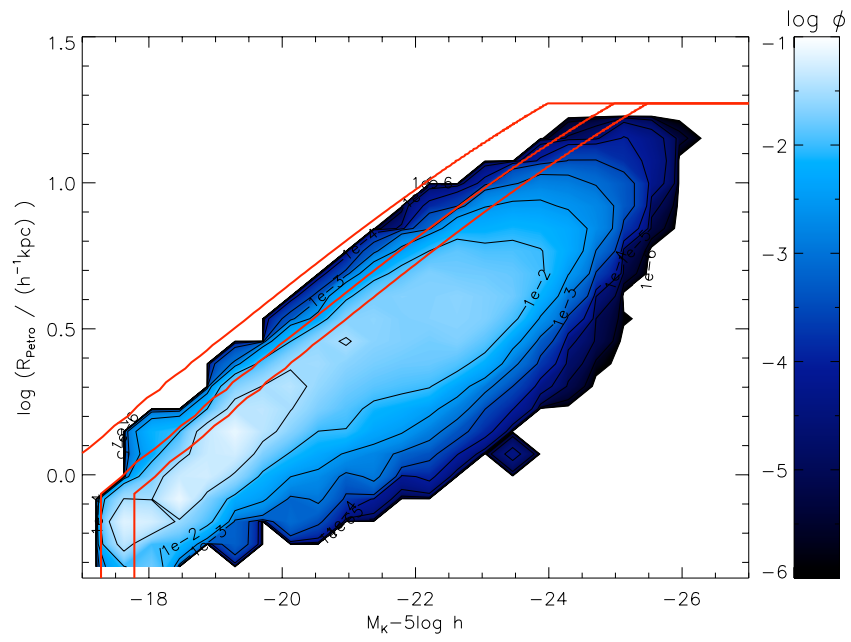


Figure 5.7: Bivariate space density of the sample in K -band absolute magnitude and K -band physical Petrosian radius. The solid red lines show the combined limit from the r -band absolute magnitude and K -band Petrosian radius for various rest-frame $r - K$ colours. Only galaxies below and to the right of the line would be visible in the survey. From the top-left to the bottom-right, the red lines correspond to $r - K = 1.5, 2.5$ and 3 , respectively.

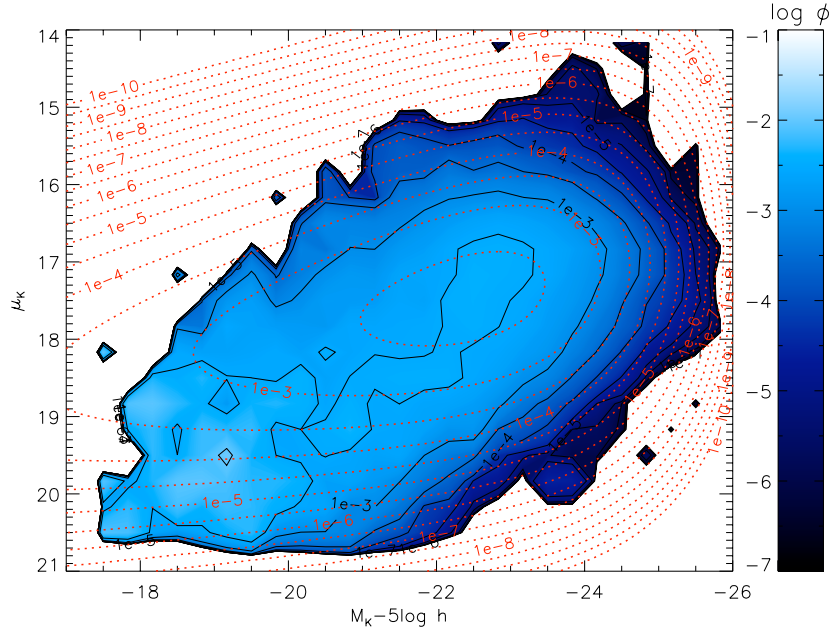


Figure 5.8: BBD with one-sigma errors subtracted, as Fig. 4.7.

This issue could be resolved with greater r -band depth and a higher maximum-redshift limit, while still maintaining the large-radius limit. However, the larger the range in redshift spanned by the sample the greater the uncertainty caused by modelling of the evolution corrections, so it would be preferable to relax the large-radius limit rather than the high-redshift limit. However, this would be possible only by running the LAS data through a different source-extraction pipeline. This will be discussed further in Chapter 6.

5.1.3 Chołoniewski fit

Fig. 5.2 also shows the best-fitting Chołoniewski fit to the BBD, performed as described in Section 4.7. It can be seen that the function provides a poor fit to the data, being unable to fit simultaneously the decline at high surface brightness and high luminosity, the broadening of the surface brightness distribution at faint luminosity, or the slope of the luminosity–surface brightness relation, which flattens at high luminosity. These features have also been seen in optical determinations of the BBD (e.g., Driver, 1999; Blanton et al., 2001; Cross & Driver, 2002; Driver et al., 2005)

Figs. 5.8 and 5.9 respectively show the BBD with one-sigma (χ^2) errors subtracted and added. Even within the uncertainties a Chołoniewski function still provides a poor fit.

The data used to find the fit are taken only from those regions of the BBD unlikely to be affected by significant incompleteness (see Sections 5.1.1 and 5.1.2). The limits chosen are $M_K - 5 \log h < -20$ and $\mu_{e,abs} < 19$, although it is clear by visual inspection that a different choice of

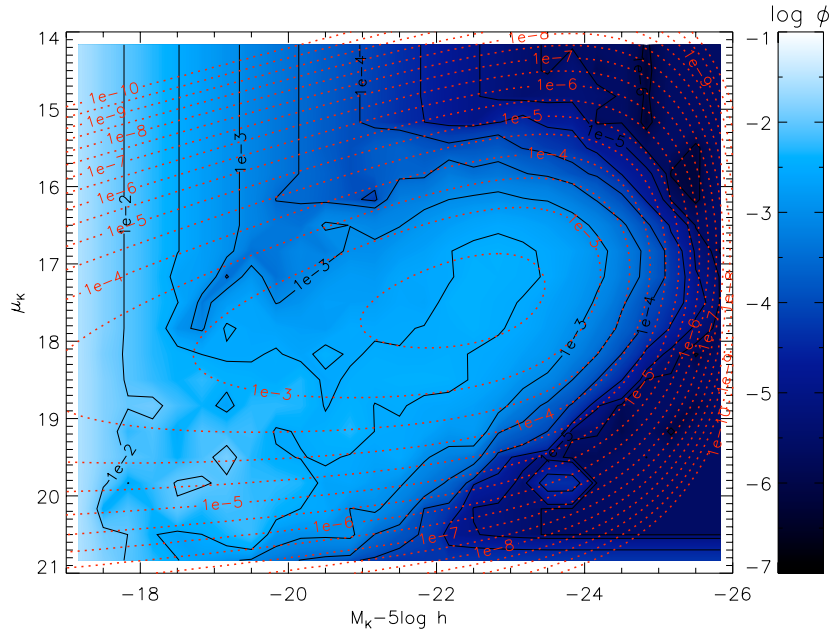


Figure 5.9: BBD with one-sigma errors added, as Fig. 4.8.

limits cannot yield a Chołoniewski function that fits the data well.

5.2 *K*-band luminosity function and luminosity density

The luminosity function is found by summing the galaxy weights ($w = 1/V_{\text{max}}$) in bins of luminosity. This is equivalent to a marginalization of the BBD along the surface brightness axis. As such, any incompleteness in surface brightness will also be reflected in the luminosity function. This will have an effect at all values of the luminosity, but the effect is likely to be small, except at low luminosity, where most galaxies have low surface brightness.

Fig. 5.10 shows the *K*-band luminosity function for the whole sample. The parameters of the best-fitting Schechter function correlate strongly, with $\text{corr}(M^*, \alpha) = 0.92$, $\text{corr}(M^*, \phi^*) = 0.97$ and $\text{corr}(\alpha, \phi^*) = 0.91$.

The most significant deviation from the published *K*-band luminosity functions is at the bright end, where my LF has a very steep decline compared with those of Bell et al. (2003b), Eke et al. (2005) and Jones et al. (2006). There are several possible explanations for this, for example:

1. Differences in the evolution corrections used (0.3 mag at $z = 0.3$ for my value of $Q = 1$), affecting the high-luminosity galaxies (cf. Fig. 5.1). Jones et al. (2006) applied no evolution corrections, although they have a redshift limit of $z < 0.2$. Eke et al. (2005), for $z < 0.12$, and Bell et al. (2003b) applied combined *K*- and evolution-corrections similar to my own. There is better agreement between my results and those of Kochanek et al. (2001), although

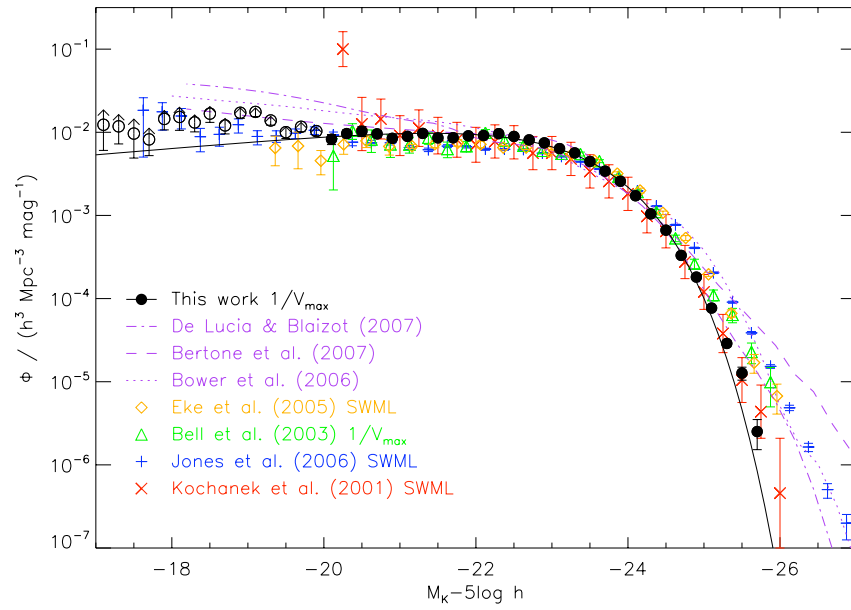


Figure 5.10: K -band luminosity function for the whole sample, with a compendium of published results from observations or semi-analytic models. Only the filled points are used in the Schechter function fit, i.e., $M_K - 5 \log h < -20$; the unfilled points are likely to suffer from some incompleteness of low-surface brightness or red, low-luminosity galaxies. Schechter function parameters are $M^* - 5 \log h = -23.17 \pm 0.04$, $\alpha = -0.81 \pm 0.04$ and $\phi^* = (0.0176 \pm 0.0009)h^3 \text{ Mpc}^{-3}$.

they applied no evolution corrections.

2. I have used Petrosian magnitudes, which are significantly fainter than total magnitudes for galaxies with a high Sérsic index (Graham et al., 2005), for example, by 0.24 mag for a de Vaucouleurs profile (Blanton et al., 2001). Kochanek et al. (2001) used isophotal magnitudes with a correction of 0.20 ± 0.04 mag to estimate total magnitudes, but a larger correction may be needed for the most luminous galaxies (Blanton et al., 2001), which tend to have a higher Sérsic index. Bell et al. (2003b) used the 2MASS Kron magnitudes with a correction of 0.1 mag, Eke et al. (2005) applied a similar correction based on the $J - K_S$ colour, while Jones et al. (2006) used the 2MASS total (extrapolated) magnitudes. However, this effect is countered to an extent by the effects of seeing, which, for a de Vaucouleurs profile, increases the fraction of the galaxy’s flux recovered by the Petrosian magnitude when the angular size of the galaxy is small (Blanton et al., 2001). This is the case for most of the luminous galaxies in the sample, which are generally observed at higher redshift.

3. Unidentified sources of incompleteness or a poorly understood selection function, given the non-trivial limits on my sample and the dependence on completeness in SDSS.

4. Improved photometry: significant magnitude errors at the bright end will lead to an overestimate of the space density at high luminosities (Jones et al., 2006). The results of Kochanek et al. (2001), Bell et al. (2003b), Eke et al. (2005) and Jones et al. (2006) are all based on 2MASS magnitudes, with much shallower imaging than the UKIDSS LAS. Moreover, the most luminous galaxies in my sample are all observed at magnitudes much brighter than the K -band flux limit, as can be seen from Fig. 5.1, so the photometric errors are very small.

Fig. 5.11 shows a comparison between K -band 2MASS Kron magnitudes and UKIDSS Petrosian magnitudes for sources in this sample with counterparts in the 2MASS XSC (Jarrett et al., 2000). It can be seen that a significant number of faint 2MASS sources have 2MASS Kron magnitudes that are much brighter than UKIDSS Petrosian magnitudes, sometimes by over 0.5 mag. If the UKIDSS magnitudes are more accurate, then this suggests that 2MASS luminosity functions will over-estimate the bright end, since high luminosity galaxies are most often found near the faint magnitude limit. For a more in-depth comparison between UKIDSS and 2MASS photometry, see Cross et al. (in preparation).

5. Missing deblended sources. My sample excludes all sources flagged as deblended in the LAS. If, as discussed in Section 4.1, these sources are in fact preferentially high-luminosity galaxies, then my LF would be underestimated at the bright end.

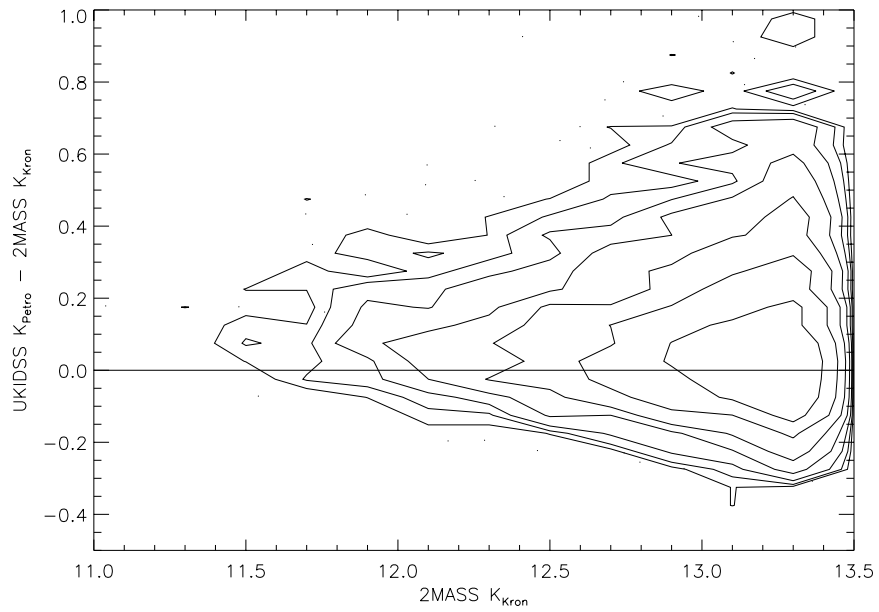


Figure 5.11: Residual K -band magnitudes between 2MASS Kron and UKIDSS Petrosian magnitudes for 5463 sources matched between my sample and the 2MASS XSC, with 2MASS $K_{\text{Kron}} < 13.5$ mag. Dust reddening corrections have not been applied. Contours show the density of sources on a logarithmic scale, with sources shown as points where the density is low. The mean offset is 0.096 mag, such that 2MASS magnitudes are brighter on average, with a standard deviation of 0.19 mag.

6. Incompleteness at very large physical radius. As shown in Fig. 5.7, the limits in K -band Petrosian radius and redshift combine to set an upper limit to the physical radius of galaxies in the sample. It appears likely that this will lead to some incompleteness for $M_K - 5 \log h < -24$, and significant incompleteness for $M_K - 5 \log h < -25.5$.

There is general agreement at the bright end of the (Petrosian) r -band LF (see below), which is consistent with the possible explanations listed above, although it does suggest that the effect of missing deblended galaxies is not the most significant factor. Further investigation is required to resolve this.

The steeper bright end leads to a fainter value of M^* compared with previous results, and consequently (given the strong correlation between M^* and α) a shallower value for the faint-end slope, α . It should be clear from Fig. 5.10 that the faint end of the Schechter function does not coincide with the faint end of the luminosity function. It should also be noted that the $1/V_{\max}$ method is sensitive to large-scale structure, which will be particularly apparent at the faint end of the LF.

The Schechter function fit used only those points with $M_K - 5 \log h < -20$. However the resulting parameters of the Schechter function are largely independent of the range of values included in the fit (see Fig. 6.1).

At intermediate magnitudes, $M_K - 5 \log h \simeq -22$, my LF is noticeably higher than those from the literature. This is likely to be due to an overdensity of galaxies at $z \lesssim 0.1$, as seen in Fig. 5.1.

The K -band luminosity density, with jackknife errors, is found to be $j = (6.375 \pm 0.078) \times 10^8 L_{\odot} h \text{ Mpc}^{-3}$ from the Schechter function, or $j = (6.500 \pm 0.073) \times 10^8 L_{\odot} h \text{ Mpc}^{-3}$ from the galaxy weights. The true luminosity density is likely to be higher than these values since the Schechter function provides a poor fit at both the faint end and the bright end, and since there are known sources of incompleteness in absolute magnitude, colour and surface brightness at the faint end. For comparison, measurements of the luminosity density from the published results shown in Fig. 5.10 lie within approximately $(5.8-7.6) \times 10^8 L_{\odot} h \text{ Mpc}^{-3}$ (see Jones et al., 2006, fig. 15).

5.3 r -band luminosity function and luminosity density

The r -band luminosity function is found using the same sample of galaxies as above, but with the weights ($1/V_{\max}$) summed in bins of r -band luminosity.

Fig. 5.12 shows the r -band luminosity function. While there is excellent agreement at the bright end with the LF of Blanton et al. (2003b), my LF is over-dense at intermediate redshifts,

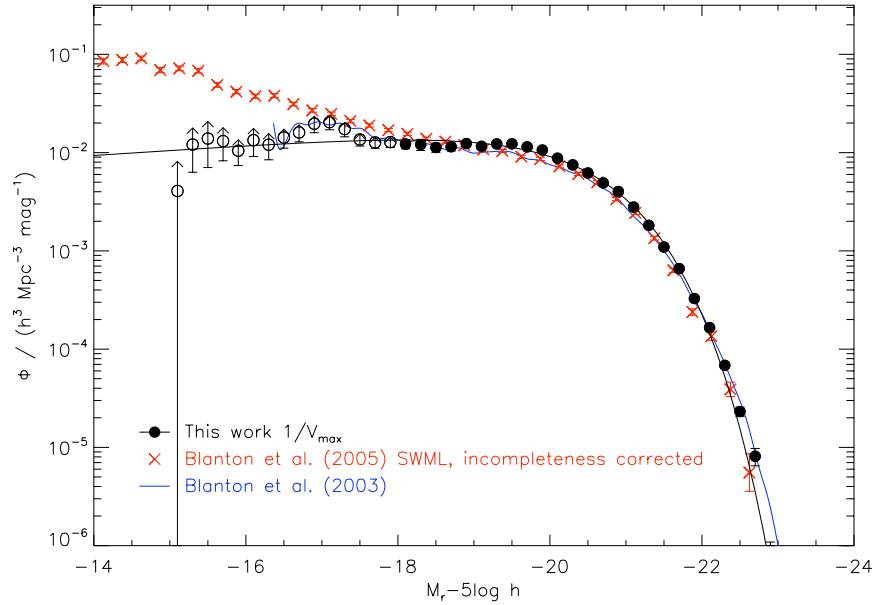


Figure 5.12: r -band luminosity function. Only the filled points are used in the Schechter function fit, i.e., $M_r - 5 \log h$ brighter than -18 ; the unfilled points are likely to suffer from some incompleteness of low-surface brightness galaxies or low-luminosity blue galaxies. The LF of Blanton et al. (2003b) has been adjusted from the $^{0.1}r$ -band at $z = 0.1$ to the r -band at $z = 0$ by (1) shifting Qz mag fainter in magnitude, with $Q = 1.62$ and $z = 0.1$, (2) reducing the number density by $0.4Pz$ dex, with $P = 0.18$, and (3) making the LF 0.22 mag brighter in magnitude, to convert from $^{0.1}r$ to r . Schechter function parameters are $M^* - 5 \log h = -20.32 \pm 0.04$, $\alpha = -0.87 \pm 0.05$ and $\phi^* = (0.0216 \pm 0.0010)h^3 \text{ Mpc}^{-3}$.

probably due to large-scale structure. The deficit of blue low-luminosity galaxies, identified in Fig. 5.4, is clearly evident.

The r -band luminosity density is found to be $j = (1.964 \pm 0.028) \times 10^8 L_\odot h \text{ Mpc}^{-3}$ by extrapolating the Schechter function, or $j = (1.977 \pm 0.021) \times 10^8 L_\odot h \text{ Mpc}^{-3}$ from the galaxy weights, assuming a solar absolute magnitude of 4.64 (Blanton & Roweis, 2007). Again, the true luminosity density is likely to be higher than these values given the incompleteness at the faint end. This is somewhat higher than the r -band $z = 0$ luminosity density of Blanton et al. (2003b), $-15.90 + 2.5 \log h$ mag in a Mpc^3 , or $1.64 \times 10^8 L_\odot h \text{ Mpc}^{-3}$.

Given that I use the same source of data as Blanton et al. (2003b), but over a smaller area, with a more complex selection function and with an inferior luminosity function estimator, the r -band results I find should not be interpreted as being more than a consistency check on my analysis.

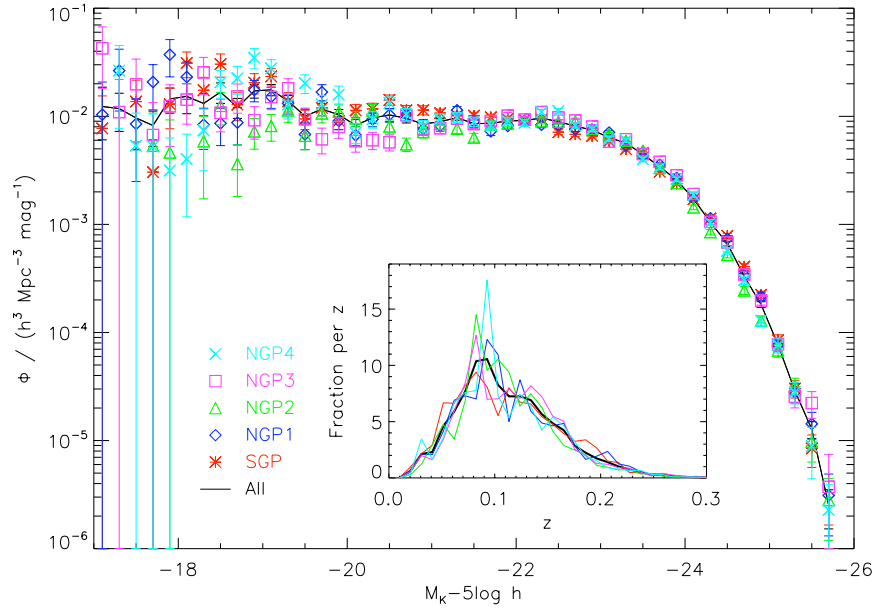


Figure 5.13: K -band LF for the NGP and SGP regions, also showing the LF for the whole sample. Jackknife errors are shown for the whole sample and Poisson errors for the sub-samples. The inset shows the redshift distribution of galaxies in each sub-sample, with the line colours corresponding to the colours of the symbols in the main figure.

5.4 Large-scale structure

In order to show the effect of large-scale structure on my results, I have estimated the luminosity function for the five principal subdivisions of the sample (see Fig. 4.1).

Fig. 5.13 shows the K -band LF for the SGP region and the four NGP regions, together with the LF for the whole sample and the redshift distribution for each sub-sample. I note the following:

1. There is a large scatter at low luminosities, illustrating the limitations of the $1/V_{\text{max}}$ method.
2. Each sub-sample is normalized according to the number of galaxies rather than the area covered, so a genuine over-density at a certain redshift will be artificially compensated for by an apparent under-density at other redshifts, and *vice versa*. Given the correlation between redshift and absolute luminosity (Fig. 5.1), this means that large-scale structure at intermediate redshifts will distort the luminosity function at both ends. This may be seen for the SGP region, where an under-density at $z \simeq 0.1$, seen at $M_K - 5 \log h$ between -22.5 and -23 , may have led to an over-estimate of the LF at both the faint end and the bright end, while for the NGP2 region, an over-density at $z \lesssim 0.1$ ($M_K - 5 \log h \simeq -22.5$), the Sloan Great Wall (Gott et al., 2005), may have led to an under-estimate of the LF at both ends.

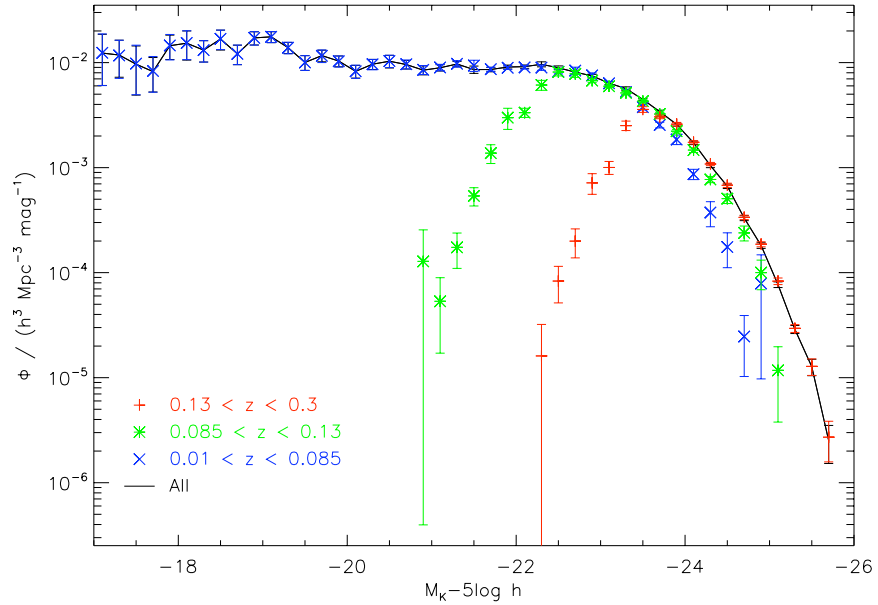


Figure 5.14: K -band luminosity function for galaxies in three different redshift intervals, also showing the LF for the whole sample.

5.5 Variation with redshift

In order to identify further sources of bias I split the sample into three bins in redshift, containing approximately equal numbers of galaxies.

Fig. 5.14 shows the K -band LF estimated for the low-, mid- and high-redshift sub-samples. There is disagreement at the bright end between the different slices, with ϕ getting progressively higher at higher redshift. This trend could be a result of the evolution corrections being too small, but the trend is still present when a correction as strong as $Q = 2$ is applied. It is more likely to be a result of (1) an over-simplistic form for the evolution corrections, $E(z) = Qz$, independent of galaxy type, (2) a decreasing apparent Sérsic index with decreasing angular size, as a result of convolution with the PSF, causing a greater fraction of a galaxy's flux to be recovered by the Petrosian magnitude when the galaxy is observed at higher redshift, (3) poorly understood limits to the sample, or (4) large-scale structure affecting both the shape and normalization of the LF for each redshift slice.

This test does make it clear that my results are to an extent dependent on the redshift limits chosen: choosing a lower value than 0.3 for the maximum redshift would have given an even steeper bright-end slope for the LF. This is likely to be due to the limit on the large physical radius, which is dependent on the maximum redshift of the sample; this translates to an approximate limit to the maximum luminosity, given the correlation between luminosity and physical size.

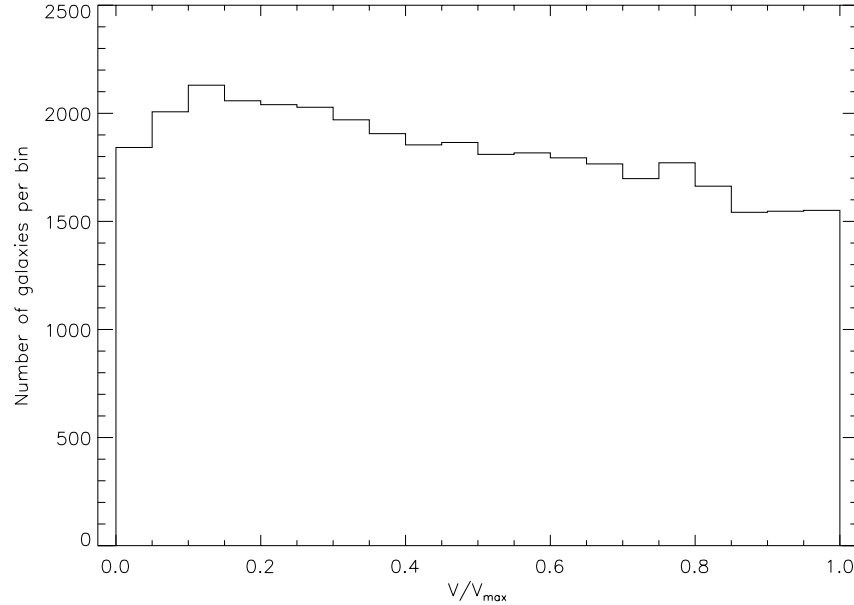


Figure 5.15: Histogram of V/V_{\max} for 36 659 galaxies in the sample. 425 of the galaxies were assigned values of the redshift limits such that $V > V_{\max}$; these have been given values of $V/V_{\max} = 1$.

Fig. 5.15 shows the values of V/V_{\max} for the sample. The mean value is $\langle V/V_{\max} \rangle = 0.476$, which is 16σ below the expected value of 0.5, using $\sigma = (12n)^{-1/2}$ and $n = 36659$. (This result remains the same with $Q = 2$.) This provides clear evidence for some incompleteness, or some inadequacies with the modelling of the evolution corrections, or both.

5.6 Subdividing the sample

The bimodality of the galaxy population has been recognized by many authors (see [Driver et al., 2006](#); [Ball et al., 2006](#), and references therein). This may be visualized by subdividing the LF or BBD in various ways, for example, according to colour, concentration or spectral class. Of these properties, I find, following [Driver et al. \(2006\)](#), that the $u - r$ core (PSF) colour gives a particularly sharp dichotomy.

The space density for each sub-population is estimated by summing the weights of galaxies of that type, with jackknife errors estimated subsequently.

5.6.1 Subdividing by colour

[Driver et al. \(2006\)](#) demonstrated that the central (PSF) $u - r$ colour is very effective for subdividing the sample. They suggest that this is because it reflects the colour of the bulge of the system,

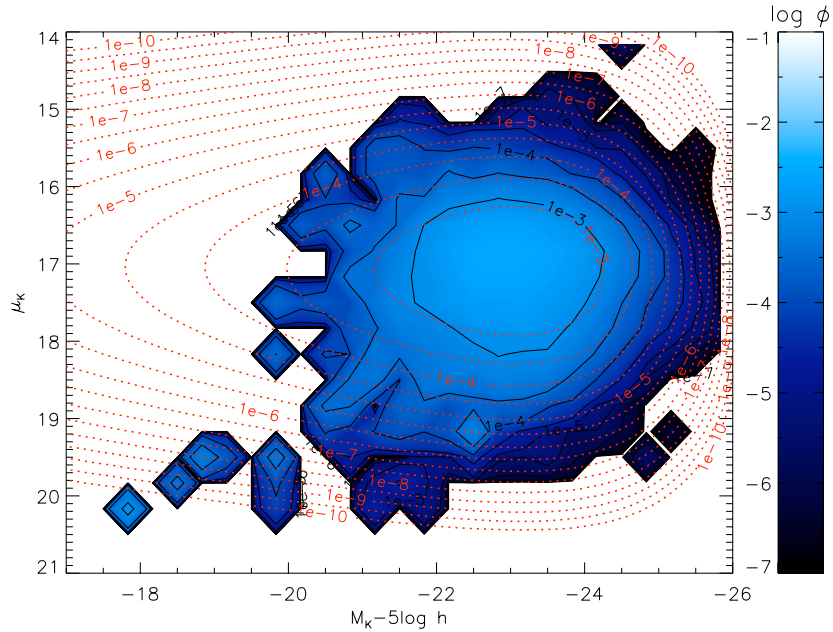


Figure 5.16: BBD for red galaxies, with $(u - r)_{\text{PSF}} > 2.35$. The best-fitting Chołoniewski function, estimated using $M_K - 5 \log h < -20$ and $\mu_{e,\text{abs}} < 19$, is shown by the red dotted contours. Parameters of the fit are $M^* - 5 \log h = -22.88$ mag, $\alpha = 0.17$, $\phi^* = 0.0121 h^3 \text{ Mpc}^{-3}$, $\mu_{e,\text{abs}}^* = 17.09 \text{ mag arcsec}^{-2}$, $\sigma_{\mu_{e,\text{abs}}} = 0.570 \text{ mag arcsec}^{-2}$ and $\beta = -0.012$.

rather than the combined colour of the bulge and disc.

Figs. 5.16–5.18 show the K -band BBD and K - and r -band LFs split by the SDSS rest-frame $u - r$ PSF colour with $u - r > 2.35$ or $u - r < 2.35$ for red and blue galaxies respectively.

The BBD for red galaxies, excluding outliers, shows no evidence of a correlation between luminosity and surface brightness, while the BBD for blue galaxies shows no flattening off of the luminosity–surface brightness relation at high luminosities, suggesting that this division reflects a property of the underlying population. Moreover, the Chołoniewski function appears to fit the blue BBD much better than it fits the BBD for the whole sample. However, I caution that the lack of red-core galaxies with $M_K - 5 \log h > -20$ could be a symptom of the incompleteness identified in Fig. 5.3, while the lack of such galaxies with $\mu_{e,\text{abs}} > 19.5 \text{ mag arcsec}^{-2}$ could be due to the low-surface brightness limit for de Vaucouleurs profile galaxies.

The LFs show a sharp division, with red-core galaxies more abundant than blue-core galaxies by an order of magnitude at high luminosity (and *vice versa* at low luminosity), and with the bright end of the LF around 0.8 mag more luminous in the K -band for red-core galaxies.

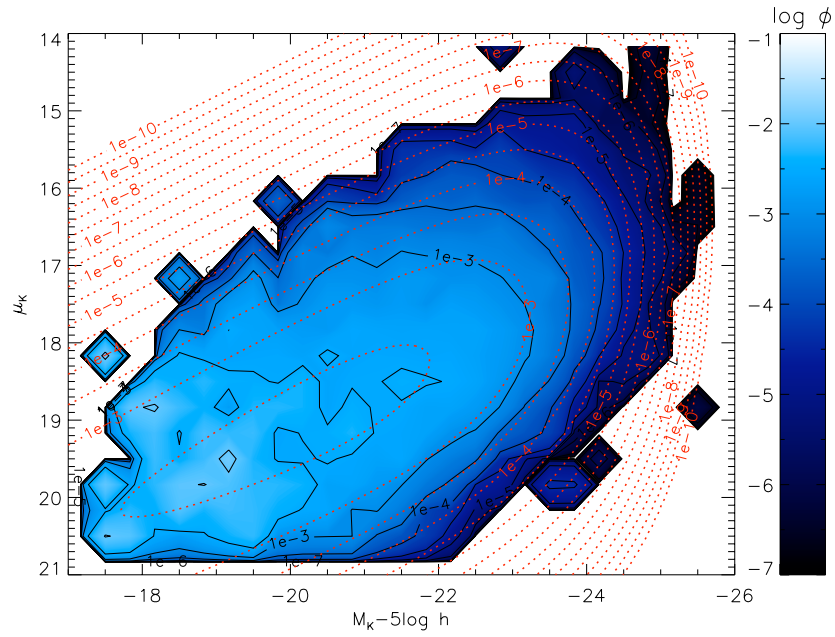


Figure 5.17: BBD for blue galaxies, with $(u - r)_{\text{PSF}} < 2.35$. The best-fitting Choloniewski function, estimated using $M_K - 5 \log h < -20$ and $\mu_{e,\text{abs}} < 19$, is shown by the red dotted contours. Parameters of the fit are $M^* - 5 \log h = -22.52$ mag, $\alpha = -0.86$, $\phi^* = 0.0131 h^3 \text{ Mpc}^{-3}$, $\mu_{e,\text{abs}}^* = 17.94 \text{ mag arcsec}^{-2}$, $\sigma_{\mu_{e,\text{abs}}} = 0.807 \text{ mag arcsec}^{-2}$ and $\beta = 0.463$.

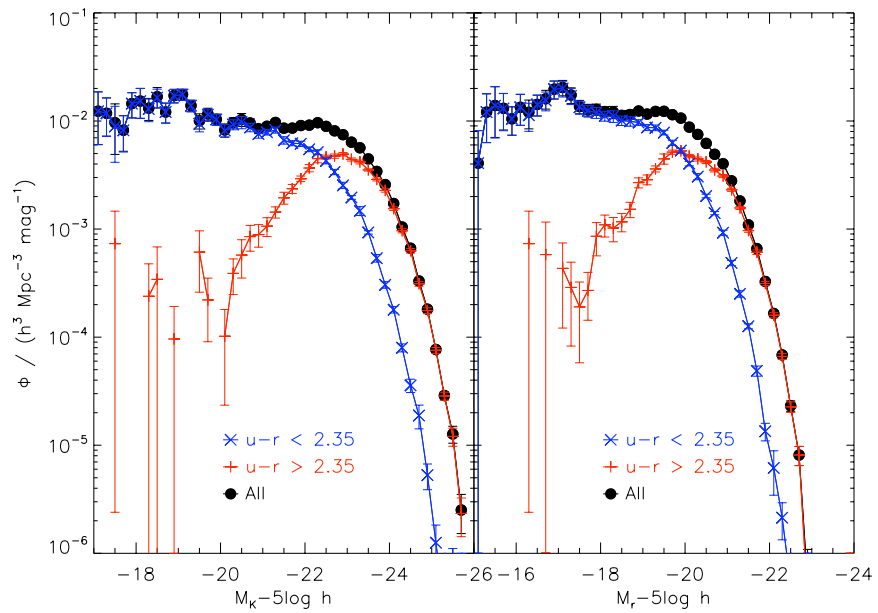


Figure 5.18: K -band (left) and r -band (right) LFs for red and blue galaxies, showing the total LF as well.

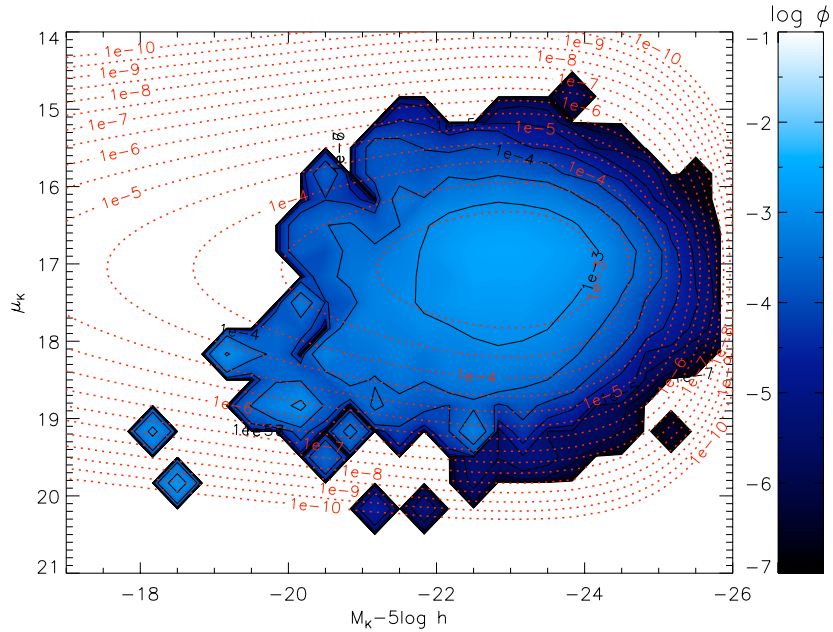


Figure 5.19: BBD for galaxies with early spectral type, defined as $eClass < -0.1$. As Fig. 5.16. Parameters of the fit are $M^* - 5 \log h = -22.91$ mag, $\alpha = 0.09$, $\phi^* = 0.0103 h^3 \text{ Mpc}^{-3}$, $\mu_{e,abs}^* = 17.08$ mag arcsec $^{-2}$, $\sigma_{\mu_{e,abs}} = 0.553$ mag arcsec $^{-2}$ and $\beta = -0.004$.

5.6.2 Subdividing by spectral type

The SDSS eClass (Connolly et al., 1995; Connolly & Szalay, 1999) is a continuous measure of the spectral class of the spectral class of a galaxy, based on a principal component analysis of SDSS spectra. Ball et al. (2006) demonstrated that this can be used effectively to subdivide the galaxy population.

Figs. 5.19 and 5.20 show the K -band BBD for early and late spectral type respectively, while Fig. 5.21 shows the K -band and r -band LFs for the subdivisions. It can be seen that the same general trends are present, as when the sample is subdivided by colour, although the separation is less pronounced at the bright end of the luminosity functions.

5.6.3 Subdividing by concentration

The SDSS inverse concentration index (e.g., Blanton et al., 2001) may also be used to subdivide the sample (Ball et al., 2006). This is defined as $CI_{inv} = R_{50}/R_{90}$, where R_{50} and R_{90} are respectively the radii containing 50 and 90 per cent of the r -band Petrosian flux. An inverse concentration index is chosen so that $0 < CI_{inv} < 1$. In order to give a clear distinction, I divide the sample here into high concentration ($CI_{inv} < 0.36$) and low concentration ($CI_{inv} > 0.36$).

Figs. 5.22 and 5.23 show the K -band BBD for high and low concentrations respectively, while

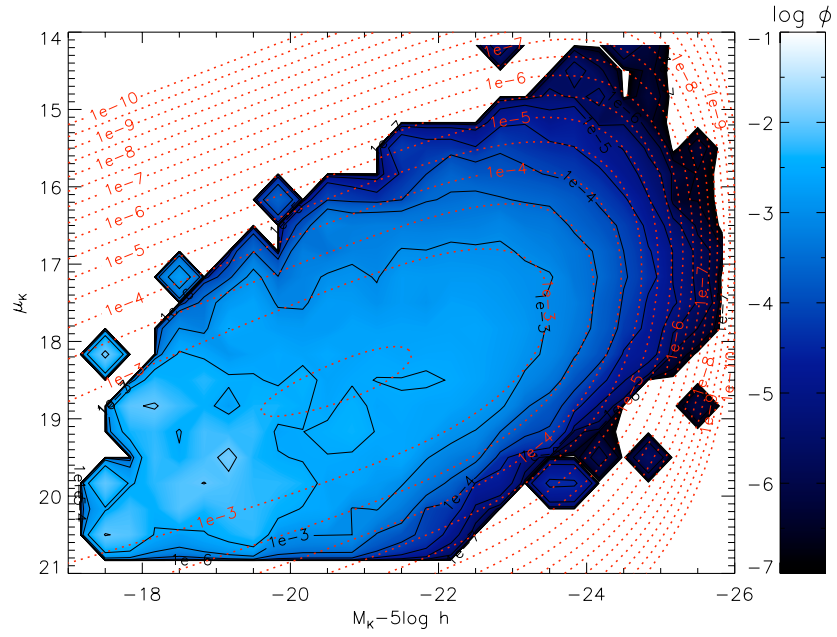


Figure 5.20: BBD for galaxies with late spectral type, defined as $\text{eClass} > -0.1$. As Fig. 5.16. Parameters of the fit are $M^* - 5 \log h = -22.93 \text{ mag}$, $\alpha = -0.86$, $\phi^* = 0.0109 h^3 \text{ Mpc}^{-3}$, $\mu_{\text{e,abs}}^* = 17.71 \text{ mag arcsec}^{-2}$, $\sigma_{\mu_{\text{e,abs}}} = 0.794 \text{ mag arcsec}^{-2}$ and $\beta = 0.355$.

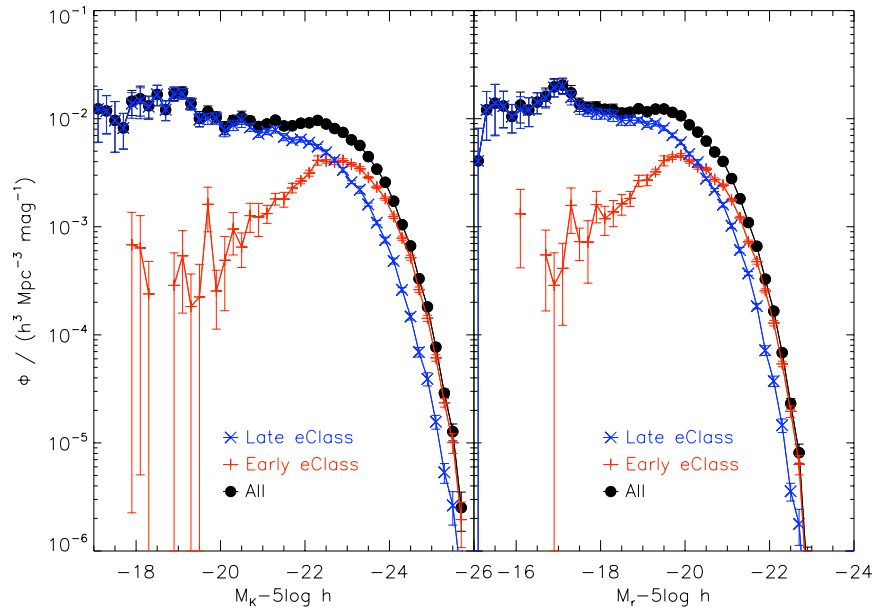


Figure 5.21: As Fig. 5.18 but for early and late spectral class.

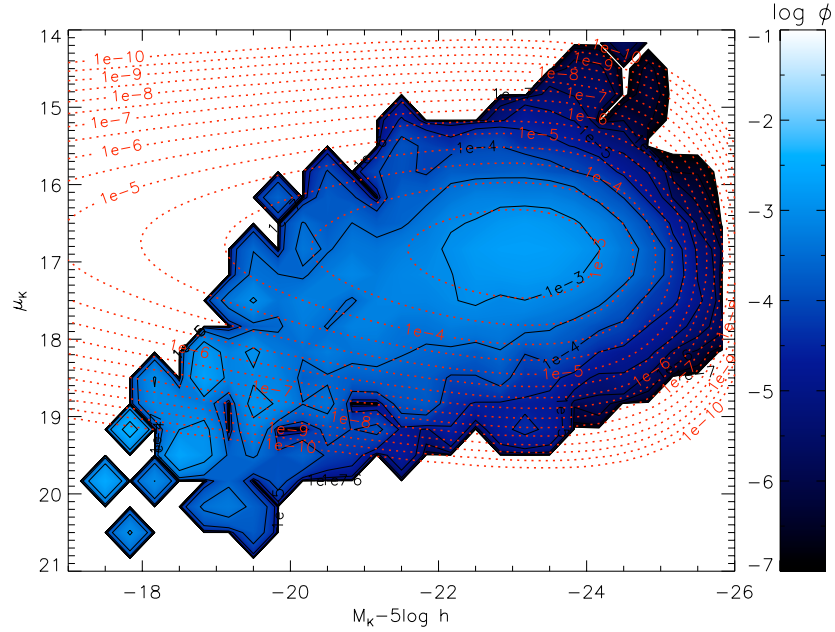


Figure 5.22: BBD for galaxies with high concentration, as Fig. 5.16. Parameters of the fit are $M^* - 5 \log h = -22.98 \text{ mag}$, $\alpha = 0.12$, $\phi^* = 0.0072 h^3 \text{ Mpc}^{-3}$, $\mu_{e,\text{abs}}^* = 16.89 \text{ mag arcsec}^{-2}$, $\sigma_{\mu_{e,\text{abs}}} = 0.476 \text{ mag arcsec}^{-2}$ and $\beta = -0.029$.

Fig. 5.24 shows the K -band and r -band LFs for the subdivisions. Once again, the same general trends are apparent, as with previous subdivisions, but there is greater overlap between the populations.

5.7 Stellar mass function

Stellar masses are derived from the KCORRECT template fits, which are based on the $ugriz$ photometry. Described in more detail by Blanton & Roweis (2007), these templates are generated from Bruzual & Charlot (2003) stellar population synthesis models with a Chabrier (2003) initial mass function (IMF).

The M/L ratio varies less at near-infrared than at optical wavelengths (Bell & de Jong, 2001) so, assuming the uncertainty in stellar mass is dominated by the uncertainty in the M/L ratio, it makes sense to estimate the stellar mass from the near-infrared absolute magnitude and the M/L ratio at that wavelength. Ideally the M/L ratio could be found by fitting a template to all available photometry, i.e., $ugrizYJHK$ for SDSS and UKIDSS. However, for this analysis, where I have good optical colours (with consistent apertures) but poor near-infrared and optical–near-infrared colours, I find it is best to estimate the K -band M/L ratio using the optical colours only.

Fig. 5.25 shows the stellar mass function, with stellar masses estimated from the K -band

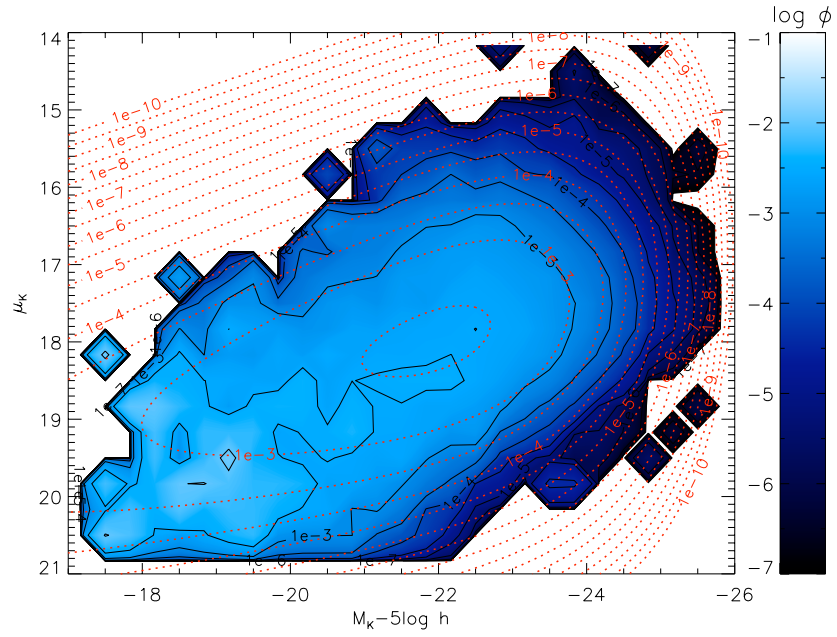


Figure 5.23: BBD for galaxies with low concentration, as Fig. 5.16. Parameters of the fit are $M^* - 5 \log h = -22.73 \text{ mag}$, $\alpha = -0.52$, $\phi^* = 0.0166 h^3 \text{ Mpc}^{-3}$, $\mu_{e,\text{abs}}^* = 17.77 \text{ mag arcsec}^{-2}$, $\sigma_{\mu_{e,\text{abs}}} = 0.717 \text{ mag arcsec}^{-2}$ and $\beta = 0.275$.

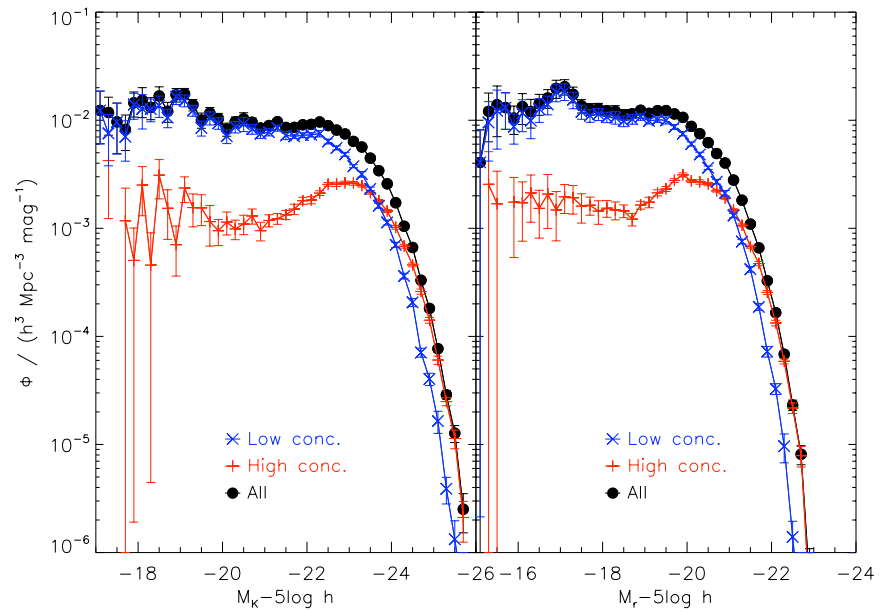


Figure 5.24: As Fig. 5.18 but for high- or low-concentration galaxies.

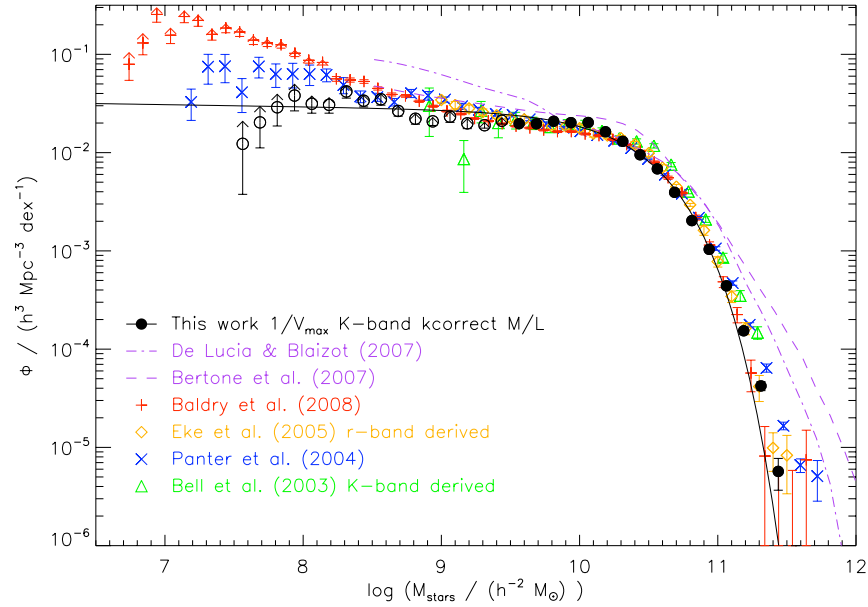


Figure 5.25: Stellar mass function. Only the filled points are used in the Schechter function fit, i.e., stellar mass greater than $10^{9.5}h^{-2}M_{\odot}$; the unfilled points are likely to suffer from some incompleteness of low-surface brightness galaxies. Masses calculated from the K -band KCORRECT mass-to-light ratios have been increased by 0.1 dex. Schechter function parameters are found to be $\log(\mathcal{M}^*h^2/M_{\odot}) = 10.44 \pm 0.02$, $\alpha = -1.02 \pm 0.04$ and $\phi^* = (0.0112 \pm 0.0007)h^3 \text{ Mpc}^{-3}$. Stellar masses based on other IMFs have been reduced for comparison with my results, based on the Chabrier IMF: Salpeter IMF (Panter et al., 2004) by 0.3 dex, ‘diet’ Salpeter (Bell et al., 2003b) by 0.15 dex, and no conversion has been applied for the Kennicutt (Eke et al., 2005) or Kroupa (Baldry et al., 2008) IMFs.

absolute magnitudes and the K -band mass-to-light (M/L) ratios from KCORRECT. At the high-mass end my results agree well with previously-published stellar mass functions, while at the low-mass end the discrepancy could be a result of incompleteness or large-scale structure. The underdensity at intermediate masses ($\mathcal{M} \simeq 10^9h^{-2}M_{\odot}$) could be due to inappropriate M/L ratios or large-scale structure.

Fig. 5.26 shows the stellar mass function calculated using the default mass and the various mass-to-light ratios, all given by KCORRECT. I found there to be an offset in the masses derived from the K -band M/L ratios compared with the masses derived from the optical bands. The precise cause of this offset is not known, but it has been compensated for by increasing the K -band-derived masses by 0.1 dex. The greater uncertainty in blue M/L ratios would be expected to stretch the high-mass end of the SMF, as seen in the u - and g -bands SMFs. Smaller uncertainty in the red or near-infrared M/L ratios may be responsible for the disagreement at

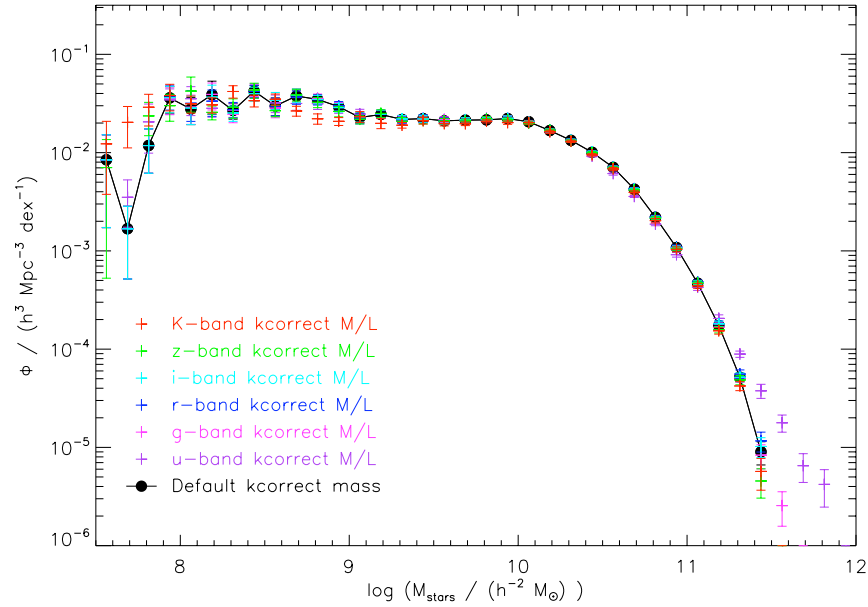


Figure 5.26: Stellar mass function, using various mass-to-light ratios and the default KCORRECT mass, which is derived from the template fit to the input (*ugriz*) absolute magnitudes. Masses calculated from the *K*-band KCORRECT mass-to-light ratios have been increased by 0.1 dex.

low masses. However, there is some uncertainty in the *K*-band *M/L* ratios caused by the emission from thermally-pulsating asymptotic giant branch (TP-AGB) stars, which are very difficult to model (Bruzual, 2007).

The stellar mass density is found to be $(3.12 \pm 0.05) \times 10^8 h M_{\odot} \text{ Mpc}^{-3}$ by extrapolating the Schechter function or $(3.09 \pm 0.04) \times 10^8 h M_{\odot} \text{ Mpc}^{-3}$ from the galaxy weights. Due to incompleteness this is likely to be an underestimate, and a different IMF could increase this substantially, for example, by a factor of 2 for a Salpeter IMF.

5.8 Summary

The results presented in this chapter appear to confirm previous findings for the (optical) bivariate brightness distribution, *K*- and *r*-band luminosity functions and the stellar mass function. However, various sources of uncertainty have been identified, which suggest that additional work is needed in order to obtain precise results for the space density of galaxies from the LAS.

Chapter 6

Conclusions

Il n'est pas certain que tout soit incertain.

Having already compared my results with those of other authors in Chapter 5, I now consider some implications of those findings along with various possible improvements and extensions.

6.1 Summary of results and interpretation

I have presented an estimate of the space density of galaxies from the UKIDSS Large Area Survey. The $1/V_{\max}$ space density estimator has been used in a multi-dimensional form to produce results for the K -band and r -band luminosity functions and the stellar mass function that are consistent with previous findings. I have presented the first K -band bivariate brightness distribution in K -band absolute magnitude and effective surface brightness, which shows similar trends to the optical BBD.

The multiple limits on the survey have been taken into account. For example, deblended sources have been excluded, and limits in K -band and r -band magnitude, K -band Petrosian radius and K -band surface brightness have been used to estimate the volume within which each galaxy would have been visible.

Given the general consistency with previous results, my findings serve to reinforce the conclusions of others, which are described here.

6.1.1 Interpreting the luminosity function

The LF in both the K - and r -bands has been found to be close to a Schechter function. The exponential decline at high luminosity suggests a galaxy formation mechanism that limits the maximum permissible size of a galaxy, while the almost flat slope at low luminosity is likely to reflect a low star formation efficiency in low-mass objects (Baldry et al., 2008). Both of these features can be

explained by feedback: from AGN, limiting the maximum size, or from supernovae, expelling the gas from small galaxies and thus limiting their star formation (see Section 1.4.3).

6.1.2 Interpreting the bivariate brightness distribution

The BBD displays a clear correlation between luminosity and surface brightness, which flattens at high luminosity and broadens at low luminosity. Along with the associated luminosity–size correlation, this has been seen previously by other authors (see Sections 1.7.2 and 5.1.3).

These relations have been associated with various theoretical models for galaxy formation. The properties of a galaxy are expected to be related to the properties of the dark matter halo in which it formed. So the mass of the galaxy will be related to the mass of the halo, and the size (and hence the surface brightness) of the galaxy disc will be related to the angular momentum of the halo, as described by the dimensionless spin parameter of the dark matter halo (Peebles, 1969). This connection has been made in various studies, with good agreement with observations (e.g., Fall & Efstathiou, 1980; Dalcanton et al., 1997; Mo et al., 1998; de Jong & Lacey, 2000; Bell et al., 2003a; Somerville et al., 2008).

Freeman’s law for galaxy discs (Freeman, 1970) expresses the observation that galaxy discs are found to have essentially the same surface brightness, independent of luminosity. However, this observation can be explained as a selection effect (e.g., Phillipps & Disney, 1986). de Jong & Lacey (2000), by fitting a Chołoniewski function to their data, have found a relation between the luminosity and the effective radius for spiral galaxies. This is shown by Driver et al. (2005, table 2, fig. 19) to be equivalent to an increasing surface brightness with increasing luminosity. For galaxy spheroids, the Kormendy relation (Kormendy, 1977) relates the effective surface brightness to the effective radius. This may be converted to a surface brightness–luminosity relation (using $L = 2\pi\mu_e r_e^2$, in appropriate units), and corresponds to higher luminosity galaxies having a lower surface brightness (Driver et al., 2005). For elliptical galaxies, the Kormendy relation may be understood as a projection of the Fundamental Plane onto the plane of effective radius and surface brightness (Waddington et al., 2002).

The observed BBD may arise through a combination of these trends, and as a result of a change in the typical bulge-to-disc ratio as a function of luminosity, with higher luminosity galaxies generally being increasingly bulge-dominated (Driver et al., 2005).

6.1.3 Interpreting the bimodality

Results presented in Section 5.6 suggest that core colour, spectral class and concentration each reflect the bimodality of the galaxy population (see also Sections 1.4.2 and 1.7.2). This is consistent

with previous results that have found that galaxies may be divided into two broadly distinct but overlapping populations: one consisting of early-type, red, luminous, passive, concentrated galaxies, and the other consisting of late-type, blue, low-luminosity, star-forming, low-concentration galaxies (Driver et al., 2006; Ball et al., 2006), with the fraction of red or early-type galaxies increasing towards higher-density environments (Baldry et al., 2006; Bamford et al., 2008).

Driver et al. (2006, 2007a,b) have shown that the bimodality of the galaxy population is strongly related to the two-component nature of galaxies, with (classical) bulge-dominated galaxies forming the red population and disc(+pseudobulge)-dominated galaxies forming the blue population. (See Kormendy & Kennicutt, 2004, and Section 1.3.5 for an explanation of the difference between classical bulges and pseudobulges.) By performing bulge-disc decomposition on a large sample of galaxies, they have shown that the dichotomy between the bulge population and the disc population is stronger than the dichotomy between the red population and the blue population. This suggests that bulges and discs have different formation mechanisms and histories.

The (classical) bulge-disc nature of galaxies would not in itself produce the observed bimodality of galaxies, since no bimodality will be seen if the bulge-to-total fraction varies smoothly from 0 to 1. Rather, the galaxy bimodality will arise only if the bulge-to-total fraction is itself bimodal, with galaxies dominated by either their bulge or their disc. One possible explanation is that galaxies with a classical bulge have undergone a major merger whereas galaxies with no classical bulge have not (Drory & Fisher, 2007).

6.2 Measuring the evolution of the Schechter function

The properties and evolution of the galaxy population are conventionally quantified using simple functional (e.g., Schechter function) fits to the data. With a large sample of low-redshift galaxies, it is possible to test for expected bias when such a simple form is assumed for the luminosity function at high redshift.

Fig. 6.1 shows the Schechter function fits to K -band luminosity function of Section 5.2, restricting the fit to various portions of the luminosity function. This is intended to mimic surveys at higher redshift, where only the bright end of the luminosity function is visible. These results suggest that, even for a non-evolving luminosity function, with increasing redshift one would observe (1) a steeper faint-end slope, (2) a brighter characteristic magnitude and (3) a decreasing number density at the characteristic magnitude. This arises because of systematic deviations from the Schechter function form and highlights the danger of relying too strongly on the Schechter function fits.

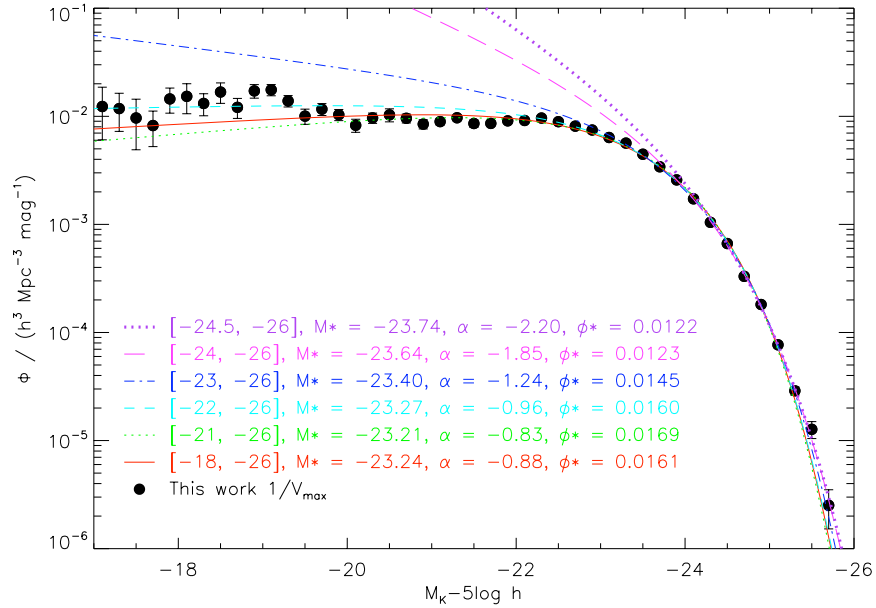


Figure 6.1: K -band luminosity function, showing Schechter function fits restricting the fitting range in absolute magnitudes, intended to mimic surveys at higher redshifts.

6.3 Possible improvements to UKIDSS data

The most significant difficulties encountered in this work have been related to the UKIDSS source extraction, particularly the large radius limit corresponding to a 24 arcsec diameter aperture. This is likely to be a matter of optimizing the parameters of the source extractor to suit a particular purpose: with ground-based deep near-infrared imaging, the sky is so bright that there can be no ‘one size fits all’ source extractor. If the software is well suited to point sources, then it is unlikely to be ideally suited to extended sources. Most of the applications of UKIDSS are for compact sources or point sources, such as stars in the Galactic Plane Survey (GPS) and Galactic Clusters Survey (GCS), brown dwarfs and quasars in the Large Area Survey (LAS) and distant galaxies in the Deep eXtragalactic Survey (DXS) and Ultra-Deep Survey (UDS).

For these reasons, the best course of action seems to be to generate an extended source catalogue using a different source extraction pipeline. Others are taking this approach. For example, [La Barbera et al. \(2008\)](#) have investigated the SDSS-UKIDSS Fundamental Plane of elliptical galaxies using their own source extractor on UKIDSS images at the locations of known SDSS galaxies.

A more suitable UKIDSS extractor for extended sources would have the following characteristics.

1. The sky subtraction would be suitable for large extended objects. [Dye et al. \(2006\)](#) explain

that the UKIDSS source extractor traces variations in the sky brightness on a scale of 25.6 arcsec, which ‘represents a compromise between being small enough to follow rapidly varying background and large enough to sample sufficient sky pixels outside objects’. However, in order to extract properties of galaxies larger than 25.6 arcsec, a different compromise is needed, with a larger scale and consequently a less precise determination of the sky level at a particular location. Ideally, the pixel scale for the sky subtraction should be adjusted according to the size of each galaxy.

2. Elliptical apertures would be used in addition to circular apertures. These allow a more precise determination of the sky background contribution to the measured flux of a galaxy.
3. Sérsic (1968) profile fits would be performed, providing structural information about the galaxies as well as an estimate of the total galaxy magnitude, which would improve estimates of the bright end of the luminosity function (Section 5.2). Other model fits to the galaxy brightness profile would also be performed, such as bulge-to-disc decomposition (Allen et al., 2006; Driver et al., 2007a,b).
4. Galaxy colours would be made available. In order to find the colour of an extended source, it is necessary to have the same aperture in both wavebands, or the same parameters for the model fit (e.g. Sérsic). Currently the UKIDSS Petrosian and Kron magnitudes do not use the same apertures between the different bands, while the Sérsic magnitudes are not available. This means that the fixed aperture magnitudes must be used to determine the galaxy colours, which is problematic since different regions of the galaxy will be included in the aperture depending on the angular size or aperture size. As well as NIR colours, which could reflect various physical properties of the galaxies (Eminian et al., 2008), good optical–near-infrared colours should be made available, in conjunction with the SDSS images or catalogues. This would lead to better template fits for the galaxies, resulting in better K -corrections, mass-to-light ratios and photometric redshifts.

All of these features were implemented for the 2MASS eXtended Source Catalogue (XSC; Jarrett et al., 2000): galaxies as large as 150 arcsec are fully imaged in 2MASS (the limit in UKIDSS is 24 arcsec), elliptical apertures are used, extrapolated (Sérsic) total magnitudes are provided (Cole et al., 2001) and accurate colours are made available by using consistent (‘fiducial’) apertures over the three JHK bands.

6.4 Possible improvements to methods

In this section I consider some alternative approaches that could be used to estimate the space density of galaxies.

6.4.1 Selection of the sample

The sample was selected through the UKIDSS-SDSS cross-matched table provided on the WSA. An alternative method would be to find all the target SDSS main galaxies within the LAS field and search for matches to those galaxies within the LAS. This would make clear the instances where an SDSS galaxy has not been detected at all in the LAS, giving a clearer understanding of the completeness of the matched sample.

6.4.2 Peculiar motions of nearby galaxies

Peculiar motions have not been considered in this work. However, in order to probe fainter absolute magnitudes, it is necessary to reduce the low-redshift limit. But peculiar velocities can no longer be neglected at very low redshift ($z < 0.01$). This was discussed in Section 3.5.4.

6.4.3 Magnitude errors

In this work the uncertainties in the measured quantities have been largely ignored and assumed to be negligible. However, the presence of magnitude errors can have a significant effect on the shape of the luminosity function, particularly at the bright end, where the luminosity function is steepest.

The effect of magnitude errors can be modelled for simple cases. In a magnitude-limited sample, assuming Poisson-type errors in flux, $\Delta f = \alpha\sqrt{f}$, Equation (3.11) can be rearranged to find α given a certain magnitude error, Δm , at the magnitude limit, m_{lim} :

$$\alpha = \frac{1 - 10^{(-0.4\Delta m)}}{10^{0.2m_{\text{lim}}}}. \quad (6.1)$$

Fig. 6.2 shows estimates of the luminosity function from a theoretical magnitude-limited sample at three discrete values of the redshift, with appropriate Poisson-like errors incorporated. It can be seen that, depending on the properties of the particular survey, magnitude errors may have a significant effect on the bright end of the luminosity function and on estimates of M^* .

One method of incorporating magnitude errors into the $1/V_{\text{max}}$ estimator would be as follows. Rather than dividing the galaxies into bins and summing their values of $1/V_{\text{max}}$, each individual galaxy could be represented as a Gaussian, according to the uncertainty in that galaxy's magnitude, such that the contribution of that galaxy to the luminosity function is $1/V_{\text{max}}$. With a sufficiently

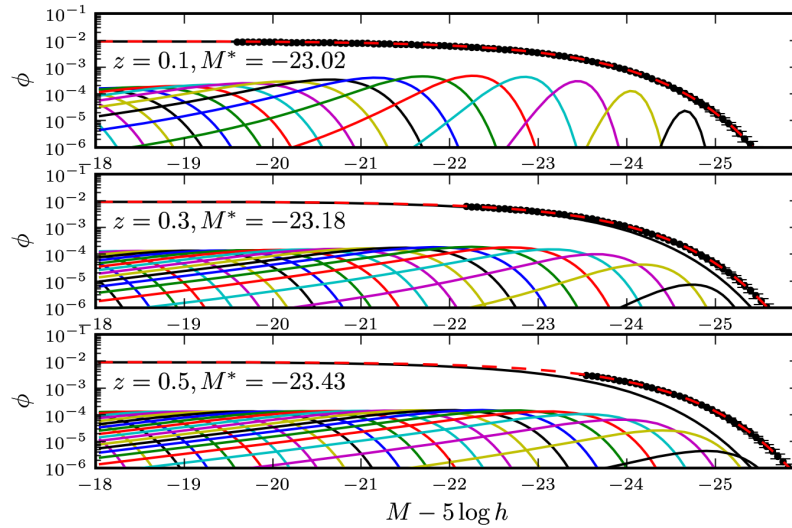


Figure 6.2: Theoretical predictions for observations of a Schechter luminosity function in the presence of magnitude errors. The original Schechter function has parameters $M^* - 5 \log h = -23$, $\alpha = -1$ and $\phi^* = 10^{-2}$ (black solid line). ϕ has units $h^3 \text{Mpc}^{-3} \text{mag}^{-1}$. Galaxies are observed in a magnitude-limited survey with $m < 18$ at three discrete redshifts, $z = 0.1$ (top), 0.3 (middle) and 0.5 (bottom). Each point on the original Schechter function is convolved with a Gaussian in flux appropriate to Poisson-type errors, with an uncertainty of 0.3 mag at the magnitude limit, $m_{\text{lim}} = 18$, as described in the text. The ‘observed’ luminosity function is constructed by summing these Gaussians. There is a large number of Gaussians; every tenth is shown in the plots as a coloured curve. The volume of the survey has not been included in the modelling. For galaxies that would be brighter than the magnitude limit at that redshift, the ‘observed’ luminosity function is shown (black dots), with errors chosen to be proportional to $\sqrt{\phi}$. Recovered best-fitting Schechter functions are shown (red dashed curves), with the values of M^* found to be -23.02 ($z = 0.1$), -23.18 ($z = 0.3$) and -23.43 ($z = 0.5$).

large sample, this will produce a smooth luminosity function. However, this would not resolve the bias at the bright end. To do so would require some kind of deconvolution of the luminosity function (e.g., Richardson–Lucy deconvolution, [Lucy, 1974](#)). However, my attempts at deconvolving the luminosity function have thus far not produced useful results.

A better method of including magnitude errors is used by [Blanton et al. \(2003b\)](#), who include magnitude errors in their version of the stepwise maximum likelihood (SWML) method. At each stage of the iteration, and for each galaxy, the current estimate for the luminosity function is convolved with a Gaussian in magnitude to find the likelihood that the particular galaxy was drawn from that luminosity function. Note that they used a Gaussian in magnitude, compared to a Gaussian in flux in [Fig. 6.2](#), which would give slightly different results.

These methods have not yet been used for a multivariate analysis, incorporating errors in size and surface brightness as well as errors in magnitude, with appropriate covariances.

6.4.4 Compensating for incompleteness

The bivariate brightness distribution in [Section 5.1](#) was expected to be incomplete at low surface brightness. There are various methods of compensating for such incompleteness when marginalizing the BBD to obtain the luminosity function. If sufficient information is available about completeness as a function of surface brightness, then an appropriate correction can be made to the LF ([Blanton et al., 2005b](#)). Alternatively, the assumption could be made that the distribution in surface brightness is Gaussian for fixed values of absolute magnitude. A Gaussian can then be fit to the regions of high completeness ([Driver et al., 2005](#)). The LF may then be found by extrapolating the Gaussians to high and low surface brightness. However, [Driver et al. \(2005\)](#) have shown that a Gaussian provides a poor fit at intermediate absolute magnitudes, so this method would be likely to return biased results.

6.4.5 Compensating for large-scale structure

The $1/V_{\max}$ estimator used in this thesis is sensitive to changes in the galaxy density with redshift. With a larger sample this will be less apparent, but other methods are less sensitive to large-scale structure, e.g., a multivariate version of the SWML method could be used, or a correction for variation with redshift can be applied to the $1/V_{\max}$ method ([Cross et al., 2001](#)). I have found the SWML method difficult to implement over four dimensions, due to the complexity of the parameter space, but these problems may not be insurmountable.

An additional source of uncertainty comes from the method used to divide the sample for jackknife error estimation. As discussed in [Section 5.4](#), it would be better to divide the sample

into regions of equal area, rather than regions of (approximately) equal numbers of galaxies.

6.5 Future surveys

In closing, I now consider how future surveys may help in building a census of K -band galaxies.

This work has used DR3 from UKIDSS and DR5 from SDSS, with the effective area (563 deg²) limited partly by those regions in SDSS that had spectroscopic coverage in DR5. The full LAS is anticipated to cover 4000 deg², all within the SDSS field. This would give a final sample around seven times as large as that used here. As well as reducing the statistical errors on the results and extending the range in absolute magnitude, this would open up various possibilities for subdividing the sample.

The LAS may be combined with other redshift and imaging surveys, such as the Millennium Galaxy Catalogue (MGC) survey (Liske et al., 2003), a 37.5 deg² B -band imaging survey, with redshifts for the 10 095 $B < 20$ galaxies. The MGC lies within the LAS area, and its greater depth compared with SDSS could overcome some of the problems experienced as a result of the SDSS r -band limit.

Another approach is to supplement SDSS (or other) redshifts with spectroscopy of galaxies within some K -band completeness limit. For example, redshifts from the recently-started Galaxy And Mass Assembly (GAMA)¹ spectroscopic survey could be used to probe regions of the parameter space not currently sampled.

Other imaging surveys have been proposed with UKIRT and with VISTA, the Visible and Infrared Survey Telescope for Astronomy, currently being commissioned at Paranal Observatory in Chile.² These include the UKIRT Hemisphere Survey (UHS)³ and the VISTA Hemisphere Survey (VHS),⁴ which would combine to give an all-sky NIR imaging survey in J and K .

Proposed surveys such as the VISTA Kilo-degree INfrared Galaxy survey (VIKING)⁵ would cover some 1500 deg² to a depth approximately 1.4 mag deeper than the LAS, providing measurements to a fainter surface brightness limit than is possible with the UKIDSS LAS.

It is hoped that these and other surveys, combined with new techniques for analysing the data, will provide insights into the processes that shaped the current galaxy population.

¹<http://www.eso.org/~jliske/gama/>

²<http://www.vista.ac.uk/>

³<http://wiki.astrogrid.org/pub/UKIDSS/SurveyCall/UHS-proposal-2006Nov-V3.pdf>

⁴<http://www.ast.cam.ac.uk/~rgm/vhs/>

⁵<http://www.ast.cam.ac.uk/~rgm/vista/vhs/viking/viking10.pdf>

Bibliography

- Abazajian K., et al. (2003). The First Data Release of the Sloan Digital Sky Survey. *AJ*, 126:2081–2086. [26](#)
- Adelman-McCarthy J. K., et al. (2007). The Fifth Data Release of the Sloan Digital Sky Survey. *ApJS*, 172:634–644, arXiv:0707.3380. [19](#)
- Albrecht A., Steinhardt P. J. (1982). Cosmology for grand unified theories with radiatively induced symmetry breaking. *Phys. Rev. Lett.*, 48:1220–1223. [2](#)
- Allen P. D., Driver S. P., Graham A. W., Cameron E., Liske J., de Propris R. (2006). The Millennium Galaxy Catalogue: bulge-disc decomposition of 10095 nearby galaxies. *MNRAS*, 371:2–18, arXiv:astro-ph/0605699. [128](#)
- Andreon S. (2002). Dim galaxies and outer halos of galaxies missed by 2MASS? The near-infrared luminosity function and density. *A&A*, 382:495–502, arXiv:astro-ph/0111207. [15](#), [30](#)
- Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T. (2006). Galaxy bimodality versus stellar mass and environment. *MNRAS*, 373:469–483. [126](#)
- Baldry I. K., Glazebrook K., Driver S. P. (2008). On the galaxy stellar mass function, the mass-metallicity relation and the implied baryonic mass function. *MNRAS*, 388:945–959, arXiv:0804.2892. [82](#), [89](#), [122](#), [124](#)
- Ball N. M., Loveday J., Brunner R. J., Baldry I. K., Brinkmann J. (2006). Bivariate galaxy luminosity functions in the Sloan Digital Sky Survey. *MNRAS*, 373:845–868, arXiv:astro-ph/0507547. [7](#), [18](#), [32](#), [87](#), [90](#), [115](#), [118](#), [126](#)
- Balogh M., et al. (2004). Galaxy ecology: groups and low-density environments in the SDSS and 2dFGRS. *MNRAS*, 348:1355–1372, astro-ph/0311379. [7](#)
- Bamford S. P., et al. (2008). Galaxy Zoo: the independence of morphology and colour. *ArXiv e-prints*, arXiv:0805.2612. [7](#), [126](#)

- Bell E. F., de Jong R. S. (2001). Stellar Mass-to-Light Ratios and the Tully-Fisher Relation. *ApJ*, 550:212–229, arXiv:astro-ph/0011493. [13](#), [120](#)
- Bell E. F., Baugh C. M., Cole S., Frenk C. S., Lacey C. G. (2003a). The properties of spiral galaxies: confronting hierarchical galaxy formation models with observations. *MNRAS*, 343:367–384, arXiv:astro-ph/0303531. [125](#)
- Bell E. F., McIntosh D. H., Katz N., Weinberg M. D. (2003b). The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. *ApJS*, 149:289–312, arXiv:astro-ph/0302543. [15](#), [16](#), [73](#), [74](#), [92](#), [107](#), [109](#), [122](#)
- Bell E. F., et al. (2004). Nearly 5000 Distant Early-Type Galaxies in COMBO-17: A Red Sequence and Its Evolution since $z \sim 1$. *ApJ*, 608:752–767, arXiv:astro-ph/0303394. [7](#), [10](#)
- Bell E. F., et al. (2006a). Dry Mergers in GEMS: The Dynamical Evolution of Massive Early-Type Galaxies. *ApJ*, 640:241–251, arXiv:astro-ph/0506425. [7](#), [10](#), [11](#)
- Bell E. F., Phleps S., Somerville R. S., Wolf C., Borch A., Meisenheimer K. (2006b). The Merger Rate of Massive Galaxies. *ApJ*, 652:270–276, arXiv:astro-ph/0602038. [7](#), [11](#)
- Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S. (2003). What Shapes the Luminosity Function of Galaxies? *ApJ*, 599:38–49, arXiv:astro-ph/0302450. [7](#), [77](#)
- Bertin E., Arnouts S. (1996). SExtractor: Software for source extraction. *A&AS*, 117:393–404. [30](#)
- Bertone S., De Lucia G., Thomas P. A. (2007). The recycling of gas and metals in galaxy formation: predictions of a dynamical feedback model. *MNRAS*, 379:1143–1154, arXiv:astro-ph/0701407. [8](#), [14](#)
- Binggeli B., Sandage A., Tammann G. A. (1988). The luminosity function of galaxies. *ARA&A*, 26:509–560. [7](#), [14](#)
- Blanton M. R., et al. (2001). The Luminosity Function of Galaxies in SDSS Commissioning Data. *AJ*, 121:2358–2380. [17](#), [18](#), [29](#), [31](#), [60](#), [62](#), [63](#), [82](#), [87](#), [106](#), [109](#), [118](#)
- Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J. (2003a). An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey “Tiling” Algorithm. *AJ*, 125:2276–2286, arXiv:astro-ph/0105535. [32](#), [68](#)

- Blanton M. R., et al. (2003b). The Galaxy Luminosity Function and Luminosity Density at Redshift $z = 0.1$. *ApJ*, 592:819–838, arXiv:astro-ph/0210215. [62](#), [73](#), [78](#), [86](#), [87](#), [88](#), [111](#), [112](#), [131](#)
- Blanton M. R., et al. (2003c). The Broadband Optical Properties of Galaxies with Redshifts $0.02 < z < 0.22$. *ApJ*, 594:186–207, arXiv:astro-ph/0209479. [18](#)
- Blanton M. R., et al. (2005a). New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *AJ*, 129:2562–2578, arXiv:astro-ph/0410166. [30](#), [37](#), [63](#)
- Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J. (2005b). The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. *ApJ*, 631:208–230, arXiv:astro-ph/0410164. [68](#), [82](#), [83](#), [131](#)
- Blanton M. R., Roweis S. (2007). K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared. *AJ*, 133:734–754, arXiv:astro-ph/0606170. [28](#), [39](#), [72](#), [112](#), [120](#)
- Bonifacio P., Monai S., Beers T. C. (2000). A Search for Stars of Very Low Metal Abundance. V. Photoelectric UBV Photometry of Metal-weak Candidates from the Northern HK Survey. *AJ*, 120:2065–2081, arXiv:astro-ph/0006178. [33](#)
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G. (2006). Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, arXiv:astro-ph/0511338. [9](#), [14](#)
- Bromm V., Larson R. B. (2004). The First Stars. *ARA&A*, 42:79–118, arXiv:astro-ph/0311019. [4](#)
- Bruzual G., Charlot S. (1993). Spectral evolution of stellar populations using isochrone synthesis. *ApJ*, 405:538–553. [73](#)
- Bruzual G., Charlot S. (2003). Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028, arXiv:astro-ph/0309134. [77](#), [120](#)
- Bruzual G. (2007). On TP-AGB stars and the mass of galaxies. In Vazdekis A., Peletier R. F., editors, *Stellar Populations as Building Blocks of Galaxies*, volume 241 of *IAU Symposium*, pages 125–132. arXiv:astro-ph/0703052. [123](#)
- Bundy K., Treu T., Ellis R. S. (2007). The Mass Assembly History of Spheroidal Galaxies: Did Newly Formed Systems Arise via Major Mergers? *ApJ*, 665:L5–L8, arXiv:0705.1007. [7](#), [11](#)

- Cameron E., Driver S. P. (2007). The galaxy luminosity-size relation and selection biases in the Hubble Ultra Deep Field. *MNRAS*, 377:523–534, arXiv:astro-ph/0701419. [18](#)
- Casali M., et al. (2007). The UKIRT wide-field camera. *A&A*, 467:777–784. [19](#), [23](#)
- Chabrier G. (2003). Galactic Stellar and Substellar Initial Mass Function. *PASP*, 115:763–795, arXiv:astro-ph/0304382. [120](#)
- Chołowiecki J. (1985). Bivariate luminosity function of E and S0 galaxies. *MNRAS*, 214:197–202. [16](#), [17](#), [83](#)
- Cirasuolo M., et al. (2007). The evolution of the near-infrared galaxy luminosity function and colour bimodality up to $z \simeq 2$ from the UKIDSS Ultra Deep Survey Early Data Release. *MNRAS*, 380:585–595, arXiv:astro-ph/0609287. [12](#), [28](#)
- Cole S., et al. (2001). The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions. *MNRAS*, 326:255–273, arXiv:astro-ph/0012429. [15](#), [16](#), [128](#)
- Cole S., Helly J., Frenk C. S., Parkinson H. (2008). The statistical properties of Λ cold dark matter halo formation. *MNRAS*, 383:546–556, arXiv:0708.1376. [10](#)
- Connolly A. J., Szalay A. S., Bershadsky M. A., Kinney A. L., Calzetti D. (1995). Spectral Classification of Galaxies: an Orthogonal Approach. *AJ*, 110:1071–1082, arXiv:astro-ph/9411044. [118](#)
- Connolly A. J., Szalay A. S. (1999). A Robust Classification of Galaxy Spectra: Dealing with Noisy and Incomplete Data. *AJ*, 117:2052–2062, arXiv:astro-ph/9901300. [118](#)
- Cooray A., Milosavljević M. (2005). What is L_* ? Anatomy of the Galaxy Luminosity Function. *ApJ*, 627:L89–L92, arXiv:astro-ph/0504580. [77](#)
- Cowie L. L., Songaila A., Hu E. M., Cohen J. G. (1996). New Insight on Galaxy Formation and Evolution From Keck Spectroscopy of the Hawaii Deep Fields. *AJ*, 112:839–864, arXiv:astro-ph/9606079. [4](#)
- Cross N., et al. (2001). The 2dF Galaxy Redshift Survey: the number and luminosity density of galaxies. *MNRAS*, 324:825–841, arXiv:astro-ph/0012165. [18](#), [89](#), [131](#)
- Cross N., Driver S. P. (2002). The bivariate brightness function of galaxies and a demonstration of the impact of surface brightness selection effects on luminosity function estimations. *MNRAS*, 329:579–587, arXiv:astro-ph/0110402. [18](#), [106](#)

- Croton D. J., et al. (2006). The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365:11–28, arXiv:astro-ph/0508046. [5](#), [8](#), [9](#), [14](#)
- Dalcanton J. J., Spergel D. N., Summers F. J. (1997). The Formation of Disk Galaxies. *ApJ*, 482:659–676, arXiv:astro-ph/9611226. [125](#)
- Davis M., Huchra J. (1982). A survey of galaxy redshifts. III - The density field and the induced gravity field. *ApJ*, 254:437–450. [87](#)
- de Jong R. S., Lacey C. (2000). The Local Space Density of SB-SDM Galaxies as Function of Their Scale Size, Surface Brightness, and Luminosity. *ApJ*, 545:781–797, arXiv:astro-ph/0008071. [17](#), [125](#)
- De Lucia G., Blaizot J. (2007). The hierarchical formation of the brightest cluster galaxies. *MNRAS*, 375:2–14, arXiv:astro-ph/0606519. [8](#), [14](#)
- de Vaucouleurs G. (1953). On the distribution of mass and luminosity in elliptical galaxies. *MNRAS*, 113:134–161. [27](#)
- Driver S. P., Phillipps S., Davies J. I., Morgan I., Disney M. J. (1994). Dwarf Galaxies at $Z=0.2$ - Photometry of the Cluster ABELL:963. *MNRAS*, 268:393–404. [83](#)
- Driver S. P. (1999). The Contribution of Normal, Dim, and Dwarf Galaxies to the Local Luminosity Density. *ApJ*, 526:L69–L72, arXiv:astro-ph/9909469. [17](#), [106](#)
- Driver S. P., Liske J., Cross N. J. G., De Propris R., Allen P. D. (2005). The Millennium Galaxy Catalogue: the space density and surface-brightness distribution(s) of galaxies. *MNRAS*, 360:81–103, arXiv:astro-ph/0503228. [18](#), [75](#), [83](#), [87](#), [88](#), [106](#), [125](#), [131](#)
- Driver S. P., et al. (2006). The Millennium Galaxy Catalogue: morphological classification and bimodality in the colour-concentration plane. *MNRAS*, 368:414–434, arXiv:astro-ph/0602240. [6](#), [7](#), [10](#), [18](#), [28](#), [115](#), [126](#)
- Driver S. P., Allen P. D., Liske J., Graham A. W. (2007a). The Millennium Galaxy Catalogue: The Luminosity Functions of Bulges and Disks and Their Implied Stellar Mass Densities. *ApJ*, 657:L85–L88. [126](#), [128](#)
- Driver S. P., Popescu C. C., Tuffs R. J., Liske J., Graham A. W., Allen P. D., de Propris R. (2007b). The Millennium Galaxy Catalogue: the B-band attenuation of bulge and disc light and

- the implied cosmic dust and stellar mass densities. *MNRAS*, 379:1022–1036, arXiv:0704.2140. [13](#), [33](#), [126](#), [128](#)
- Driver S. P., Popescu C. C., Tuffs R. J., Graham A. W., Liske J., Baldry I. (2008). The Energy Output of the Universe from 0.1 to 1000 μm . *ApJ*, 678:L101–L104, arXiv:0803.4164. [6](#), [33](#)
- Drory N., Fisher D. B. (2007). A Connection between Bulge Properties and the Bimodality of Galaxies. *ApJ*, 664:640–649, arXiv:0705.0973. [6](#), [10](#), [126](#)
- Dye S., et al. (2006). The UKIRT Infrared Deep Sky Survey Early Data Release. *MNRAS*, 372:1227–1252, arXiv:astro-ph/0603608. [13](#), [24](#), [26](#), [29](#), [30](#), [32](#), [37](#), [127](#)
- Efstathiou G., Ellis R. S., Peterson B. A. (1988). Analysis of a complete galaxy redshift survey. II - The field-galaxy luminosity function. *MNRAS*, 232:431–461. [14](#), [84](#), [86](#), [87](#)
- Eggen O. J., Lynden-Bell D., Sandage A. R. (1962). Evidence from the motions of old stars that the Galaxy collapsed. *ApJ*, 136:748–766. [5](#)
- Einstein A. (1915). Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (Berlin)*, pages 844–847. [2](#)
- Eke V. R., Baugh C. M., Cole S., Frenk C. S., King H. M., Peacock J. A. (2005). Where are the stars? *MNRAS*, 362:1233–1246, arXiv:astro-ph/0412049. [15](#), [16](#), [107](#), [109](#), [122](#)
- Eminian C., Kauffmann G., Charlot S., Wild V., Bruzual G., Rettura A., Loveday J. (2008). Physical interpretation of the near-infrared colours of low-redshift galaxies. *MNRAS*, 384:930–942, arXiv:0709.1147. [128](#)
- Fall S. M., Efstathiou G. (1980). Formation and rotation of disc galaxies with haloes. *MNRAS*, 193:189–206. [125](#)
- Ferrarese L., Merritt D. (2000). A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies. *ApJ*, 539:L9–L12, arXiv:astro-ph/0006053. [5](#)
- Freeman K. C. (1970). On the Disks of Spiral and S0 Galaxies. *ApJ*, 160:811–830. [125](#)
- Freeman K. C., Bland-Hawthorn J. (2002). The New Galaxy: Signatures of Its Formation. *ARA&A*, 40:487–537, arXiv:astro-ph/0208106. [5](#), [10](#)
- Friedmann A. (1922). Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10:377–386. [2](#)
- Friedmann A. (1924). Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21:326–332. [2](#)

- Frieman J., Turner M., Huterer D. (2008). Dark Energy and the Accelerating Universe. *ArXiv e-prints*, arXiv:0803.0982. [2](#)
- Gardner J. P., Sharples R. M., Frenk C. S., Carrasco B. E. (1997). A Wide-Field K-band Survey: The Luminosity Function of Galaxies. *ApJ*, 480:L99–L102, arXiv:astro-ph/9702178. [14](#), [16](#)
- Gebhardt K., et al. (2000). A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion. *ApJ*, 539:L13–L16, arXiv:astro-ph/0006289. [5](#)
- Gilmore G., Wilkinson M. I., Wyse R. F. G., Kleyna J. T., Koch A., Evans N. W., Grebel E. K. (2007). The Observed Properties of Dark Matter on Small Spatial Scales. *ApJ*, 663:948–959, arXiv:astro-ph/0703308. [10](#)
- Gilmore G. (2008). The Science of Galaxy Formation. *ArXiv e-prints*, arXiv:0808.3188. [10](#)
- Glazebrook K., Peacock J. A., Miller L., Collins C. A. (1995). An imaging K-band survey - II. The redshift survey and galaxy evolution in the infrared. *MNRAS*, 275:169–184, arXiv:astro-ph/9502094. [14](#), [16](#), [73](#)
- Gott J. R. I., Jurić M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J. (2005). A Map of the Universe. *ApJ*, 624:463–484, arXiv:astro-ph/0310571. [113](#)
- Graham A. W., Driver S. P. (2005). A Concise Reference to (Projected) Sérsic $R^{1/n}$ Quantities, Including Concentration, Profile Slopes, Petrosian Indices, and Kron Magnitudes. *Publ. Astron. Soc. Aust.*, 22:118–127, arXiv:astro-ph/0503176. [30](#)
- Graham A. W., Driver S. P., Petrosian V., Conselice C. J., Bershady M. A., Crawford S. M., Goto T. (2005). Total Galaxy Magnitudes and Effective Radii from Petrosian Magnitudes and Radii. *AJ*, 130:1535–1544, arXiv:astro-ph/0504287. [29](#), [60](#), [109](#)
- Gunn J. E., Gott J. R. I. (1972). On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, 176:1–19. [6](#)
- Guth A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356. [2](#)
- Hambly N. C., et al. (2008). The WFCAM Science Archive. *MNRAS*, 384:637–662, arXiv:0711.3593. [20](#)
- Heavens A., Panter B., Jimenez R., Dunlop J. (2004). The star-formation history of the Universe from the stellar populations of nearby galaxies. *Nat*, 428:625–627. [5](#)

- Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T. (2006). The UKIRT Infrared Deep Sky Survey ZY JHK photometric system: passbands and synthetic colours. *MNRAS*, 367:454–468, arXiv:astro-ph/0601592. [19](#), [21](#), [23](#), [28](#), [72](#), [91](#)
- Hogg D. W. (1999). Distance measures in cosmology. *ArXiv e-prints*, arXiv:astro-ph/9905116. [2](#)
- Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J. (2002). The K correction. *ArXiv e-prints*, arXiv:astro-ph/0210394. [72](#)
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D. (2008a). A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. I. Galaxy Mergers and Quasar Activity. *ApJS*, 175:356–389, arXiv:0706.1243. [9](#)
- Hopkins P. F., Cox T. J., Kereš D., Hernquist L. (2008b). A Cosmological Framework for the Co-evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. II. Formation of Red Ellipticals. *ApJS*, 175:390–422, arXiv:0706.1246. [9](#)
- Huang J.-S., Cowie L. L., Gardner J. P., Hu E. M., Songaila A., Wainscoat R. J. (1997). The Hawaii K-Band Galaxy Survey. II. Bright K-Band Imaging. *ApJ*, 476:12–21, arXiv:astro-ph/9610084. [79](#)
- Huang J.-S., Glazebrook K., Cowie L. L., Tinney C. (2003). The Hawaii+Anglo-Australian Observatory K-Band Galaxy Redshift Survey. I. The Local K-Band Luminosity Function. *ApJ*, 584:203–209, arXiv:astro-ph/0209440. [15](#), [16](#)
- Hubble E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173. [2](#)
- Impey C., Bothun G. (1997). Low Surface Brightness Galaxies. *ARA&A*, 35:267–307. [17](#)
- Irwin M. J. (1985). Automatic analysis of crowded fields. *MNRAS*, 214:575–604. [46](#)
- Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P. (2000). 2MASS Extended Source Catalog: Overview and Algorithms. *AJ*, 119:2498–2531, arXiv:astro-ph/0004318. [14](#), [85](#), [109](#), [128](#)
- Johnston R., Teodoro L., Hendry M. (2007). Completeness - I. Revisited, reviewed and revived. *MNRAS*, 376:1757–1766, arXiv:astro-ph/0703040. [59](#)
- Jones D. H., Peterson B. A., Colless M., Saunders W. (2006). Near-infrared and optical luminosity functions from the 6dF Galaxy Survey. *MNRAS*, 369:25–42, arXiv:astro-ph/0603609. [15](#), [16](#), [63](#), [80](#), [88](#), [107](#), [109](#), [111](#)

- Juneau S., et al. (2005). Cosmic Star Formation History and Its Dependence on Galaxy Stellar Mass. *ApJ*, 619:L135–L138, arXiv:astro-ph/0411775. [4](#)
- Kashlinsky A., Arendt R. G., Mather J., Moseley S. H. (2005). Tracing the first stars with fluctuations of the cosmic infrared background. *Nat*, 438:45–50, arXiv:astro-ph/0511105. [4](#)
- Kauffmann G., Haehnelt M. (2000). A unified model for the evolution of galaxies and quasars. *MNRAS*, 311:576–588, arXiv:astro-ph/9906493. [8](#), [9](#)
- Kauffmann G., et al. (2003). Stellar masses and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey. *MNRAS*, 341:33–53, arXiv:astro-ph/0204055. [34](#)
- Kennicutt Jr. R. C. (1998). Star Formation in Galaxies Along the Hubble Sequence. *ARA&A*, 36:189–232, arXiv:astro-ph/9807187. [5](#)
- Kirby E., Jerjen H., Ryder S., Driver S. (2008). Deep Near-IR Surface Photometry of 57 Galaxies in the Local Sphere of Influence. *ArXiv e-prints*, arXiv:0808.2529. [15](#)
- Kochanek C. S., et al. (2001). The K-Band Galaxy Luminosity Function. *ApJ*, 560:566–579, arXiv:astro-ph/0011456. [15](#), [16](#), [87](#), [107](#), [109](#)
- Kormendy J. (1977). Brightness distributions in compact and normal galaxies. II - Structure parameters of the spheroidal component. *ApJ*, 218:333–346. [125](#)
- Kormendy J., Kennicutt Jr. R. C. (2004). Secular Evolution and the Formation of Pseudobulges in Disk Galaxies. *ARA&A*, 42:603–683, arXiv:astro-ph/0407343. [6](#), [126](#)
- Kron R. G. (1980). Photometry of a complete sample of faint galaxies. *ApJS*, 43:305–325. [30](#)
- La Barbera F., Busarello G., Merluzzi P., de la Rosa I., Coppola G., Haines C. P. (2008). The SDSS-UKIDSS Fundamental Plane of Early-type Galaxies. *ArXiv e-prints*, arXiv:0807.3829. [127](#)
- Lawrence A., et al. (2007). The UKIRT Infrared Deep Sky Survey (UKIDSS). *MNRAS*, 379:1599–1617, arXiv:astro-ph/0604426. [19](#), [23](#), [24](#), [46](#)
- Lemaître G. (1931). Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *MNRAS*, 91:483–490. [2](#)
- Liddle A., Mukherjee P., Parkinson D. (2006). Model selection in cosmology. *Astronomy and Geophysics*, 47:4.30–4.33, arXiv:astro-ph/0608184. [78](#)

- Lin L., et al. (2008). The Redshift Evolution of Wet, Dry, and Mixed Galaxy Mergers from Close Galaxy Pairs in the DEEP2 Galaxy Redshift Survey. *ApJ*, 681:232–243, arXiv:0802.3004. [7](#), [11](#)
- Linde A. D. (1982). A new inflationary Universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Phys. Lett. B*, 108:389–393. [2](#)
- Linde A. D. (1983). Chaotic inflation. *Phys. Lett. B*, 129:177–181. [2](#)
- Liske J., Lemon D. J., Driver S. P., Cross N. J. G., Couch W. J. (2003). The Millennium Galaxy Catalogue: $16 \leq B_{\text{MGC}} < 24$ galaxy counts and the calibration of the local galaxy luminosity function. *MNRAS*, 344:307–324, arXiv:astro-ph/0207555. [132](#)
- Loeb A., Barkana R. (2001). The Reionization of the Universe by the First Stars and Quasars. *ARA&A*, 39:19–66, arXiv:astro-ph/0010467. [5](#)
- Loveday J. (1997). The Local Space Density of Dwarf Galaxies. *ApJ*, 489:29–36, arXiv:astro-ph/9703022. [83](#)
- Loveday J. (2000). The K-band luminosity function of nearby field galaxies. *MNRAS*, 312:557–566, arXiv:astro-ph/9907179. [14](#), [16](#), [17](#), [63](#), [73](#), [90](#)
- Lucy L. B. (1974). An iterative technique for the rectification of observed distributions. *AJ*, 79:745–754. [131](#)
- Lupton R. (1993). *Statistics in theory and practice*. Princeton, N.J.: Princeton University Press. [80](#)
- Magorrian J., et al. (1998). The Demography of Massive Dark Objects in Galaxy Centers. *AJ*, 115:2285–2305, arXiv:astro-ph/9708072. [5](#)
- Maller A. H., Berlind A. A., Blanton M. R., Hogg D. W. (2008). The Intrinsic Properties of SDSS Galaxies. *ArXiv e-prints*, arXiv:0801.3286. [13](#)
- Mannucci F., Basile F., Poggianti B. M., Cimatti A., Daddi E., Pozzetti L., Vanzi L. (2001). Near-infrared template spectra of normal galaxies: k-corrections, galaxy models and stellar populations. *MNRAS*, 326:745–758, arXiv:astro-ph/0104427. [13](#)
- Mo H. J., Mao S., White S. D. M. (1998). The formation of galactic discs. *MNRAS*, 295:319–336, arXiv:astro-ph/9707093. [125](#)

- Mobasher B., Sharples R. M., Ellis R. S. (1993). A Complete Galaxy Redshift Survey - Part Five - Infrared Luminosity Functions for Field Galaxies. *MNRAS*, 263:560–574. [14](#), [16](#)
- Panter B., Heavens A. F., Jimenez R. (2004). The mass function of the stellar component of galaxies in the Sloan Digital Sky Survey. *MNRAS*, 355:764–768, arXiv:astro-ph/0406299. [122](#)
- Peebles P. J. E. (1969). Origin of the Angular Momentum of Galaxies. *ApJ*, 155:393–401. [125](#)
- Peebles P. J. E. (2004). The Future of Astronomy and Physical Cosmology. *Publ. Astron. Soc. Aust.*, 21:385–389. [3](#)
- Peebles P. J. E. (2005). Open problems in cosmology. *Nuclear Physics B Proceedings Supplements*, 138:5–9. [3](#)
- Petrosian V. (1976). Surface brightness and evolution of galaxies. *ApJ*, 209:L1–L5. [29](#)
- Petrosian V. (1998). New and Old Tests of Cosmological Models and the Evolution of Galaxies. *ApJ*, 507:1–15, arXiv:astro-ph/9807189. [17](#)
- Phillipps S., Disney M. (1986). Selection effects in the bivariate brightness distribution for spiral galaxies. *MNRAS*, 221:1039–1048. [17](#), [125](#)
- Phillipps S., Davies J. I., Disney M. J. (1990). The visibility of high-redshift galaxies. *MNRAS*, 242:235–240. [17](#)
- Poggianti B. M., et al. (2006). The Evolution of the Star Formation Activity in Galaxies and Its Dependence on Environment. *ApJ*, 642:188–215, arXiv:astro-ph/0512391. [4](#)
- Richstone D., et al. (1998). Supermassive black holes and the evolution of galaxies. *Nat*, 395:A14–A19, arXiv:astro-ph/9810378. [5](#)
- Robertson H. P. (1935). Kinematics and World-Structure. *ApJ*, 82:284–301. [2](#)
- Rowan-Robinson M. (1969). Local Theory for Quasars. *Nat*, 224:1094. [14](#), [88](#)
- Sandage A., Tammann G. A., Yahil A. (1979). The velocity field of bright nearby galaxies. I - The variation of mean absolute magnitude with redshift for galaxies in a magnitude-limited sample. *ApJ*, 232:352–364. [14](#), [84](#), [85](#)
- Schechter P. (1976). An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306. [14](#), [80](#)

- Schlegel D. J., Finkbeiner D. P., Davis M. (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *ApJ*, 500:525–553, arXiv:astro-ph/9710327. [33](#)
- Schmidt M. (1968). Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *ApJ*, 151:393–409. [14](#), [84](#), [88](#), [89](#)
- Sérsic J. L. (1968). *Atlas de galaxias australes*. Cordoba, Argentina: Observatorio Astronomico. [27](#), [30](#), [60](#), [128](#)
- Shen Y., et al. (2007). Clustering of High-Redshift ($z \geq 2.9$) Quasars from the Sloan Digital Sky Survey. *AJ*, 133:2222–2241, arXiv:astro-ph/0702214. [5](#)
- Sheth R. K., Jimenez R., Panter B., Heavens A. F. (2006). Environment and the Cosmic Evolution of Star Formation. *ApJ*, 650:L25–L28, arXiv:astro-ph/0604581. [4](#)
- Smith A. J., Loveday J., Cross N. J. G. (2008). Luminosity and surface brightness distribution of K-band galaxies from the UKIDSS Large Area Survey. *ArXiv e-prints*, arXiv:0806.0343. [ii](#)
- Sodré L. J., Lahav O. (1993). The bivariate diameter-magnitude function of galaxies in the ESO-LV catalogue. *MNRAS*, 260:285–298. [17](#), [18](#)
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L. (2008). A Semi-Analytic Model for the Co-evolution of Galaxies, Black Holes, and Active Galactic Nuclei. *ArXiv e-prints*, arXiv:0808.1227. [125](#)
- Spergel D. N., et al. (2007). Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170:377–408, arXiv:astro-ph/0603449. [2](#), [3](#), [5](#)
- Springel V., et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nat*, 435:629–636, arXiv:astro-ph/0504097. [4](#), [8](#)
- Springel V., Frenk C. S., White S. D. M. (2006). The large-scale structure of the Universe. *Nat*, 440:1137–1144, arXiv:astro-ph/0604561. [4](#)
- Stoughton C., et al. (2002). Sloan Digital Sky Survey: Early Data Release. *AJ*, 123:485–548. [23](#), [28](#), [30](#)
- Strauss M. A., et al. (2002). Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *AJ*, 124:1810–1824, arXiv:astro-ph/0206225. [32](#), [62](#)

- Strigari L. E., Bullock J. S., Kaplinghat M., Simon J. D., Geha M., Willman B., Walker M. G. (2008). A common mass scale for satellite galaxies of the Milky Way. *Nat*, 454:1096–1097, arXiv:0808.3772. [10](#)
- Szokoly G. P., Subbarao M. U., Connolly A. J., Mobasher B. (1998). The Near-Infrared Number Counts and Luminosity Functions of Local Galaxies. *ApJ*, 492:452–460, arXiv:astro-ph/9801132. [14](#), [16](#)
- Takeuchi T. T., Yoshikawa K., Ishii T. T. (2000). Tests of Statistical Methods for Estimating Galaxy Luminosity Function and Applications to the Hubble Deep Field. *ApJS*, 129:1–31, arXiv:astro-ph/0003127. [84](#)
- Tegmark M., et al. (2004). Cosmological parameters from SDSS and WMAP. *Phys. Rev. D*, 69(10):103501, arXiv:astro-ph/0310723. [2](#), [3](#)
- Turner M. S., Tyson J. A. (1999). Cosmology at the millennium. *Reviews of Modern Physics Supplement*, 71:S145–S164, arXiv:astro-ph/9901113. [3](#)
- Waddington I., et al. (2002). Old elliptical galaxies at $z \simeq 1.5$ and the Kormendy relation. *MNRAS*, 336:1342–1350, arXiv:astro-ph/0207663. [125](#)
- Walker A. G. (1935). On the formal comparison of Milne’s kinematical system with the systems of general relativity. *MNRAS*, 95:263–269. [2](#)
- Wall J. V., Jenkins C. R. (2003). *Practical Statistics for Astronomers*. Princeton Series in Astrophysics. [88](#)
- Warren S. J., et al. (2007). The United Kingdom Infrared Telescope Infrared Deep Sky Survey First Data Release. *MNRAS*, 375:213–226, arXiv:astro-ph/0610191. [24](#)
- White S. D. M., Rees M. J. (1978). Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358. [5](#)
- White S. D. M., Frenk C. S. (1991). Galaxy formation through hierarchical clustering. *ApJ*, 379:52–79. [5](#), [8](#)
- Willick J. A., Strauss M. A., Dekel A., Kolatt T. (1997). Maximum Likelihood Comparisons of Tully-Fisher and Redshift Data: Constraints on Omega and Biasing. *ApJ*, 486:629–664, arXiv:astro-ph/9612240. [63](#)
- Willmer C. N. A. (1997). Estimating galaxy luminosity functions. *AJ*, 114:898–912, arXiv:astro-ph/9707246. [84](#), [85](#)

York D. G., et al. (2000). The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120:1579–1587, arXiv:astro-ph/0006396. [19](#), [22](#)