Analysis of the two degree field galaxy redshift survey

Norberg, Iohn Peder Ragnar

Norberg, Iohn Peder Ragnar (2001) *Analysis of the two degree field galaxy redshift survey*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/3859/

Analysis of The Two Degree Field Galaxy Redshift Survey

by Iohn Peder Ragnar Norberg

September 2001

Abstract

In this Thesis we analyse in detail the Anglo-Australian 2 degree Field Galaxy Redshift Survey (2dFGRS). The goal of this survey is to measure all galaxy redshifts for the 250 000 galaxies brighter than $b_J = 19.45$ spread over $\sim$ 2000 square degrees. At present, the 2dFGRS has obtained redshifts for $\sim$ 190 000 galaxies. It is currently the biggest galaxy redshift survey in existence and represents an order of magnitude increase in size over any previous survey.

The study of the large-scale structure of the Universe is undergoing a revolution due to important technological advances in observational astronomy that make surveys like the 2dFGRS possible. This new era in mapping the Universe demands the development of new theoretical analysis tools, both to exploit the large amounts of data, and to take advantage of, for the first time in cosmology, the extraordinary opportunity to push random errors below the level of the systematic errors.

By a detailed analysis of the survey construction and observing strategy, we implement a set of maps to estimate, as precisely as possible, the selection function of the 2dFGRS. These maps, which characterize the survey completeness, enable us to estimate accurately fundamental properties of a homogeneous galaxy sample: the galaxy luminosity function and real space galaxy clustering. By combining the 2dFGRS with the near infra-red 2MASS survey, we estimate the K_S-band galaxy luminosity function, from which we infer the stellar mass function of galaxies. This yields a total mass fraction in stars between 0.1% and 0.3% of the critical cosmic density. Exploiting the size of the survey, we undertake the first precise measurement of the dependence of galaxy clustering on luminosity and spectral type. Star-forming galaxies as well as more quiescent galaxies show a clear increase in clustering strength with luminosity at a similar rate. This is the first time that we are able to examine in detail the properties of galaxies that drive their spatial distribution.

Analysis
of
The Two Degree Field
Galaxy Redshift Survey

by Iohn Peder Ragnar Norberg

A thesis submitted to the University of Durham
in accordance with the regulations for
admittance to the Degree of Doctor of Philosophy.

Department of Physics
University of Durham
September 2001

Contents

List of Figures

List of Tables

Declaration

The work described in this thesis was undertaken between October 1998 and September 2001 whilst the author was a research student under the supervision of Prof. Carlos Frenk and Dr. Shaun Cole in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or at any other University.

Portions of this work have appeared in the papers listed below. These have been adapted to highlight the portions to which the author has contributed the most.

- Colless M., Dalton G., Maddox S., Sutherland W., Norberg P., Cole S., & the 2dFGRS Team, 2001, MNRAS, 328, 1039 (Chapter 2)
- Norberg P., Cole S., Baugh C. M., Frenk C. S., & the 2dFGRS Team, 2001, MNRAS, submitted (Chapter 3)
- Cole S., Norberg P., Baugh C. M., Frenk C. S., & the 2dFGRS Team, 2001, MNRAS, 326, 255 (Chapter 4)
- Peacock J. A., Cole S., Norberg P., & the 2dFGRS Team, 2001, Nature, **410**, 169–173 (Chapter 5)
- Norberg P., Baugh C. M., Hawkins E., Maddox S. J., Peacock J. A., Cole S., Frenk C. S., & the 2dFGRS Team, 2001, MNRAS, 328, 64 (Chapter 6)
- Norberg P., Baugh C. M., Cole S., Frenk C. S., & the 2dFGRS Team, 2001, MNRAS, in press (Chapter 7)

Prologue

As major parts of this Thesis have already been written as separate papers, as indicated in the declaration, the first chapter of this Thesis is not a fully self-contained introduction to the whole Thesis. Instead, each chapter contains a separate introduction whose purpose is to explain the relevance of the work presented in that chapter. The aim of the introduction presented in the first chapter is to set up the general context of the 2 degree Field Galaxy Redshift Survey (2dFGRS).

Acknowledgments*

It is finally here that I have got a chance to acknowledge all those who have been involved, on all kind of levels, in this Thesis.

I start by expressing my biggest thanks to my supervisors Carlos Frenk and Shaun Cole. None of the work presented here would have been possible without their continuous support, enthusiasm toward research and ability of always seeing the best out of the situation, even when research shows its more darker sides. I have to thank them for their involvement in helping me to successfully obtain an Overseas Research Award and a TMR Grant from the Swiss National Science Foundation. I would like to stress my gratitude to these funding bodies for having given me the opportunity to realize one of my childhood's dreams: carry out research in cosmology.

A very special thank to Carlton Baugh, with whom I have been working rather substantially during the second half of my PhD. Many things would not have been the same if it was not for Carlton. His involvement in all the aspects of my research are without any measure. Thanks also for introducing me to five-a-side football, but also to squash... Great moments with, finally, very intense matches!

I would also like to thank my supervisors for having given me the opportunity to do research on a particularly interesting subject, the 2dF Galaxy Redshift Survey (2dFGRS). This survey has taken up a substantial fraction of my research time, but in the end it was really worth it! I take the chance here to thank the whole 2dFGRS Team, with a special mention to Matthew Colless and John Peacock, who have supported me rather substantially during this last year, which coincided with the famous '100K-release'. I would definitively say that the experience of a big collaboration is really worth the effort, even if it has to be combined with a certain amount of politics.

Last but not least on the working side, I would like to thank Alan Lotts for all his help with various computer issues, but also Nigel Metcalfe, for his recent involvements in trying to understand Himalia.

---

[*] This is probably the only text of the entire thesis Carlton has not had the opportunity to 'anglicise'... Slow-Poke would like to thank him here once again for his efforts in improving my English, and hopefully they are not in vain...

So after having thanked all those who have directly supervised my work, I must thank all of you who made the time here in Durham so enjoyable. First, all the office mates which have had to put up with me (and for sure they really know that it is not always an easy task): Chris Done and Piotr Zycki (who probably do not even remember me sharing an office with them), Steve Hatton (for our peculiar train journey discussion), Mikey Beasley (for being my first tennis partner), Scott Kay (for both his amazing ability to teach me the spoken English language and his nice will to make me taste beer), Andrew Benson (for simply being Andrew, the beer and curry lover), Cesario Lia (for being so Italian), Steve Quinney (for both his guidance through the mystery of the written English language and for his wonderful calculator) and Chris Power (for his natural all-round knowledge and friendship, which both give rise to so many fabulous discussions). A special mention to Marc Vallbe, with whom I have never shared an office, for his continuous support in my attempts to practice my singing. But obviously office mates are not the only ones who contribute to the young-trendy-astronomy life in the department, so I have to mention here: Fiona Hoyle (for me known as 'Miss Organizer' and creator of the tennis tournament I liked so much!), Eric Bell, Dave Gilbank, Kevin Pimbblet, Harold Dyson, James Turner, Geoff Busswell, Amaya Gaztelu, Richard McDermid and Graham Smith, for only citing a few of them.

Then a few words for those who have indirectly helped me with this Thesis, but without being much around in Durham: my family. They know how much their attention has been valued, so no point in over doing it here. Just a 'stort tack till er alla' for supporting me over the years, especially when it comes down to all those 'det är inte rätt!'...

And finally, thanks Joy for being here with me.

Chapter 1

Introduction to Galaxy Redshift Surveys and the 2dFGRS

The aim of undertaking the 2dF Galaxy Redshift Survey (2dFGRS) is to advance the state of the art for galaxy redshift surveys by an order of magnitude, with the goal of improving our understanding of large-scale structure, galaxy formation and cosmology. The aim of this chapter is to introduce the subject of galaxy redshift surveys. We[*] devote the first section to a brief introduction to the general concept, context and content of galaxy redshift surveys[†], followed by a succinct description of previous important galaxy redshift surveys, and ending by briefly comparing the two current major galaxy redshift surveys of the local Universe. In section 1.2, we introduce in more detail the 2dFGRS, which is the focus of our analysis in the subsequent chapters. In section 1.3, we summarize the current status of the survey, with a review of the progress of the survey with time. Finally, we end this chapter with a brief summary of the aims of this Thesis.

1.1 Introduction

The fundamental motivation behind every research project is the same regardless of the field of investigation: to increase our knowledge and to develop our understanding. Cosmology, the study of the origin of the Universe, is no different from any other scientific field in this respect. Galaxy redshift surveys are a part of our attempt to increase our general understanding of how the Universe evolves and functions. The primary goal of these surveys is to obtain a three dimensional picture of our neighbourhood in the Universe.

The desire to map our neighbourhood is possibly one of the most ancient problems

---

[*] Throughout this Thesis, I will use the conventional 'we' to refer to the first person as well as to the 2dFGRS Team, of which I am member.

[†] The aim of this section is to introduce galaxy redshift surveys in a form suitable for any reader of this Thesis. For a more scientific introduction, we refer the reader directly to §1.1.1.

tackled by humans. Long before the rules of cartography were laid down, human beings had always shown a need to identify the locations of objects with respect to each other. With time however, the aims behind constructing maps have evolved, from, for example, knowing where to find prosperous fields to where other civilisations are located. As the techniques of cartography have evolved with time, the fundamental aims have also changed, and the outcome from this map making process has evolved. Creating maps of the neighbourhood not only allows us to derive spatial information about the surroundings, but also vital information about its properties, like hydrology maps which show the region over which a river has its influence, and hence are useful, for example, during floods. Galaxy redshift surveys contain much more information than just the positions of galaxies in the universe, as we explain in the following paragraphs.

If we assume that we know the 'exact'[‡] positions of a large number of galaxies, then this information can be used in several ways.

The first and probably the most obvious use to most people, is to create a 3-D map of the spatial distribution of galaxies. With such a map, we can see how galaxies are grouped together. Indeed, the first striking conclusion anyone would have from a 3-D map is that the distribution of galaxies is clustered. However, the interpretation of a map of the local universe is not as straightforward as one may naively think. One major problem for astronomy, and cosmology in particular, is that we can only measure physical parameters of those objects that we can see. If for some reason some galaxies are not seen, then we will not be able to measure their redshift[§], and hence we will not know their position in the universe. It is extremely important to distinguish between not being able to measure something from not knowing that it is there in the first place. A large fraction of this thesis is dedicated to understanding, quantifying and correcting for objects that we have not been able to obtain a redshift for, but which we know are physically there. From time to time (or more exactly whenever it is possible), we also address the question of what fraction of galaxies we have missed, which is undeniably a very difficult question to fully assess. Hence, in order to construct a faithful 3-D map it is essential to know the probability of finding a galaxy in the survey, which we call the spatial selection function. This function can most of the time be divided into two components, the angular

---

[‡] Exact is in quotes, as we want to emphasise the problem of measuring distance in cosmology: nothing is exact, it can only be known to the precision of our current interpretation of the measurement, as explained later on in this paragraph.

[§] A galaxy's redshift is, in first approximation, a measure of its distance away from us. It is the ratio of its recession velocity to the speed of light. See appendix A for a more precise definition of redshift.

selection function and the radial selection function. Quantifying the selection function is essential for many projects. A detailed understanding of the survey selection function is essential for the interpretation of a 3-D map of the Universe. In a magnitude limited survey, consisting of galaxies brighter than a specified flux limit, the radial density of galaxies varies strongly, as it would do even in the absence of galaxy clustering. This is due to the fact that more luminous galaxies can be seen over a wider range of redshift than the fainter ones. For that reason we devote two entire chapters of this thesis to quantifying the selection function of the 2dFGRS (chapter 2 for the angular selection function and chapter 3 for the radial selection function). Once the selection function is known, various studies can be done on the galaxy sample. In this Thesis we have looked at galaxy clustering in particular, both with and without the peculiar motion of galaxies taken into account (see chapter 5 and chapters 6 and 7 respectively). The measure of the position of a galaxy is inferred from its recession velocity, which is not only due to the Hubble flow, arising from the expansion of the Universe, but contains also a component dubbed 'peculiar motion', generated by local inhomogeneities in the density distribution.

As well as looking at the three dimensional positions of galaxies, we can directly use the redshift of each galaxy to get an estimate of its intrinsic luminosity. Indeed, if we know both the distance at which a galaxy is located and its apparent magnitude (ie. a direct measure of the flux received by the observer), it is possible to calculate the galaxy's intrinsic luminosity[¶]. Perhaps the most basic and fundamental description of the local universe is to count the number of galaxies per unit volume as function of their luminosity, summarized as the galaxy luminosity function, which is therefore an extremely important constraint on models of galaxy formation. This motivates us to study in great detail, in chapter 3, the galaxy luminosity function.

Another source of information in galaxy redshift surveys is the galaxy spectrum, from which the redshift is measured. The form of a galaxy spectrum is mainly influenced by the chemical composition and the star formation history of the galaxy. The flux of a galaxy is always measured through a filter, which is only sensitive to a certain range of wavelengths, hence fluxes measured in different passbands are expected to be sensitive to different physical properties. The filter in which 2dFGRS galaxies are selected is sensitive to recent star formation, whereas the filters used in the Two Micron All Sky Survey

---

[¶] In this short introduction, we neglect the complications arising from the k-correction, ie. from band shifting due to the expansion of the Universe. This is obviously taken into account in our analysis in the forthcoming chapters.

(2MASS) are more sensitive to the assembled stellar mass. Therefore, we combine, in chapter 4, the optically selected 2dFGRS with the near infra-red selected 2MASS. This allows us to measure the galaxy luminosity function in two other passbands, but also to look at the galaxy stellar mass function, which is the mass of stars per unit volume as function of luminosity.

As mentioned earlier, the shape and the form of the galaxy spectrum tells us something about the properties of the galaxy. A purely statistical way of classifying galaxies with spectra is to optimally reduce each spectra to a continuous set of parameters. From this spectral classification scheme, samples with different spectral characteristics can be compared. In chapter 7, we use this technique to look at the clustering properties of star-forming galaxies compared to those of more quiescent galaxies.

This brief overview is a simple, but nevertheless accurate, picture of what we hope to achieve in this Thesis. In the sections that follow, we discuss galaxy redshift surveys in a more scientific way, mentioning achievements and limitations of previous surveys, with a special focus on the context of the two major galaxy redshift surveys of the local universe, the 2dFGRS and Sloan Digital Sky Survey (SDSS).

1.1.1 Galaxy redshift surveys in the XX$^{\text{th}}$ century

The first galaxy redshifts to appear in the literature are probably those measured and interpreted by Hubble (1929a)[‖], and denote the start of the compilation of the first real galaxy redshift catalogue. Even with further pioneering work from Humason (1931) and de Vaucouleurs & de Vaucouleurs (1964, 1967), galaxy redshift surveys did not really take off as a major scientific goal. This is because up until the late 1970's, most extra-galactic spectroscopy was dependent on image intensifier tubes, like Vidicon TV tubes or Robinson-Wampler scanners (Robinson & Wampler 1977). This was clearly not an optimal set up for constructing a large redshift survey, as revealed by the comment from Davis & Newman (2001), in a review of galaxy redshift surveys: 'Needless to say, one did not undertake redshift surveys with such tools'.

The real start of large and well sampled galaxy redshift surveys came with a techno-

---

[‖] Slipher (1917) provides measures of radial velocities of 'nebulae'. However, the notion of an extra-galactic object was not widespread at his time, and it is probably first with the compilation of radial velocities of 'nebulae' by Strömberg (1925) that some of the 'nebulae' observed by Slipher are called non-galactic nebulae. The current concept of galaxy redshift only appears with Hubble (1929a, 1929b) and Hubble & Humason (1931).

logical breakthrough: a photon-counting multichannel spectrometer (Shectman & Hiltner 1976). It is with a similar instrument coupled to a low dispersion spectrograph and mounted on a 1.5m telescope on Mt. Hopkins (Arizona, USA) that the first optical redshift survey of the modern area started: the Centre for Astrophysics redshift survey (Tonry & Davis 1979), with the first results reported by Davis, Huchra, Latham & Tonry (1982a). The CfA1 survey was soon followed by a large number of new, optically selected redshift surveys with various strange acronyms, among which we find CfA2 (Geller & Huchra 1989, Huchra, Vogeley & Geller 1999), AARs (Peterson et al. 1986), SSRS1&2 (da Costa et al. 1988, Alonso et al. 1993), Stromlo-APM (Loveday et al. 1992), Durham-UKST (Ratcliffe et al. 1996), LCRS (Shectman et al. 1996) and ESP (Vettolani et al. 1997). Obviously the start of optically selected redshift survey implied the start of galaxy redshift surveys selected in other wavebands, among which we have the three infra-red selected redshift surveys, 1.2Jy Survey (Fisher et al. 1995), QDOT (Lawrence et al. 1999) and PSCz (Saunders et al. 2000), which also had the advantage of being all-sky surveys. For completion of this section about redshift surveys, we also need to mention that the 1990's was also the start for medium to high redshift surveys like the CFRS (Lilly et al. 1995), LDSS (Ellis et al. 1996) and the CNOC1&2 (Yee, Ellingson & Carlberg 1996, Yee et al. 2000), and photometric techniques were developed to select Lyman break galaxies (Steidel et al. 1996) at redshifts between 2.5 and 3.5.

The galaxy redshift surveys of the 1980's and 1990's surveyed mostly the local universe, with a median depth steadily increasing with time from $z_{\rm med} \simeq 0.02$ for CfA1 to $z_{\rm med} \simeq 0.1$ for LCRS and ESP, but unlike surveys selected from the IRAS point source catalogue (Beichman et al. 1988), the optically selected galaxy redshift surveys covered a small fraction of the sky. This is partially due to the fact that there is no single optically selected galaxy catalogue which covers the entire sky, simply because no ground based telescope can observe the entire sky from a single location, and combining observations from different telescopes is a rather perilous task. Another issue is dust obscuration due to our galaxy, which is less a problem for far infra-red selected samples. Also going fainter (ie. going further out in redshift) is usually in competition with covering a larger portion of the sky. This is easily understandable as the volume surveyed increases roughly, for these local surveys, with the cube of redshift range probed. To this one should also mention that the fainter one goes, the longer the exposure times need to be to obtain a spectrum of good quality, from which, at least, a redshift can be measured.

CfA1 was the first wide angle redshift survey to reach beyond the Local Supercluster.

It provided strong evidence that the distribution of galaxies was far from homogeneous, showing instead filaments and voids, and produced the well-known picture of the stick man and Great Wall (see Fig. 1(a) of de Lapparent, Geller & Huchra 1986). The structure was still poorly defined, because of the sparseness of the sample, but at the same time the picture that emerged from this first redshift survey of the modern era was considerably different from that favoured until the late 1970's, in which clusters were to believed to be rare, isolated regions of high density in an otherwise uniform background (da Costa 1999). The SSRS1 improved our view of the local universe by obtaining an unobstructed view of the large scale structure through avoiding the Virgo cluster and testing the reproducibility of different statistics employed in the analysis of the CfA1 data. The various extensions of these two surveys told us that groups, filaments and voids are rather common in our local neighbourhood. The Great Wall in the north soon had an equivalent in the south. The larger the scale probed by the redshift surveys, the less striking these two walls became, as other 'walls' further out were discovered. As the size of the structures in these first redshift surveys were still of the order of the scales probed by the surveys, it became clear that it was necessary to extend the surveys to greater depth. Following Kaiser (1986), who advocated sparse sampling strategies for low order statistics in large scales structure studies, the Stromlo-APM survey sampled a volume about five times that of the combined CfA2-SSRS2 at the expense of small scale information. It covers $\sim 4300\ \square^\circ$ with a 1 in 20 sampling and has a median depth of $z_{\rm med} \simeq 0.05$. Soon after, the LCRS and ESP redshift surveys, constructed in the mid-1990's, covered a larger redshift range $z_{\rm med} \simeq 0.1$ with an intended one-in-one sampling over $\sim 700\ \square^\circ$ for the LCRS, but only $\sim 24\ \square^\circ$ for the ESP. Due to the substantially greater depth compensating for the comparatively small solid angle, the LCRS probes a similar sized volume as the Stromlo-APM, with, however, roughly 10 times more redshifts. The LCRS was the first large redshift survey using a multi-object-spectrometer, allowing it to measure simultaneously over 100 galaxy spectra.

It is in this context of galaxy redshift surveys of the local universe that both the 2 degree Field Galaxy Redshift Survey (2dFGRS hereafter; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS hereafter; York et al. 2000) are being undertaken to achieve over an order of magnitude improvement over any previous homogeneous galaxy redshift survey. They will target, using in a one-in-one sampling strategy, 250,000 and 1,000,000 galaxies respectively to median depths of $z_{\rm med} \simeq 0.115$ over $\sim 2300\ \square^\circ$ for the 2dFGRS and $\sim 10{,}000\ \square^\circ$ in the case of SDSS. But before entering into the description of these two very large redshift surveys, their scientific goals, and how they will supersede all previous redshift surveys of the local Universe, we focus next on the main scientific achievements made over the last 20 years using the galaxy redshift surveys presented above.

1.1.2 Achievements of local galaxy redshift surveys

Considerable progress has been made since the start of the first galaxy redshift surveys in the study of large scale structure, galaxy properties and galaxy formation. In the following paragraphs, we try to cover three areas relevant to this Thesis, in which galaxy redshift surveys of the last century have had their largest impact on our understanding of the local Universe**.

Galaxy luminosity functions

One of the most basic statistical quantities measurable from a galaxy redshift survey is the galaxy luminosity function, a measure of the number of galaxies per unit volume as function of luminosity. This quantity not only provides important information about the average galaxy population of the survey, which can be used to constrain theoretical models of galaxy formation (eg. Cole et al. 2000, and references therein), but is also a necessity for doing any clustering analysis on magnitude limited samples. As explained in more detail in chapter 3, the detailed knowledge of the galaxy luminosity function of the survey allows us to model a unbiased estimate of the survey selection function.

However, despite many estimates of the local luminosity function (eg. Marzke et al. (1998) for SSRS2, Zucca et al. (1997) for ESP, Loveday et al. (1992) for Stromlo-APM, Marzke, Huchra & Geller (1994) for CfA2 and Lin et al. (1996) for LCRS), there is still considerable debate over its shape and normalisation, particularly at the faint end. The nature of the discrepancies is probably well understood, but so far nobody has succeeded to provide an estimate of the local galaxy luminosity function with which most people agree. The issue of the overall normalisation is often related to variations in large scale structure from survey to survey, which so far have not been fully taken into account, particularly if one considers the small size of certain surveys. The techniques involved in the estimate of the overall normalisation have not yet achieved agreement among researchers. The other matter of concern is the faint end slope which certain studies claim to be much steeper

---

[**] da Costa (1999) provides a very nice overview of the progress made over the last 20 years thanks to galaxy redshift surveys, which indirectly inspired this section.

than others. This is partially due to an underestimate in certain surveys of the fraction of missed dwarf galaxies, due to for example surface brightness selection effects, as pointed out recently by Driver (1999), or simply unrepresentative samples of faint galaxies due to the small size of the survey. We revisit these issues in chapter 3 and chapter 4 in the context of the 2dFGRS and the combined 2MASS–2dFGRS catalogue.

Galaxy clustering and biased galaxy formation

The two point correlation function is the most widely used statistic to quantify galaxy clustering. It is used for magnitude limited samples, after taking into account the associated selection function, and for volume limited samples, for measuring deviations in the galaxy distribution from a uniform unclustered distribution. Unlike the case of galaxy luminosity function analyses, the clustering measurements from redshift surveys of different depths, geometry and sampling give amazingly robust and consistent results. Davis et al. (1985) made pioneering theoretical predictions of clustering in cold dark matter[††] cosmologies (CDM hereafter; Blumenthal et al. 1984). In order to match the extant robust measurements of the two-point correlation function, galaxies were required to be biased tracers of the dark matter. This was the start of biased galaxy formation models, which is still the current theoretical framework of galaxy formation.

Based on results from models of infall of matter onto clusters of galaxies (e.g Gunn & Gott 1972), Sargent & Turner (1977) show, in the context of hierarchical galaxy formation scenarios, that the apparent distance between well separated galaxies, which tend to move together because of large inhomogeneities in the density distribution, is reduced, making the large scale galaxy clustering appear flattened along the line of sight. Following these theoretical predictions, Kaiser (1987) presented a method relating this apparent flattening of structures to the total mass-density parameter, Ω. Previous redshift surveys have not been able to measure with great accuracy this apparent anisotropy of galaxy clustering on intermediate scales, mostly due to survey size and sampling. In chapter 5, we show for the first time in great detail the reality of the infall of galaxies onto large scale structure due to gravitational instability, with a precise measurement of the flattening parameter, β, which depends on the total mass-density parameter, Ω, and the 'bias' parameter of luminous galaxies, b.

[††] The name cold dark matter was invented by James Bond (Frenk private communication).

Dependence of biasing on galaxy properties

Another great success of past galaxy redshift surveys has been the study of the influence of galaxies properties on clustering. One way of comparing in detail theoretical predictions of biased galaxy formation with observational data is to study galaxy clustering as a function of luminosity, as luminosity is the simplest parameter to deal with in both theoretical models and observations. In the literature, there is still a large amount of disagreement on the dependence of clustering as function of luminosity. We show, in chapter 6, that a large homogeneous survey like the 2dFGRS allows accurate measurements which can be compared to the latest predictions of galaxy formation theories.

Most surveys, like SSRS2 or Stromlo-APM, are based on photographic surveys for which a wealth of additional information, like morphology and colour, is available for each galaxy. In order to put further constraints on models of galaxy formation, galaxy samples of different physical characteristics have been analyzed. Many trends depending on the property of the selected galaxy population have been discovered, but due to the small size of these surveys, the conclusions from these studies are not unanimous.

Certainly the difficulty of splitting samples in a homogeneous and well understood way and still having a big enough sample to be statistically representative is part of this problem. Hence in chapter 7, with large and extremely well defined galaxy samples we consider the dependence of galaxy clustering on luminosity and spectral type. These very accurate measurements allow us to consider for the first time how biasing depends on spectral class as function of luminosity. These well defined samples can be compared to theoretical models, once they become available. These measurements will, without any doubts, tighten the constraints on models of biased galaxy formation.

1.1.3 2dFGRS and SDSS

The effective context of the 2dFGRS (Colless et al. 2001) is set indirectly by the Sloan Digital Sky Survey (SDSS hereafter; York et al. 2000), which has similar scientific goals, and plans to measure redshifts for ~1 million galaxies over 10^4 □° to a comparable magnitude limit. The SDSS spectroscopic survey started in April 2000 and will take 5 years to complete. Although the 2dFGRS has a head-start, and is expected to reach its goal before SDSS achieves a similar sample size, the long-term view will surely be that the two surveys are *both* necessary and complementary, for two reasons:

1. The main scientific goals of both surveys are of fundamental importance to cos-

mology and to studies of the galaxy population. Two independent surveys, with different input catalogues, selection criteria and analysis techniques, are essential to ensure that reliable, cross-checked results are obtained. A classic example of the necessity of independent work on fundamental problems is the Sandage/deVaucouleurs controversy over the value of H_0 (eg. de Vaucouleurs 1958, Sandage & Tammann 1975, Sandage & Tammann 1990); a more recent example is the double-checking of the measurements of the acceleration of the Hubble expansion by the two supernovae teams (Riess et al. 1998, Perlmutter et al. 1999), which surely underlies the rapid acceptance of the surprising results favouring a cosmological constant.

2. Both the 2dFGRS and SDSS will provide invaluable long-term databases from which the astronomical community can select subsamples of particularly interesting objects for follow-up observations or analysis. The 2dFGRS will be far more suitable for southern-hemisphere observers, just as SDSS will be preferred by northern-hemisphere observers.

With this very brief comparison between the two ongoing large galaxy redshift surveys, 2dFGRS and SDSS, we draw to an end to this general introduction of past galaxy redshift surveys. For further details about galaxy redshift surveys of the 1980's and 1990's, we recommend the reviews of da Costa (1999) and Davis & Newman (2001). From now on, we focus our attention mainly on the 2dFGRS, and we begin by introducing the scientific goals behind the survey, followed by a quick review of the proposed observing strategy.

1.2 The 2 degree Field Galaxy Redshift Survey: 2dFGRS

The 2dF Galaxy Redshift Survey (2dFGRS) aims to achieve order-of-magnitude gains over previous redshift surveys in order to realise major qualitative and quantitative advances in understanding large-scale structure and galaxy formation (Colless 1999, Colless et al. 2001). The survey builds on the APM galaxy catalogue (Maddox et al. 1990b, 1990c, 1996), which contains image parameters for $>10^6$ galaxies to $b_J \approx 20.5$, and is made possible by the development of the 2dF 400-fibre spectrograph at the Anglo-Australian Observatory (Taylor et al. 1997). The main scientific goals of the 2dFGRS are:

1. A precise determination of the power spectrum of the galaxy distribution on all scales up to several hundred Mpc, overlapping the regime probed by CMB experi-

ments. Comparison of the large-scale power in the galaxy distribution at the present epoch and in the mass distribution at the surface of last scattering provides strong constraints on the nature of the dark matter and the form of the primordial density fluctuations.

2. A measurement of the redshift-space distortion of the clustering produced by peculiar velocities in the linear (large-scale) as well as the non-linear (small-scale) regime. The degree of distortion on large scales is a direct measure of the density parameter Ω and the biasing of the galaxies with respect to the dark matter; the distortion on small scales provides information about the non-linear evolution of clustering.

3. An analysis of higher-order clustering statistics, such as the bispectrum and the topological genus, in order to test whether the initial fluctuations were a Gaussian random field, as predicted by most inflationary cosmological models.

4. A detailed characterisation of the physical parameters of the galaxy population, such as luminosity, spectral type and star-formation rate, as a function of local environment. The variations in galaxy luminosities, ages and metallicities with local density and structure provide strong constraints on models of biasing and hierarchical galaxy formation.

5. The construction and analysis of a projection-free catalogue of groups and clusters. This will be used to measure the luminosities, masses and clustering properties of collapsed systems on all scales, as a diagnostic of large-scale environment, and to correct for non-linear effects in the analysis of redshift-space distortions.

Although these are the primary goals of the 2dFGRS, a massive uniform redshift survey has many other applications. In particular, it will be an invaluable data source for selecting samples of interesting objects for detailed follow-up based on both photometric and spectroscopic criteria, especially for southern-hemisphere observers. The potential for significant serendipitous discoveries should also not be neglected in a survey of this size.

In order to achieve these goals, we must probe deeply into the luminosity function while also covering a cosmologically representative volume. All past redshift surveys either compromised on volume, or sampled only a small fraction of the galaxies. The 2dFGRS will thus produce the first high-fidelity map of the full large-scale galaxy distribution.

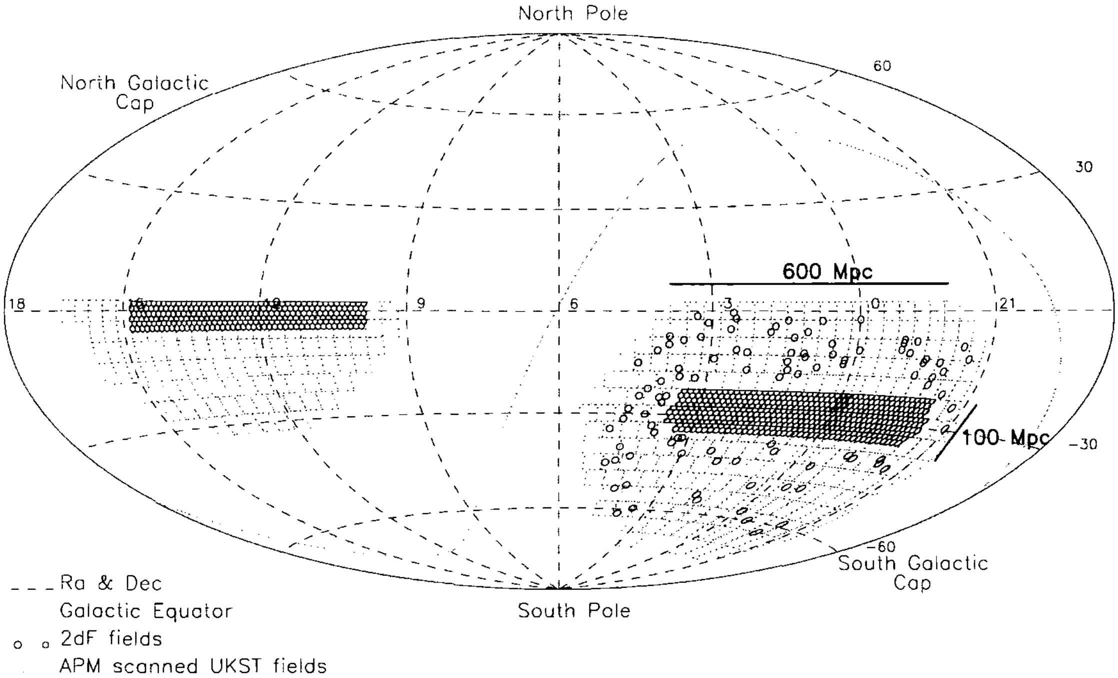


Figure 1.1: The 2dF survey geometry shown in an Aitoff projection of right ascension and declination. The grey areas show the NGP and SGP survey strips and the 99 random fields around the SGP.

1.2.1 Geometry

The various scientific goals make differing demands on the survey. For some, such as clustering on very large scales, the largest possible volume must be sampled. For this question alone, the sparse-sampling strategy used with success in some other redshift surveys might be adopted. However, many of the goals demand that the survey should achieve the highest possible space density of galaxies; for example, to test theories of galaxy formation we need to map the distribution of galaxies with the maximum possible resolution. The performance of 2dF itself also imposes constraints; the time required to configure a field is ~1 hour, so that on average the survey can cover about 7 fields per clear night.

The survey geometry we have chosen is therefore a compromise, with two contiguous strips and a surrounding 'halo' of randomly-selected fields (see Fig. 1.1). The southern strip, centred on the SGP, is $75^\circ \times 15^\circ$ (corresponding to $600\,h^{-1}\,\mathrm{Mpc} \times 120\,h^{-1}\,\mathrm{Mpc}$ at twice the median depth of the survey), while the northern, equatorial, strip is $75^\circ \times 7.5^\circ$ (corresponding to $600\,h^{-1}\,\mathrm{Mpc} \times\ 60\,h^{-1}\,\mathrm{Mpc}$). The 99 random fields, contributing about 15% of the galaxies, are drawn from the full 6400 $\square^\circ$ area of the APM galaxy catalogue, and sparse-sample a volume of approximately $10^8\ \mathrm{h}^{-3}\,\mathrm{Mpc}^3$. Their purpose is to improve the window function of the survey, giving it a better grasp on very large-scale structure. If observational selection imposes a mask $W(\mathbf{r})$, then the power spectrum of the observed galaxy distribution is related to the power spectrum in an ideal survey via convolution with the window function (the squared Fourier transform of the mask): $P_{\rm obs}(\mathbf{k}) = P_{\rm true}(\mathbf{k}) * |\tilde{W}|^2(\mathbf{k})$. The aim in choosing the survey geometry was to construct a window function $\tilde{W}$ as close as possible to a compact sphere in **k**-space—unbiased and precise estimates of $P_{\rm true}(\mathbf{k})$ require that $\tilde{W}$ is as compact as possible and nearly isotropic, without 'sidelobes' in special directions. The random fields are chosen so as to give the 2dFGRS a window function of this form.

1.2.2 Sampling

One of the key aims of the 2dFGRS is to investigate details of the galaxy density field, both in terms of its morphology, and how it varies between different galaxy types. These aims require a high space density of galaxies to sample the density field accurately, and argue strongly against any form of sparse sampling strategy in the main survey areas. For a fixed amount of telescope time, the balance between the number of fields and the integration time for each field determines the available survey volume as a function of the space density of galaxies. We found that the volume sampled at $\rho{\sim}10^{-2}\,h^3\,{\rm Mpc}^{-3}$ is not a sensitive function of magnitude limit between $b_J=19$ and $b_J=20$. We set the survey limit at $b_J=19.45$, which maximises the observational efficiency by matching the integration time necessary to achieve a suitable signal-to-noise ratio with the configuration time required to go from one field to the next. By choosing to sample all galaxies above the magnitude limit, the space density is as high as nature allows. The magnitude limit means the 2dFGRS has a median redshift of $\langle z\rangle=0.11$ and reaches out to $z \sim 0.3$, thus covering a cosmologically representative volume.

Achieving the maximum efficiency in assigning fibres to galaxies is complicated by the need to divide the sky into distinct 2°-diameter fields that 2dF can observe. Since the density of galaxies varies on the sky, the most efficient way to cover the sky with near-complete sampling is to employ adaptive tiling: a uniform grid of field centres is locally perturbed in order to increase the overlaps in high-density regions. Our algorithm allows fibres to be placed on 96% of all galaxies while efficiently covering the sky (the mean effective area per pointing is 1.8 □°, compared to 2.6 □° for an ideal hexagonal tiling).

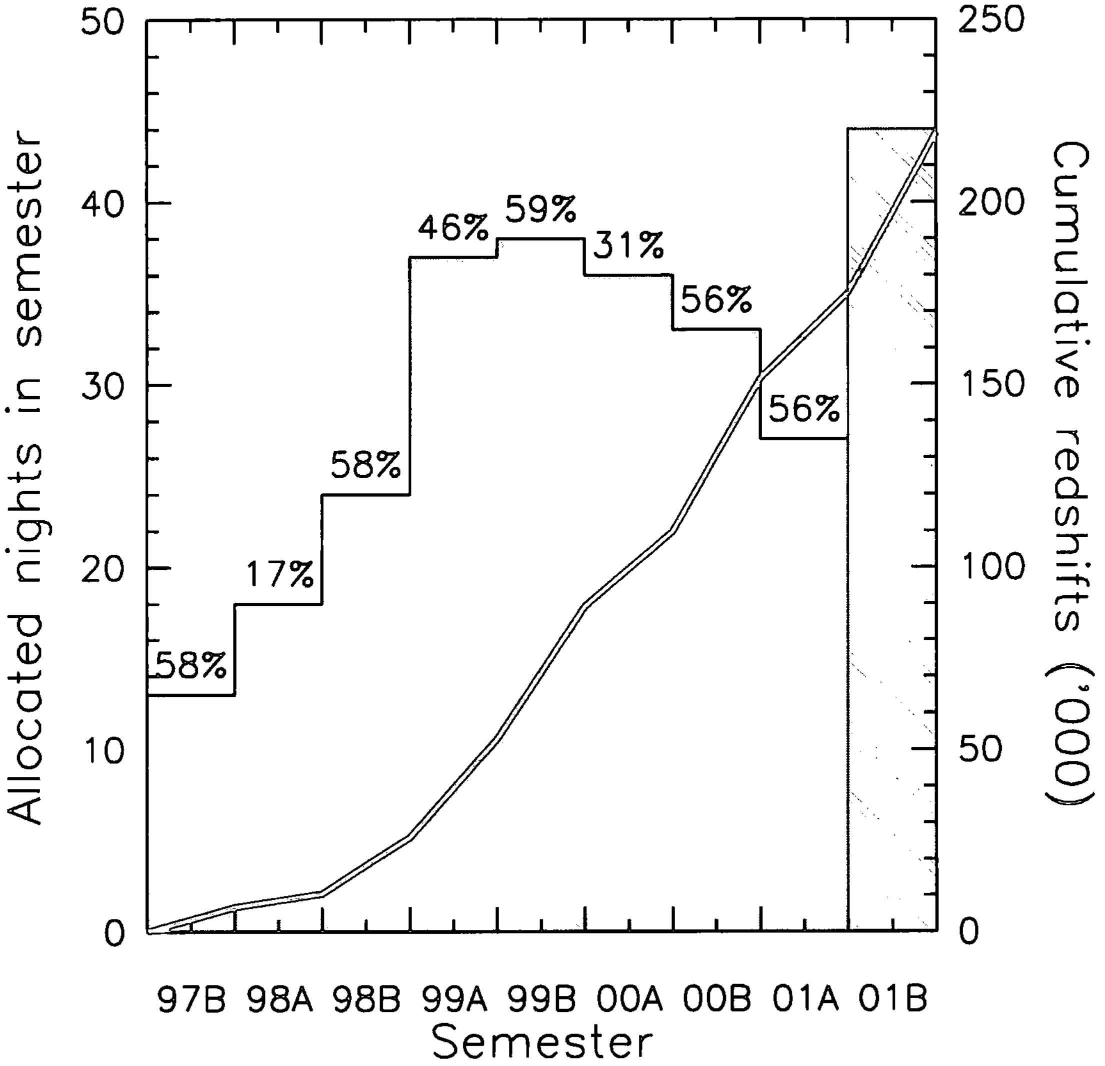


Figure 1.2: Allocated nights and survey progress. The solid histogram shows the number of nights allocated, jointly with the QSO survey, in each semester since the start of the survey. The thick line shows the accumulation of new galaxy redshifts. The dashed histogram is a projection, based on the current success rate, for the last 2dFGRS observing semester.

1.2.3 2dF QSO survey

A further increase in observational efficiency is achieved by carrying out the 2dFGRS observations simultaneously with those for the 2dF QSO redshift survey (Boyle et al. 2000), which uses a subset of the 2dFGRS fields and the same integration time. The QSO survey has about 200 targets per 2dF field that are virtually uncorrelated with the galaxies. By merging the target lists for the two surveys and generating a single tiling pattern, the total time needed to carry out both surveys is reduced by 25%. Furthermore, the additional surface density provided by the QSOs means that the field centres are closer together, making it easier to sample high-density regions and close pairs of galaxies and so achieve the goal of near-uniform sampling.

1.3 Survey Status

At the time of writing, September 2001, the survey is not yet finished. However, the 2dFGRS will not officially be allocated any more observing time beyond the end of semester 2001B. This means that observing for the current 2dFGRS will come to an end in January 2002. This section outlines the current survey status, discusses the evolution of the survey completeness, and ends with a brief note on the survey's future.

1.3.1 Summary statistics

We list some relevant summary statistics for the status of the survey as of July 2001 in Table 1.1. Statistics are given for the whole survey and separately for the NGP strip, SGP strip and random fields. The table covers semesters 1997B to 2001A, over which period the 2dFGRS was allocated 226 nights. We have been allocated a further 44 nights in 2001B.

Fig. 1.2 shows the progress of the survey to date. The black histogram shows the number of nights allocated to the 2dFGRS (jointly with the QSO survey) in each semester since the start of the survey. The percentage of usable time (i.e. excluding time lost to weather or major technical problems) is indicated; overall, 49% of the time was usable. The solid black line shows the accumulation of new galaxy redshifts. The slow initial progress was due to the delays in achieving 2dF's full positioning speed and reliability, and to very poor weather in 1998A. Unfortunately bad weather and instrument problem caused semester 2000A to be very poor. For the last two semesters the rate is close to 1000 new galaxy redshifts per *allocated* night. This rate is used to project the expected

Table 1.1: Summary statistics for the 2dFGRS as of July 2001.

	All	NGP	SGP	Random
2dF fields observed	727	311	345	71
Total Fields	1 192	450	643	99
Objects observed	213 599	89 543	99 760	24 296
Total Repeats	18 102	8 295	9 201	606
Repeated (unique) objects	16 688	7 567	8 515	606
Unique objects	195 497	81 248	90 559	23 690
Objects with redshifts	182 955	76 753	84 925	21 277
Redshift completeness	93.6%	94.5%	93.8%	89.8%
Objects with redshift problems	338	123	202	13
Stars ($z < 0.002$)	9 744	4 177	4 111	1 456
QSOs ($z > 0.75$)	53	22	22	9
Galaxies	172 817	72 430	80 589	19 798
Total targets	382 680[‡‡]	132 111	193 550	57 019
Survey completeness	47.7%	58.0%	43.8%	37.3%

[‡‡] See section 2.2 for a more detailed explanation of why this number differs from 250 000.

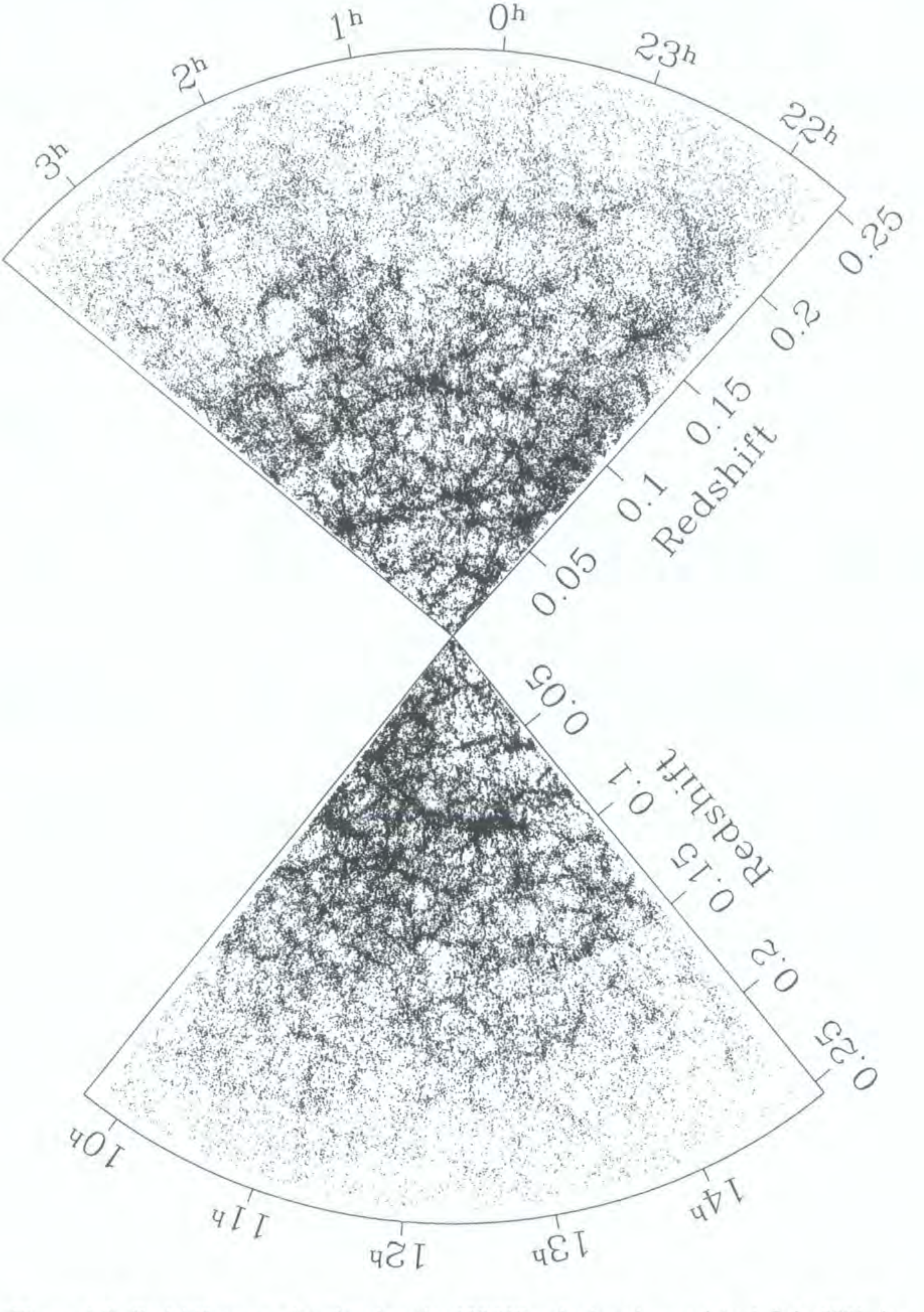


Figure 1.3: Redshift cone plot showing the 153 019 galaxies observed to 1 July 2001, in the two main strips. Top is the SGP region and bottom is the NGP region.

total of ~220,000 redshifts by the end of 2001B (faint solid line and dashed histogram).

The 2dFGRS is the first survey to measure effectively redshifts of all the galaxies that exist over a cosmologically representative volume. Fig. 1.3 is a redshift cone diagram showing all the galaxies in the main survey strips observed to date. The effects of variable sampling in the current state of the survey are still visible at some level. The detail and fidelity with which the galaxy distribution will be recovered by the full survey, and its ability to discriminate between cosmologies and galaxy formation models, is demonstrated in Fig. 1.4. The figure compares a small, and nearly complete, redshift slice with four matching mock slices constructed from ΛCDM and standard CDM simulations using bias schemes 1 and 4 of Cole et al. (1998). Scheme 1 bases the selection probability on the value of the smoothed initial density, whereas in scheme 4, a sharp cut off is applied on the final smoothed density field, prohibiting galaxies appearing in very underdense regions, and giving equal probability of formation of a galaxy once the overdensity rises above a certain threshold.

1.3.2 Evolution of the survey completeness

The overall completeness of the survey is the product of the sampling rate (the observed fraction of the input catalogue) and the spectroscopic completeness (the fraction of observed objects with redshifts). A theoretical estimate of the sampling rate is made in Colless et al. (2001) and is approximatively 93%. The current spectroscopic completeness is 93.6%. The final overall completeness is expected to be close to 85%, as the sampling rate is expected to be ~90% and the spectroscopic completeness might still increase with time, as the reliability of the whole 2dF system usually increases.

This completeness is however very time dependent, as shown by the following set of figures, showing the completeness of survey as of June 1999 (Fig. 1.5), June 2000 (Fig. 1.6) and July 2001 (Fig. 1.7). Here the completeness, defined locally, is simply the ratio of the number of redshifts obtained to the number of galaxies in the input catalogue. Details of the construction of these completeness masks is set out in chapter 2. One can clearly see that it is only within the last year that the survey has started to reach a uniform and highly complete sampling over both NGP and SGP regions. The patchiness of the survey completeness has been dramatically reduced between 1999 and 2000 for the NGP region, and similarly between 2000 and 2001 for the SGP region, as shown by panels (c) and (d) of Fig. 1.5 and Fig 1.6 and 2nd and 4th panel of Fig. 1.7.

As an example, let us discuss the completeness of the survey, for the remainder of

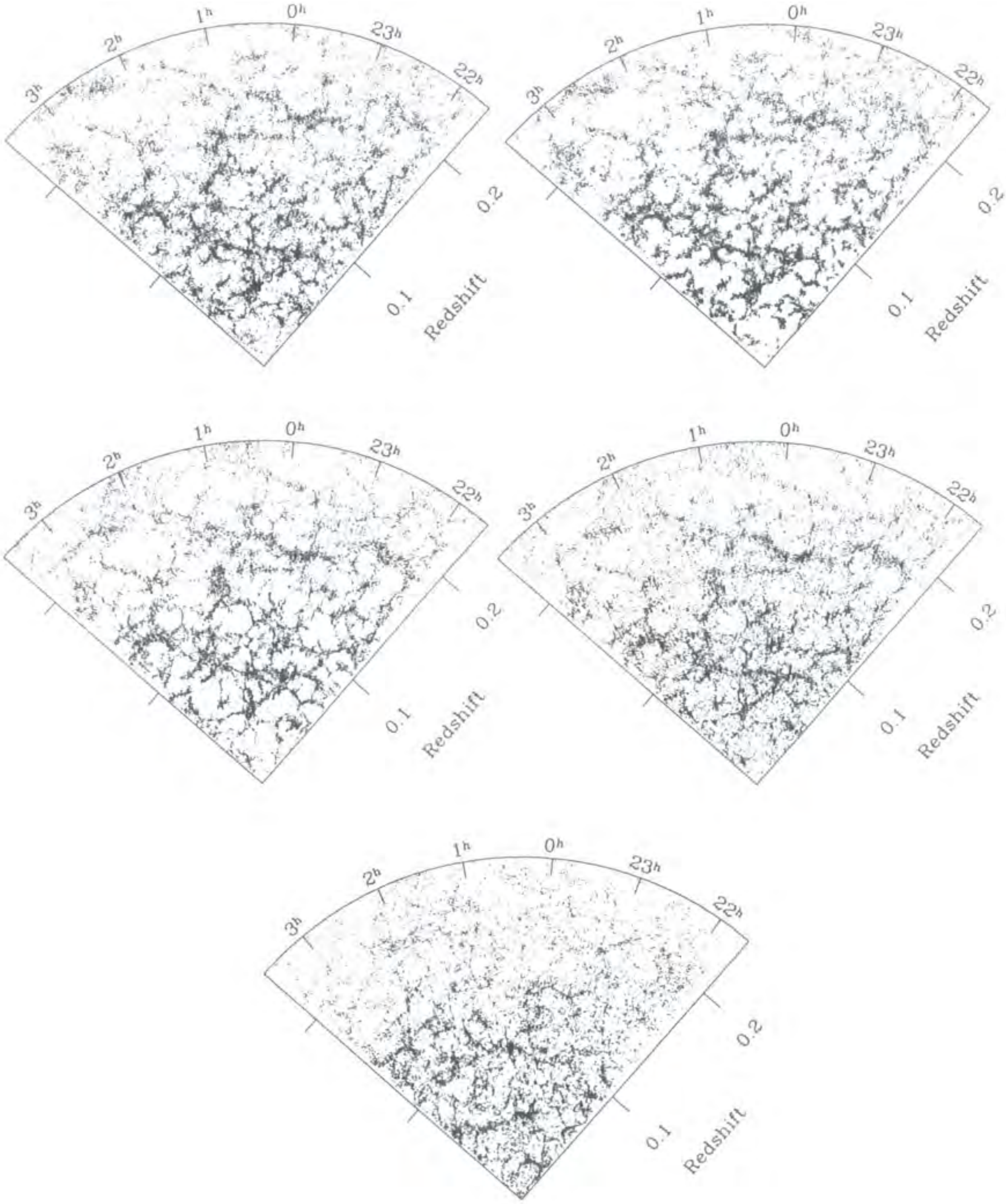
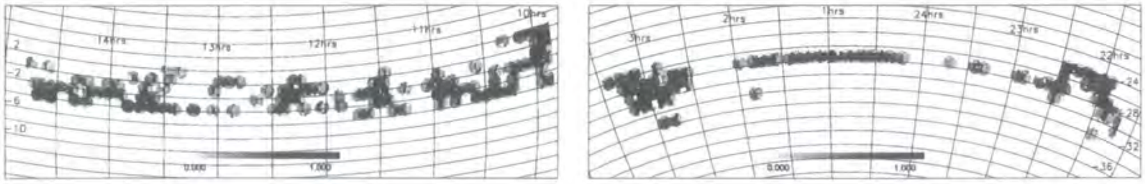
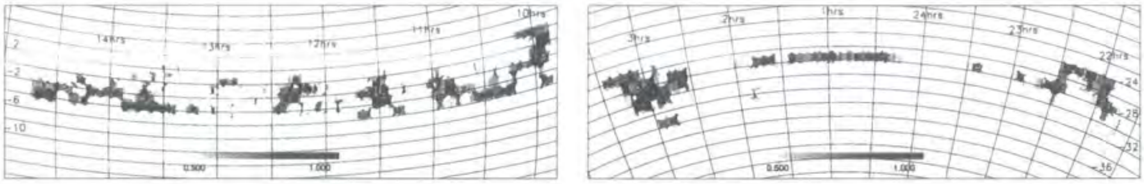


Figure 1.4: The observed distribution in a nearly-complete 2°-thick slice of the SGP strip (bottom), compared to four mock slices drawn from ΛCDM (middle) and standard CDM (top) simulations using different biasing schemes, explained in more detail in Cole et al. (1998).



(a) NGP: June 1999

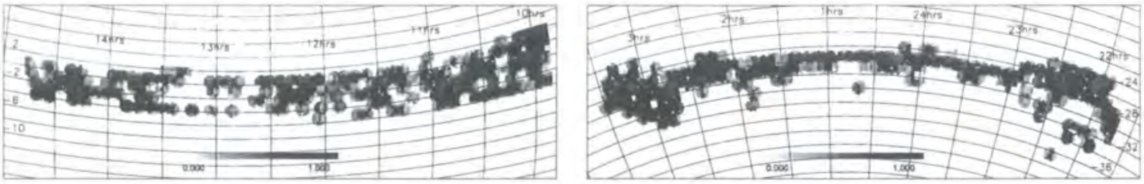
(b) SGP: June 1999



(c) NGP: June 1999

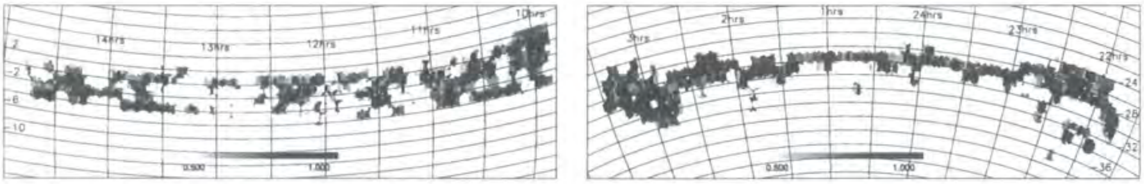
(d) SGP: June 1999

Figure 1.5: Completeness maps as of June 1999, for both NGP and SGP regions. We show in panels (c) & (d) regions in which the completeness is above 50%. The visual impression is that of a very patchy sky coverage strongly correlated with the observed fields and how they overlap with each other. Approximatively 55,000 redshifts had been measured at this time: the 2dFGRS was already the biggest single galaxy redshift survey, containing roughly half of all galaxy redshifts ever measured.



(a) NGP: June 2000

(b) SGP: June 2000



(c) NGP: June 2000

(d) SGP: June 2000

Figure 1.6: Completeness maps as of June 2000, for both NGP and SGP regions. We show in panels (c) & (d) regions in which the completeness is above 50%. These regions contains $\sim 68\,000$ of the $110\,000$ galaxy redshift measured at that time.

this section, as if we were in June 2000, which corresponds to a redshift survey with approximatively 110,000 redshifts.

Part-way through the survey, the completeness is highly non-uniform. This is an unavoidable consequence of the adaptive tiling strategy, in which uniform completeness is achieved only when all the overlapping tiles have been observed. This shows clearly why it is important to construct a completeness mask which tells us which regions of the sky can be used in statistical analyses. Panels (a) & (b) of Fig. 1.6 show the completeness in the NGP and SGP strips respectively. In panels (c) & (d) of the same figure, we show the areas of the NGP and SGP strips which are more than 50% complete—only $\sim$ 68,000 of the 110,000 galaxies observed to June 2000 lie in regions with completeness >50%. Thus for the many survey applications which require high completeness, finishing the survey will increase the size of the usable sample by at least a factor of 5, and at the same time will significantly reduce the overall patchiness of the survey.

Even applications for which variations in sampling can in principle be accounted for are problematic in the survey as of June 2000, since the sampling is still correlated with fluctuations in the galaxy distribution. The adaptive tiling algorithm generally arranges for field overlaps in high-density regions, which means that low-density regions are over-represented in the existing data (since they generally only need to be observed once, whereas high-density regions typically require overlapping 2dF fields for full sampling). This effect has unfortunately been exacerbated because for a long period in 1999 there was a relatively high fraction of broken fibres in 2dF. During that period we preferentially observed low-density fields so as not to compromise the completeness in high-density fields, which require more operational fibres. This maximised the efficiency of completing the full survey at the temporary expense of uniformity.

Thus, at this intermediate stage (June 2000), the survey was compromised by highly-variable completeness and a systematic bias towards low-density regions. Not only is it therefore necessary to finish the full survey in order to achieve a uniformly high level of completeness in our sample, but also to correct the potential remaining bias towards low-density regions of the universe.

1.3.3 Some reasons for pursuing the 2dFGRS

The survey is as of September 2001, 47.7% complete, with redshifts for $\sim$ 180 000 objects out of a target total of $\sim$ 380 000. However, if we consider the target goal of 250 000 galaxies (as intended in the accepted 2dFGRS proposal), we can confirm that we have

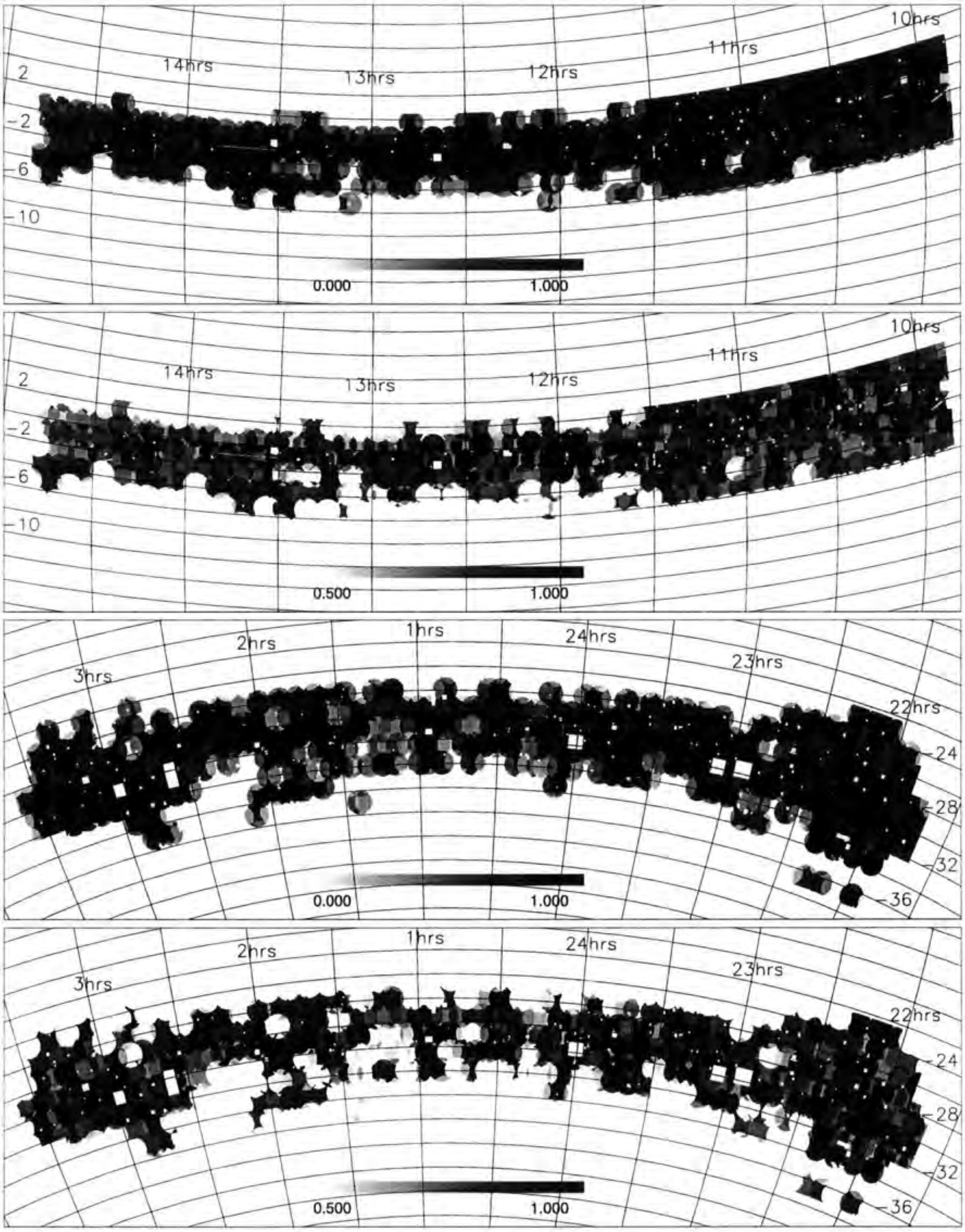


Figure 1.7: Completeness map of the 2dFGRS as of 1st of July 2001. The two top panels show the completeness of the NGP region, with the second panel showing only those regions with completeness above 50%. The bottom two panels are for the SGP region, with the lower one again showing regions with completeness above 50%.

observed ~78% of the targets and measured redshifts for 73% of the targeted goal of 250 000 objects. The goal of measuring 250 000 galaxy redshifts is probably not going to be achieved with the remaining allocated observing time. Instead, we estimate this final number to be close to 220 000 galaxy redshifts.

The observing strategy was designed to achieve uniform sampling and high completeness for the entire survey as efficiently as possible, which is shown to have been achieved in the last year of observing. However, there are still valid reasons for pursuing the 2dFGRS beyond the figure of a quarter of a million galaxies, which we briefly enumerate below:

- Feasibility. The survey strategy beyond 250 000 galaxies is already in place. Extending the survey beyond the 380 000 targets is an easy task in the SGP region, as the full photometric input catalogue, of similar quality to the one currently used, is already available. This is however not the case for the NGP, which directly limits a NGP extension.
- Representativeness. A fundamental goal of the 2dFGRS is to cover a large enough volume to be a fair sample of the universe. The existing survey will reach this goal, which is particularly important for measuring structure in the galaxy distribution in the linear regime on scales $\gtrsim 100\,h^{-1}$ Mpc, overlapping the scales probed by CMB measurements. However any further extension of the survey can only improve the statistics on large scales.
- Divisibility. An extended version of the 2dFGRS will provide a larger and even more uniform sample to permit new fine-grained and multi-dimensional analyses of the physical properties of the galaxy population, and to explore further the links between large-scale structure and galaxy formation.

In any case, finishing the full survey will significantly improve the precision with which we can measure the fundamental quantities of interest, and also reach the initial goal of representativeness and divisibility of the survey. However, there is still debate on how to pursue an extension, as this would require a substantial amount of additional telescope time.

1.4 Aims of This Work

Our aim in the subsequent chapters is to analyze the 2dFGRS. As this introduction has shown, many survey aspects need to be carefully understood and modelled before any quantitative analysis can be carried out. Once this survey modelling is finished, we can start quantifying the most important physical quantities obtainable from the whole survey.

Therefore the outline of this Thesis as follows. In chapter 2, we quantify the 2dFGRS angular selection function whereas the radial selection function, along with the 2dFGRS b_J luminosity function, is quantified in chapter 3. We pursue luminosity function estimates in chapter 4 by combining the 2dFGRS with the near infra-red selected 2MASS catalogue, in order to look at K and J-band luminosity functions as well as the stellar mass function of galaxies. In chapter 5 we introduce the two point correlation function and look at redshift space distortions, with a estimation of the flattening parameter β. In chapter 6 we consider the luminosity dependence of galaxy clustering, using real space clustering statistics. Similar methods are used in chapter 7 to study how clustering depends on luminosity and spectral class. The last chapter, naturally, summarizes our conclusions.

Chapter 2

Completeness Analysis of the 2dFGRS

This chapter introduces in detail all the aspects of the 2 degree Field Galaxy Redshift Survey (2dFGRS) which are of importance for this thesis work. In section 2.1, we describe the source catalogue and its recalibration, and in section 2.2 we present the survey design and the proposed observing strategy. In section 2.3 we define four maps or masks*, quantifying first the imperfections of the input catalogue, and then those of the redshift catalogue. A full description of the input catalogue requires a spatially varying magnitude limit mask to take into account its photometric recalibration (§2.3.1), and a map of the stellar contamination (§2.3.2). The description of the redshift catalogue requires a spatially varying redshift completeness mask to quantify the variation of the sampling rate across the sky (§2.3.3) and a spatially varying magnitude completeness mask, also called μ-mask, to quantify the magnitude dependence of the redshift success (§2.3.4). In section 2.4, we present a modification of the masks designed to correct the redshift catalogue when an upper redshift cut is applied to the input catalogue, which is relevant for clustering analysis of galaxies with spectral type. The last section of this chapter is devoted to the two main weighting schemes we have developed for this survey.

2.1 Source Catalogue

2.1.1 Initial source catalogue

The source catalogue for the survey (Maddox et al. 2001, in preparation) is a revised and extended version of the APM galaxy catalogue (Maddox et al. 1990a,b,c;1996). This catalogue is based on Automated Plate Measuring machine (APM) scans of 390 plates from the UK Schmidt Telescope (UKST) Southern Sky Survey. The extended version of the APM catalogue includes over 5 million galaxies down to b_J=20.5 in both north and south Galactic hemispheres over a region of almost 10^4 deg^2 (bounded approximately

---

*Throughout this Thesis, we use the words 'mask' and 'map' interchangeably.

by declination $\delta \leq +3°$ and Galactic latitude $b \geq 30°$). Small regions around bright stars, satellite trails and plate flaws are excluded; these are accounted for by the survey mask (see section 2.3).

The b_J magnitude system for the Southern Sky Survey is defined by the response of Kodak IIIaJ emulsion in combination with a GG395 filter. It is zeropointed to Vega—i.e. b_J is equal to Johnson B for an object with zero colour in the Johnson-Cousins system. The colour equation is normally taken to be

$$b_J = B - 0.28(B - V) \ , \tag{2.1}$$

following Blair & Gilmore (1982), who studied a sample of 85 standard stars, with a representative colour distribution in the range $-0.1 \lesssim (B-V) \lesssim +1.6$. A larger coefficient ($-0.35$) has been suggested by Metcalfe et al. (1995), but we measure -0.27 ± 0.02 in comparison with the ESO Imaging Survey (Arnouts et al. 2001), and we therefore retain the usual value of -0.28. The photometry of the catalogue is calibrated with numerous CCD sequences and, for galaxies with b_J=17–19.45, has a 68% spread of approximately 0.15 mag, but with a non-Gaussian tail to the error distribution. We emphasise that the calibration is to *total* CCD photometry, which absorbs any remaining correction to the thresholded APM magnitudes.

The star-galaxy separation is as described in Maddox et al. (1990b), with the locus dividing stars and galaxies chosen to exclude as few compact galaxies as possible, while keeping the contamination of the galaxy sample by stars to about 5%. Spectroscopic identifications of the survey objects show that the stellar contamination is in fact 5.4% (see further discussion in §2.3.2).

The source catalogue is incomplete at all magnitudes due to various effects, including the explicit exclusion of objects classified by the APM as merged images, the misclassification of some galaxies as stars, and the non-detection (or misclassification as noise) of some low surface brightness objects. We study this incompleteness in chapter 3 with respect to data from the SDSS Early Data Release. Complimentary studies using deeper wide-area CCD photometry have been done by Pimbblet et al. (2001) and Cross & Driver (2001, in preparation). The overall level of incompleteness is 10–15% and varies slightly with apparent magnitude, being largest for the brightest and faintest objects. The main classes of objects that are excluded are: (i) merged galaxy images that are explicitly excluded from the 2dFGRS source catalogue (about 60% of the missing objects); (ii) large galaxies that are resolved into components that are classified as stellar, merged or noise objects

(20%); (iii) compact normal galaxies that are detected but classified as stars (15%); and (iv) low surface brightness galaxies that are either not detected or classified as noise objects (5%). Thus the main cause of incompleteness is misclassification of objects rather than their non-detection.

The target galaxies for the 2dFGRS were selected to have extinction-corrected magnitudes brighter than b_J=19.45. The b_J extinction is taken to be A_{b_J}=4.035$E(B-V)$, where the coefficient, and the reddening $E(B-V)$ as a function of position, come from Schlegel et al. (1998). The limit of b_J=19.45 was chosen because: (i) The surface density of galaxies at b_J=19.45 (approximately 165 deg^{-2}) is sufficiently larger than the surface density of 2dF fibres on the sky (127 deg^{-2}) to allow efficient use of all fibres—few fibres are unused even in low-density fields. (ii) The time taken to configure a typical field (60–65 min) allows, with overheads, a sufficiently long exposure time to reach the desired signal-to-noise level of $S/N>10\,\mathrm{pixel}^{-1}$ for galaxies with b_J=19.45 even in rather poor conditions. This limiting magnitude corresponds to a median redshift for the survey of about $\bar{z}$=0.1, so that the 2dFGRS is essentially a survey of the *local* universe.

2.1.2 Recalibration of initial source catalogue

Although the 2dFGRS sample was originally selected to have a uniform extinction-corrected magnitude limit of b_J=19.45, in fact the survey magnitude limit varies slightly with position on the sky. There are two reasons for this. First, more extensive photometric calibrations are now available than when the parent 2dFGRS catalogue was originally defined. This has enabled us to recalibrate the whole 2dFGRS parent catalogue (Maddox et al. 2001, in preparation), and results in new zero-point offsets and linearity corrections for each of the UKST photographic plates. Second, the extinction corrections have been changed to use the final published version of the Schlegel et al. (1998) extinction maps; the original extinction corrections came from a preliminary version of those maps.

For completeness, we give here a brief outline of the recalibration procedure used. B-band data from the Millenium Galaxy Catalogue (MGC) was used to measure the relative zero-point errors and non-linearity for a strip of UKST plates in the NGP. The required corrections were found to correlate with the offsets measured at bright magnitudes relative to 2MASS data[†]. All plates for which MGC data was not available were corrected based on the 2MASS data. The overall zero-point of the survey was then set using the ESO Imaging Survey B & V-band data, from which our accurate b_J magnitude can be derived.

[†] 2MASS: two Micron All Sky Survey. For more details about this survey, see chapter 4.

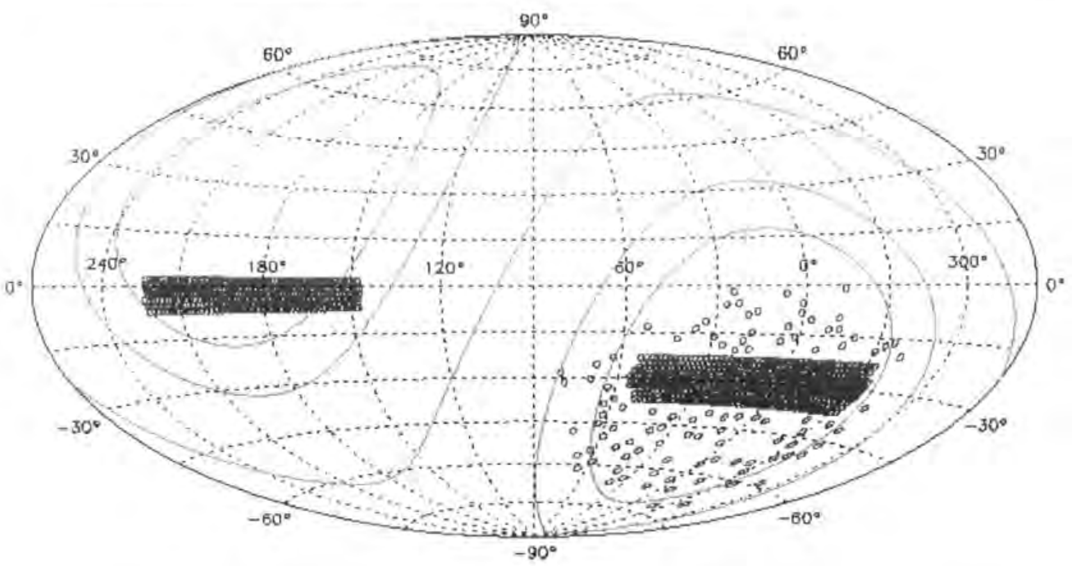


Figure 2.1: The 2dFGRS regions shown in an Aitoff projection of R.A. and Dec., with individual 2dF fields marked as small circles. Also shown are the lines of Galactic latitude $|b|=0°$, $30°$, $45°$. The numbers of survey galaxies in these regions are: 193 550 in the 643 fields of the $90° \times 15°$ SGP strip, 132 111 in the 450 fields of the $75° \times 10°$ NGP strip, and 57 019 in the 99 fields scattered around the SGP strip.

2.2 Survey Design

The areas of the sky covered by the survey were chosen so as to satisfy a number of different requirements. The first goal was to cover as large a volume as possible, in order to closely approach a statistically representative sample of the Universe on the largest possible scales. The second was to obtain near-complete sampling down to the survey limit in order to have the finest possible resolution of structure on small scales. The third requirement was to match the sample to the observational capabilities of the 2dF instrument in order to achieve high efficiency. The adopted geometry is an effective compromise between these requirements.

The survey consists of two separate declination strips of overlapping 2° fields plus 99 scattered 'random' 2° fields. One strip (the SGP strip) is in the southern Galactic hemisphere and covers approximately $90° \times 15°$ centred close to the South Galactic Pole ($21^{h}40^{m} < \alpha < 03^{h}40^{m}$, $-37.5° < \delta < -22.5°$). The other strip (the NGP strip) is in the northern Galactic hemisphere and covers $75° \times 10°$ ($09^{h}50^{m} < \alpha < 14^{h}50^{m}$, $-7.5° < \delta < +2.5°$). The 99 'random' fields are chosen from the low-extinction region of the APM catalogue in the southern Galactic hemisphere outside the survey strip (the

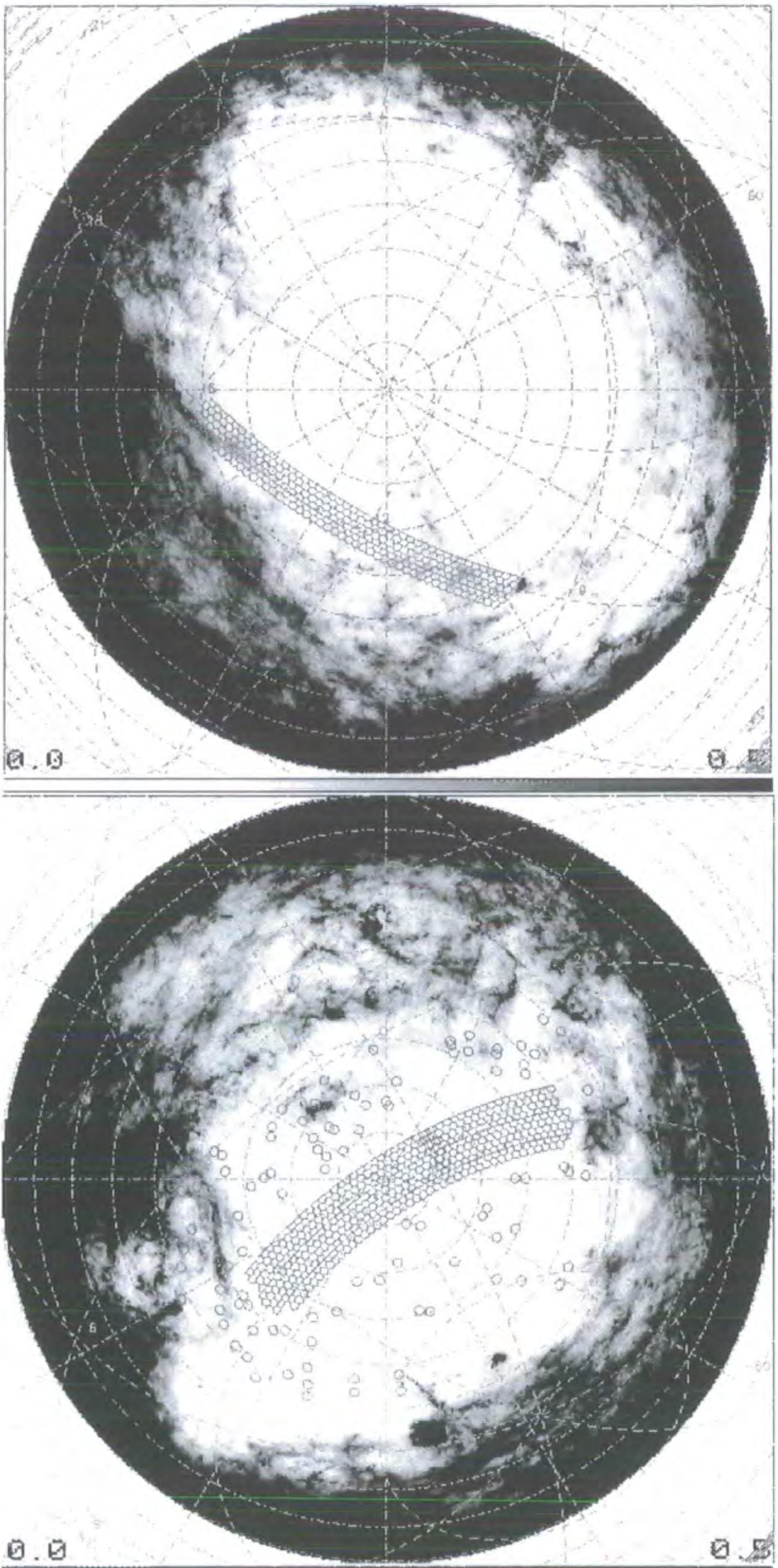


Figure 2.2: The survey fields in the NGP (top) and SGP (bottom) are superimposed on maps of the extinction coefficient A_{b_J} derived from Schlegel et al. (1998). Darker shading indicates regions of greater extinction.

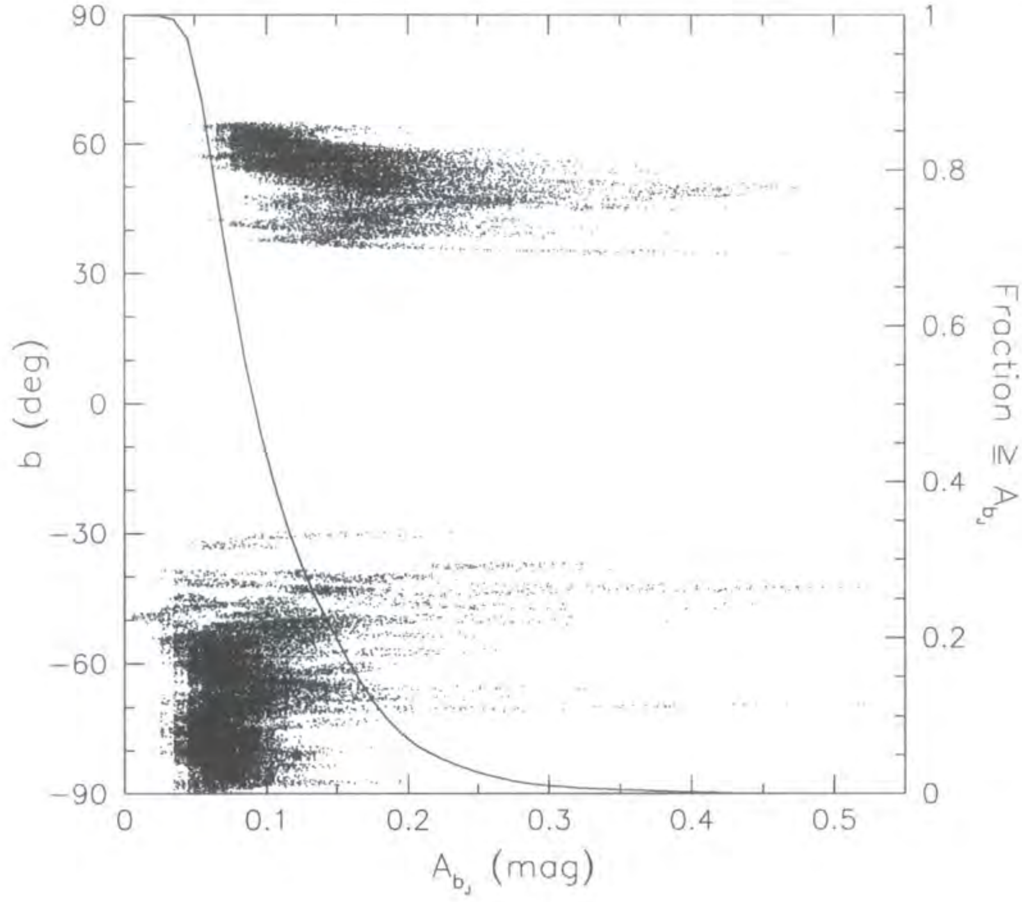


Figure 2.3: The distribution of extinction corrections A_{b_J} with Galactic latitude b (dots and left axis) and the fraction of corrections larger than a given value (line and right axis).

mean extinction over each field is required to be less than 0.2 mag—see Fig. 2 of Efstathiou & Moody (2001)). The fields are chosen pseudo-randomly within this region, with the condition that the field centres are at least 3° apart. A map of the survey fields on the sky is shown in Fig. 2.1; the locations of the fields with respect to the extinction map derived from Schlegel et al. (1998) are shown in Fig. 2.2. All the survey fields lie at Galactic latitudes greater than $|b|=30°$, and the whole of the SGP strip and most of the NGP strip and the random fields lie at Galactic latitudes greater than $|b|=45°$.

The distribution of extinction corrections as a function of Galactic latitude, and the fraction of corrections larger than a given value, are shown in Fig. 2.3. Overall, the median correction is 0.07 mag, 90% are less than 0.16 mag, and 99% are less than 0.26 mag; the corresponding quantiles in the NGP are (0.12,0.19,0.28) mag, in the SGP (0.05,0.07,0.11) mag and in the random fields (0.07,0.13,0.30) mag.

The 2dFGRS target sample of galaxies contains 193 550 galaxies in the SGP strip, 132 111 galaxies in the NGP strip down to an extinction corrected magnitude limit of b_J=19.45, and 57 019 galaxies in the random fields down to b_J=19.50[‡]. This gives a total of 382 680 possible targets, significantly more than the survey goal of 250 000 galaxies. Survey observations of the NGP and SGP strips are proceeding outwards in declination from the centre of each strip towards this goal.

At the median redshift of the survey, the SGP strip extends over $400\,h^{-1}\,\mathrm{Mpc}\times75\,h^{-1}\,\mathrm{Mpc}$, and the NGP strip over $375\,h^{-1}\,\mathrm{Mpc}\times50\,h^{-1}\,\mathrm{Mpc}$. Out to the effective limit of the survey at $z\approx0.3$, the strips contain a volume of $1.2\times10^8\,h^{-3}\,\mathrm{Mpc}^3$(for Ω_m=0.3, Ω_Λ=0.7); the volume sparsely sampled by including the random fields is between two to three times larger.

2.2.1 Tiling the survey

The survey limit of b_J=19.45 was chosen, in part, because it gives a good match between the surface density of galaxies and the surface density of 2dF fibres. Due to clustering, however, the number of galaxies in a given field varies considerably. The rms variation in the number of galaxies per randomly-placed 2° field is 140 at b_J=19.5, and is largely independent of the choice of magnitude limit over the range considered here. To make efficient use of 2dF we therefore require an algorithm for tiling the sky with 2° fields that allows us to cover the survey area at a high, and nearly uniform, sampling rate with the minimum number of 2dF fields.

So long as the sampling of the source catalogue is not biased in any way that depends on the photometric or spectroscopic properties of the galaxies, we can always use the source catalogue to accurately determine the sampling rate as a function of position (see §2.3.3). The sampling can then be accounted for in any analysis. However to keep such corrections to a minimum, considerable effort has been invested in making the sampling as complete and uniform as possible.

There are a number of possible approaches to laying down target field centres. The simplest is to adopt a uniform grid of equally spaced centres and then either randomly sample each field with the number of available fibres or observe each field several times

---

[‡] There is no official reason behind the differing magnitude limits for the main survey strips and the random fields, as this quote from one anonymous 2dFGRS team member reveals: "...the random fields actually go to 19.5 and not 19.45 due to a glitch on someone's part somewhere in the process (no idea where or when!)..." (7.06.2001)

until all the galaxies have been observed. The second of these options is clearly inefficient, as it will give rise to a large number of fields being observed with significantly less than the full complement of fibres, while the first is undesirable as it gives a different sampling factor for each field. A more efficient solution is to use an adaptive tiling strategy, where we allow each field centre to drift from the regular grid so that we maximise the number of targets that are assigned to each field, subject to the constraint that the number of targets assigned to any one field should not exceed the number of available fibres, N_f.

We found that using a fixed separation of 1.5° in declination, and adjusting the tile positions in right ascension only, provided sufficient flexibility to achieve uniform high completeness without a large increase in the total number of fields and without leaving gaps in the sky coverage.

In practice, it is found that the above prescription requires a further modification to account for the position of each object within the field. We apply this constraint by dividing each field into 36 sub-fields and restricting the number of targets that can be assigned to each sub-field to 16. Without this extra constraint the algorithm tends to place large clusters in the overlapping areas of neighbouring fields where, although there are more available fibres because of the overlap, it rapidly becomes impossible to use these fibres because of the high density of targets close to the edge of each field. Limiting the number of targets within each sector effectively removes this problem, and so increases the uniformity of the survey completeness.

The adaptive tiling algorithm also needed to cope with the requirement that the galaxy redshift survey be merged with the concurrent survey of QSO candidates (the 2dF QSO Redshift Survey: Boyle et al. 2000; Croom et al. 2001). This results in a higher surface density of targets in the region of overlap, for which we compensate by reducing the separation in declination of the tiling strips in the QSO survey regions to 75% of the original value (i.e. to 1.125°). As the QSO survey occupies the central declination strip of our SGP survey region this results in a 3-4-3 arrangement of tiling strips over the three rows of UKST fields used in the SGP survey. A similar consideration applied to the NGP survey region gives a 3-4 arrangement if we consider the full areas of the two rows of UKST fields used.

With these additional modifications, our tiling algorithm was able to achieve an overall sampling rate of 99.8%, dropping to 93% after the assignment of fibres to targets has been made (see §2.2.3).

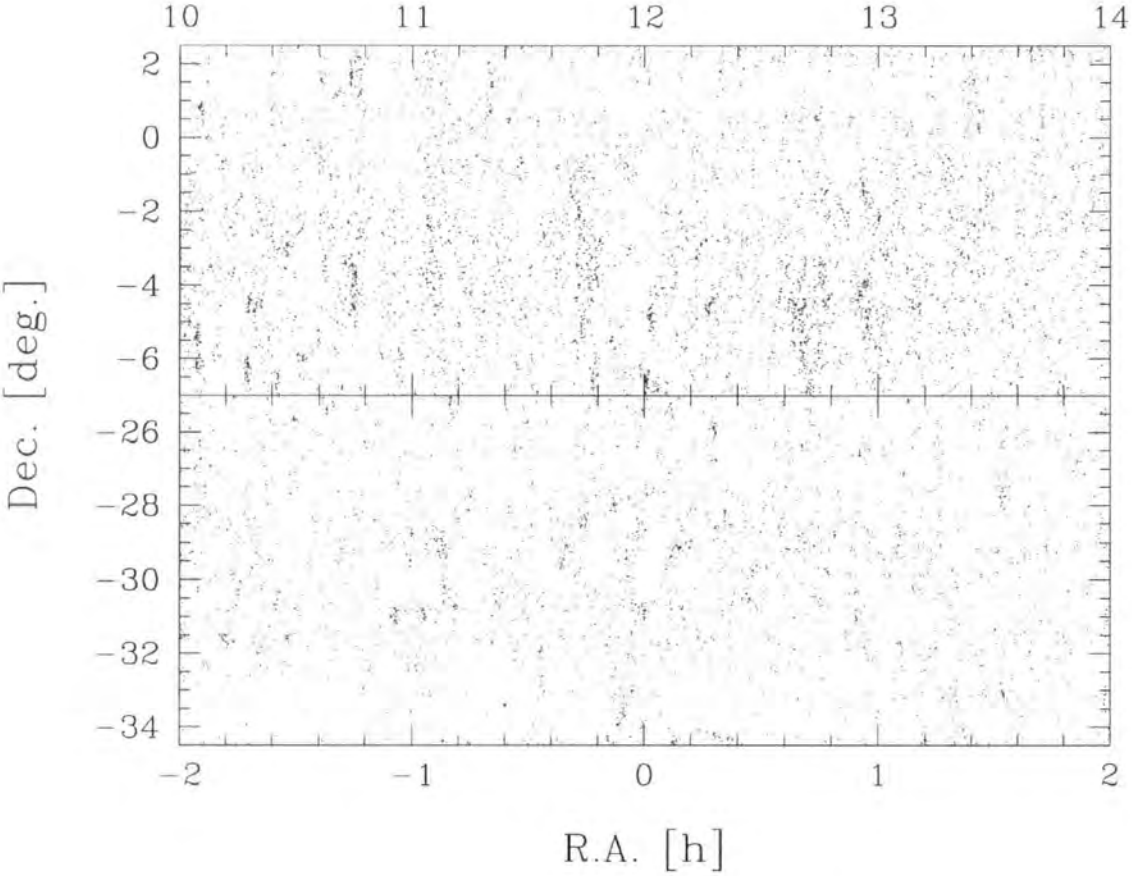


Figure 2.4: The distribution of the 7% of objects in the sample that are in theory not allocated to fibres (NGP strip at top, SGP strip at bottom).

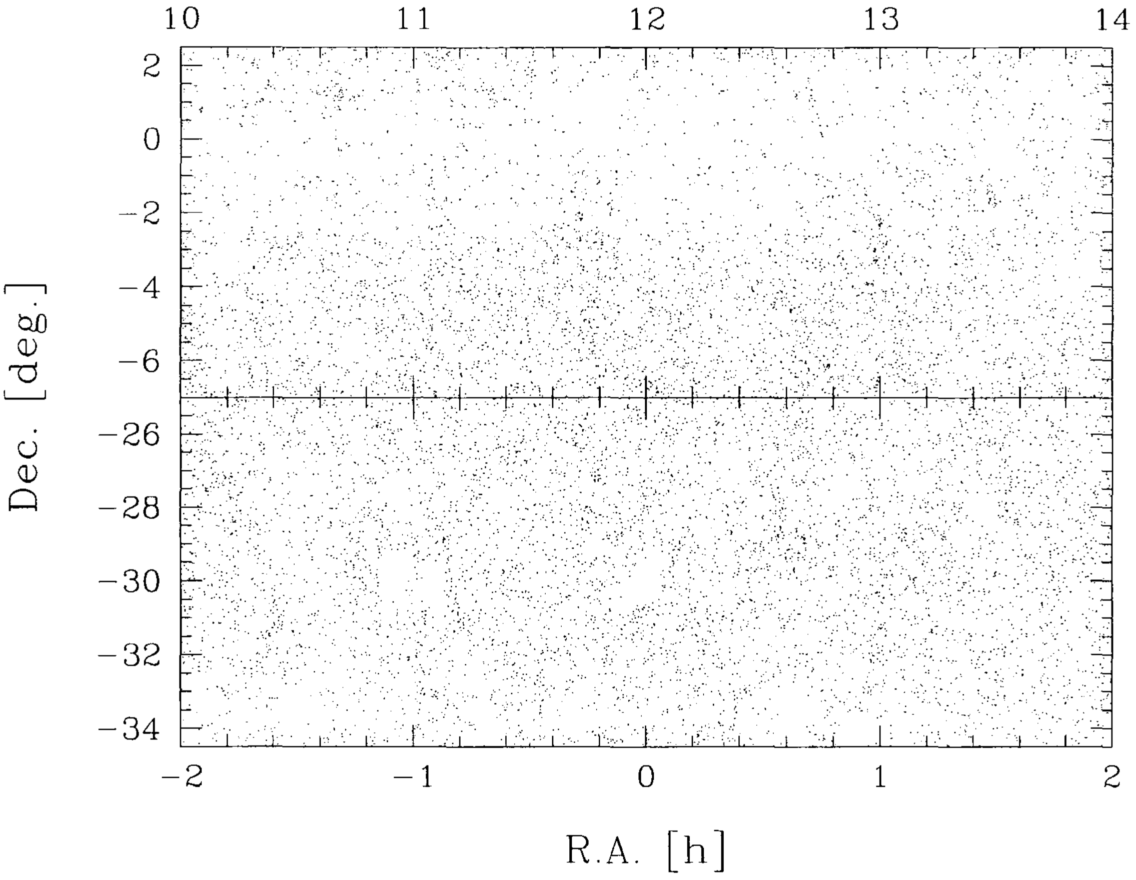


Figure 2.5: The distribution of 7% of randomly chosen objects from the input catalogues. (NGP strip at top, SGP strip at bottom). Fig. 2.4 contains more small scale structures than this random sample of the input catalogue, Nowhere does it look like that the non-targeted objects could represent a dominant fraction of all objects in the input catalogue, which means that the schemes developped in sections 2.3 and 2.5 should be able to describe this incompleteness accurately.

2.2.2 Fibre assignment

The tiling procedure described above fixes the location of the 2dF target fields on the sky and provides a first pass priority scheme for use in assigning targets to fibres. However, the mechanical constraints of 2dF imply that we cannot usually allocate fibres to all the targets assigned to a particular field, and so we consider all targets that lie within each field boundary. Our tiling scheme implies that many of the survey targets are found on more than one field, so we adopt a priority scheme as follows: targets which are unique to the field in question are assigned the highest priority, then targets which are assigned to the field, but which can also be reached from neighbouring fields, then targets which fall within the field, but which are assigned to a different, overlapping field. The priorities of the QSO targets are increased one step to ensure that we do not imprint the strong clustering pattern of the galaxy distribution on the weak clustering expected from the QSO sample.

2.2.3 Fibre assignment completeness

The results of this procedure are that we are able to allocate fibres to 93% of the source catalogue objects. The distribution of the unallocated objects on the sky are shown in Fig. 2.4. The most prominent features visible in these distributions are occasional localised clusters of unallocated objects. These are due to over-dense regions where the geometrical packing constraints imposed by the fibre button dimensions mean that it has not been possible to assign every fibre to a target, even though there are enough fibres available as a result of the tiling algorithm. This effect is enhanced by the relative increase in the number of close pairs in strongly clustered regions. This is clearly demonstrated by comparing Fig. 2.4 with Fig. 2.5 which shows a random distribution of 7% of the objects listed in the input catalogues.

It should be noted that Fig. 2.4 does not represent the final incompleteness of the survey, since individual target allocations are adjusted immediately prior to observation to account for the actual number of fibres available to 2dF on any given night (typically, about 1% of the fibres are broken at any given time, though they are continually replaced). The overall incompleteness at this stage, before any observations have been made, is 7%. The level of this incompleteness is such that the pattern is dominated by residuals due to clustering within individual two-degree fields, implying that an attempt to correct for this effect, by adding some additional fields over regions with 'groups' of untargeted objects,

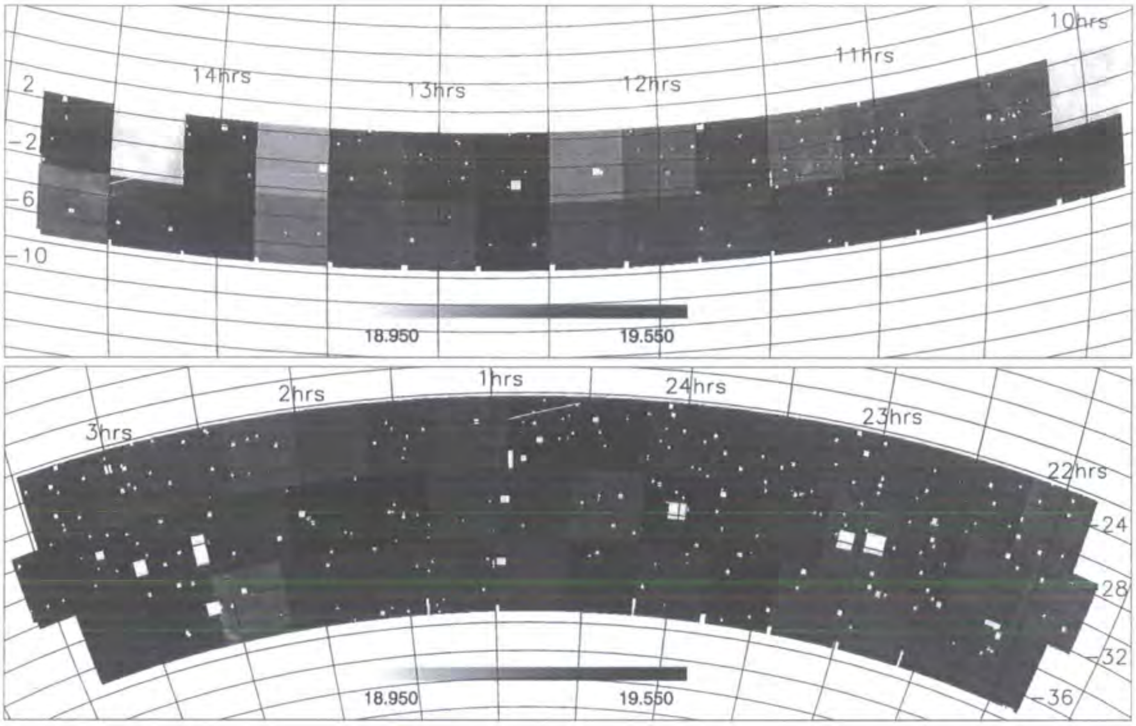


Figure 2.6: The magnitude limit masks for the NGP strip (top) and the SGP strip (bottom), plotted in a zenithal equal area projection.

would effectively imprint the observed pattern onto the data. We therefore conclude that we have achieved our goal of reducing the mean incompleteness to an acceptably low level, comparable to the level of incompleteness due to the individual uncertainties in galaxy magnitudes and the incompleteness due to not measuring redshifts because of inadequate signal-to-noise. We emphasise that neither the tiling nor the fibre allocation depend on the photometric or spectroscopic properties of the objects, so that we can use the source catalogue to determine accurately the sampling of the survey, as discussed in §2.3.3.

2.3 Survey Masks

For accurate statistical analysis of the 2dFGRS it is essential to fully understand the criteria that define its parent photometric catalogue and also the spatial and magnitude-dependent completeness of the redshift catalogue. For this purpose we have defined four maps or masks characterising this information as a function of position on the sky:

(1) The magnitude limit mask, which gives the extinction-corrected magnitude limit of the survey at each position.

(2) The stellar contamination mask, which gives the probability that an object from

the input catalogue is not a galaxy.

(3) The redshift completeness mask, which gives the fraction of measured redshifts at each position.

(4) The magnitude completeness mask, or μ-mask, which gives a parameter defining how the redshift success rate depends on apparent magnitude.

Each mask has its own use, but for some analyses it is necessary to make use of two, three or even all four masks. We now describe in detail how each one of these masks is defined and briefly outline some of their uses.

2.3.1 Magnitude limit mask

Since the targets were selected, improvements to the photometric calibrations and revised extinction corrections (see §2.1.2) have resulted in slight variations to the magnitude limit over the survey regions. The aim of the magnitude limit mask is therefore to return, for each position on the sky, the revised magnitude limit.

Construction of the mask

In order to create a magnitude limit mask, we need to implement, over the whole sky, the same transformation which has been applied to objects in the input catalogue when they were recalibrated. As described in §2.1.2, the recalibration procedure, on non-dust corrected objects, is a linear operation on each UKST photographic plate. Therefore we need to know for each UKST plate the linear transformation used, and then apply correctly the new dust corrections.

Hence for each UKST plate i the following procedure can be applied. Given a set of N objects with two non-dust corrected magnitudes ($b_{J g_{\rm old}}$ and $b_{J g_{\rm new}}$, which respectively corresponds to the old and the new non-dust corrected magnitudes), we calculate by χ^2 minimisation the best fitting values for α_i and β_i:

$$b_{J g_{\rm new}} = \alpha_i \, b_{J g_{\rm old}} + \beta_i \tag{2.2}$$

The relation between dust corrected magnitudes and non-dust corrected ones is simply:

$$b_{J_{\rm new}} = b_{J g_{\rm new}} - A_{b_{J_{\rm new}}}(\theta) \, , \tag{2.3}$$

where the $A_{b_{J_{\rm new}}}(\theta)$ is the dust extinction, which varies across the sky (see for example Fig. 2.2). It becomes thereafter straightforward to relate, on each UKST plate i, the new magnitude limit, $b_{Jlim_{\rm new}}$, to the old magnitude limit, $b_{Jlim_{\rm old}}$:

$$b_{Jlim_{\rm new}} = \alpha_i\, b_{Jlim_{\rm old}} + \beta_i + \alpha_i\, A_{b_{J_{\rm old}}}(\theta) - A_{b_{J_{\rm new}}}(\theta) \qquad (2.4)$$

From Eq. 2.4, it is clear that the magnitude limit is a function of the position on the sky because of the spatial variation of the dust extinction, and because of the different coefficients α_i and β_i that apply for each UKST plate i. It is also worth noting that the relation between $b_{Jlim_{\rm new}}$ and $b_{Jlim_{\rm old}}$ is not as simple as the linear relation given in Eq. 2.2.

Conclusion

The magnitude limit mask is therefore defined by the change in the photometric calibration of each UKST photographic plate and the change in the dust extinction correction at each position on the sky. The magnitude limit masks for the NGP and SGP strips are shown in Figure 2.6; note that the mask also accounts for the holes in the source catalogue around bright stars and plate flaws.

In the SGP, which is a subset of the APM galaxy survey (Maddox et al. 1990a,b,c), the rms change in plate zero-point is only 0.03 mag. However, in the NGP region the original calibration was less accurate and the change in zero-points have an rms of 0.08 mag. The change in the dust corrections are also less in the SGP, as the extinction is generally lower in this region. In the SGP the rms magnitude change due to improved dust corrections is 0.01 mag while in the NGP it is 0.02 mag.

The magnitude limit distribution over the NGP and SGP strips is shown in Figure 2.7. In the SGP the median limiting magnitude is b_J=19.40 with an rms about this value of 0.05 mag; in the NGP the median limiting magnitude is b_J=19.35 with an rms of 0.11 mag.

For accurate statistical analysis of the 2dF survey the magnitude limits defined by this mask should be used. It is always possible to analyse the data with a fixed magnitude limit if one is prepared to omit both the areas of the survey that have magnitude limits brighter than the chosen limit and also all the galaxies in the remaining areas with magnitudes fainter than the chosen limit.

2.3.2 Stellar contamination mask

As well as suffering from a varying magnitude limit over the sky (see §2.3.1), the parent catalogue is also contaminated to some degree by objects which are not galaxies. As a large majority of these objects are stars (see Table 1.1 for explicit numbers), we tend to

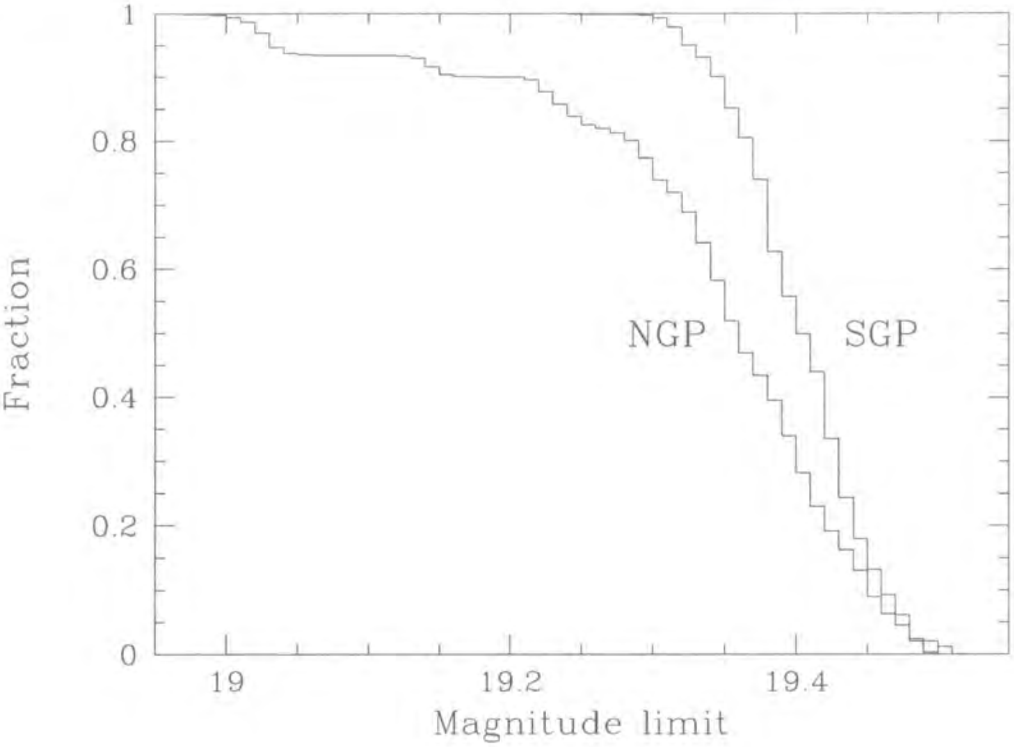


Figure 2.7: The fraction of the sky in the NGP and SGP survey strips where the survey limit is fainter than a given b_J magnitude.

call this overall contamination 'stellar contamination'. For the same reason, we have the tendency to call every object which is not a galaxy a 'star'.

It is expected from the specifications used to disentangle stars from galaxies when creating the 2dFGRS input catalogue that its overall stellar contamination should be about 5%. Indeed as described in Maddox et al. (1990b), the locus dividing stars and galaxies was chosen to exclude as few compact galaxies as possible while keeping the stellar contamination of the galaxy sample as small as possible.

Construction of the mask

The geometrical construction of the stellar contamination mask is identical to that of the redshift completeness mask. We can only estimate the stellar contamination of a region for which we have obtained redshifts. As the stellar contamination mask is a by-product of the redshift completeness mask, we choose to describe the construction of these two masks in §2.3.3.

Once the geometry of the stellar contamination mask is known, we need to find an unbiased estimator of the stellar contamination. For this purpose, it is easier to talk in

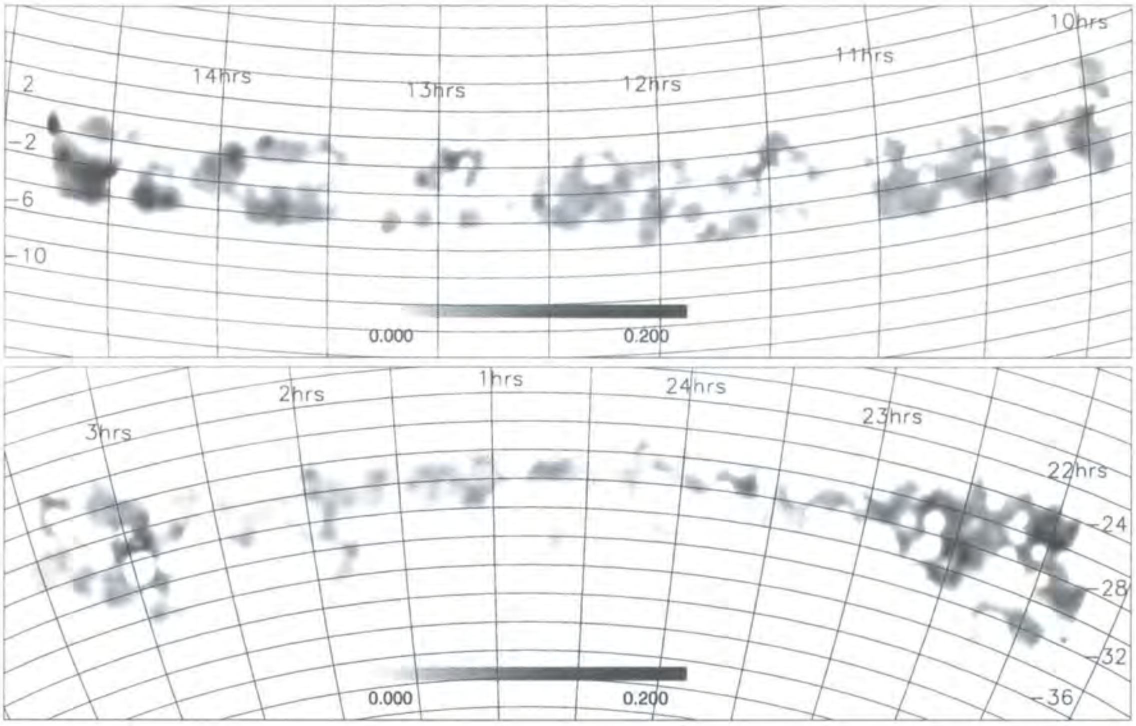


Figure 2.8: The stellar contamination mask as a function of position for the 100k public release dataset (2dFGRS data from October 1997 to February 2000). The top panel is for the NGP, bottom panel for the SGP, plotted in a zenithal equal area projection. The greyscale shows the stellar contamination, ranging from 0. to 20%.

terms of probabilities. The probability of an object in the input catalogue being a galaxy is equivalent to the complementary probability of that same object being a 'star'. In other words, when we talk about the stellar contamination, we talk, in fact, about the probability of an object in the input catalogue being a 'star'. That probability will be dependent, amongst other things, on the object's position on the sky. Let us denote that probability by $P_{\rm star}(\theta)$, where θ is a generalized coordinate on the sky (see §2.3.3 for a more specific definition of θ). The simplest and most robust unbiased estimator of $P_{\rm star}(\theta)$ is:

$$P_{\rm star}(\theta) = \frac{Nz_{\rm star}(\theta)}{Nz_{\rm star}(\theta) + Nz_{\rm gal}(\theta)} \tag{2.5}$$

where $Nz_{\rm star}(\theta)$ is the number of objects in region θ with redshift outside the galaxy redshift range, and $Nz_{\rm gal}(\theta)$ is the number of objects in region θ with redshift within the galaxy redshift range. The redshift limits which define the galaxy redshift range are determined by plotting a histogram of all measured redshifts. From this histogram it is clear that objects with redshift less than 0.002 are very likely to be stars and objects with redshift greater than 0.75 are very likely to be QSOs. Applying these redshift limits, we can plot $P_{\rm star}(\theta)$ for each position θ, which results in the stellar contamination mask shown in Fig. 2.8. Equivalently, one can define the probability of an object being a galaxy, $P_{\rm gal}(\theta)$, as $1 - P_{\rm star}(\theta)$.

The estimator given in Eq. 2.5 for the stellar contamination can suffer from three problems if not dealt with correctly.

First, $P_{\rm star}(\theta)$ can only be robustly measured in regions where the sum of $Nz_{\rm star}(\theta)$ and $Nz_{\rm gal}(\theta)$ is sufficiently large and of the order of $N_{\rm p}(\theta)$, which is the number of objects in the input catalogue in region θ. Indeed, if $N_{\rm p}(\theta)$ is too small, then the Poisson noise of the estimator will be so large that $P_{\rm star}(\theta)$ will not accurately reflect the true stellar contamination. Hence it is important to ensure that $N_{\rm p}(\theta)$ is sufficiently large. In our case, this is done automatically when creating the geometry of the stellar contamination mask, as explained in §2.3.3.

Second, observations done under poor conditions can bias the estimator high, by overrepresenting $Nz_{\rm star}(\theta)$ compared to $Nz_{\rm gal}(\theta)$. Indeed, it can be shown that under poor observing conditions, it is easier statistically to obtain a stellar spectrum of good enough quality to measure a redshift than it is for a galaxy spectrum. This is simply related to the fact that the stars within the 2dFGRS are systematically brighter than the galaxies. In terms of relative numbers, there are more stars than galaxies at magnitudes brighter than $b_J \simeq 18.5$, which in turn implies, as the redshift success is dependent on the apparent magnitude of the targeted object (as shown in more detail in §2.3.4), that poor observing conditions bias the total number of redshifts obtained towards stellar objects. This can, however, be corrected for by excluding any field which was observed under poor conditions. We have implemented this extra condition when constructing the geometry of the mask, as explained in more detail in §2.3.3.

Finally, the stellar contamination can be biased by a misinterpretation of the spectrum, which can happen in cases where the input catalogue suffers from poor astrometry or when the moon is bright. In such cases contamination of the spectra by moonlight together with inaccurate sky subtraction can lead to a well defined spectrum with 'zero' redshift: all that has been detected is the solar spectrum. This will be interpreted in our stellar contamination algorithm as a star, as the redshift is not in the galaxy redshift range, whereas it should be classified as targeted object without redshift. We know that this can happen, as shown in Colless et al. (2001), where we analyze in detail the redshift accuracy. Unfortunately this problem cannot be corrected for, unless we re-examine the spectra of all objects with 'zero' redshift. As this has not yet been done for the whole

2dFGRS redshift catalogue, we cannot yet quantify the effect of this possible bias on our estimate of the stellar contamination.

Conclusion

The stellar contamination mask spatially quantifies the success of the star-galaxy separation, as implemented by Maddox et al. (1990b), and used to construct the input catalogues. Spectroscopic identifications of the survey objects show that the stellar contamination, half-way through the survey, is in fact on average 5.4%, with the following distributions for each region: 5.5% (NGP), 4.9% (SGP) and 6.9% (random fields). In Fig. 2.9 we show the overall distribution of stellar contamination for both NGP and SGP region. It is interesting to see that the NGP region seems to have a slightly different distribution of stellar contamination compared to the SGP region. 20% of both regions are in areas of strong stellar contamination (ie. above 8%). However, only 46% of the NGP has a stellar contamination less than 4%, whereas for the SGP this same threshold corresponds to 62% of the area. The bottom panel of Fig. 2.9 shows that the stellar contamination distribution of the SGP is radically different to that of NGP. The later is mostly dominated by regions of 2% to 6% contamination, whereas the SGP is most dominated by regions with less than 4% contamination. The difference between the two regions might very well be related to the fact that the NGP region is generally closer to the galactic plane than the SGP. This explanation is also consistent with the fact that the random fields are more contaminated than the SGP, as they are on average slightly closer to the galactic plane than the SGP region. Finally it is worth noting that the stellar contamination in the SGP is slightly greater at ends of strips of the survey. This could be due to an increase in the 'true' stellar contamination or possibly poorer observing conditions for these extremities.

2.3.3 Redshift completeness mask

The motivating factors behind the construction of a redshift completeness mask for a survey of the size of the 2dFGRS are obvious. Among the many physically motivated reasons, we have:

- The selection of unbiased samples. Due to the observing strategy used for the 2dFGRS, we require, for any analysis that depends on the spatial coverage of the survey, the ability to select unbiased galaxy samples. This demands that we can

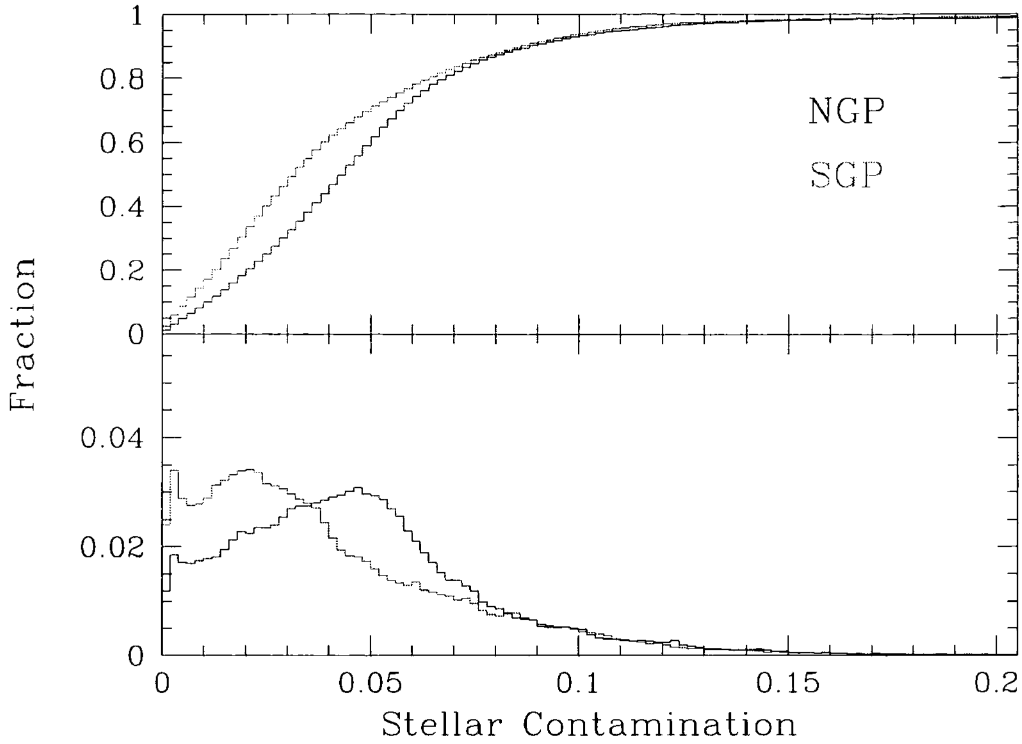


Figure 2.9: The upper panel shows the total fraction of the sky in the NGP and SGP survey strips for which the stellar contamination is less than a given threshold. The lower panel shows the histogram of stellar contaminations, in fractional units.

assess locally the current observing success rate, and also correct for the current incompleteness in regions where it is low.

- A statistical measure of the survey completeness. For all surveys it is important to be able to assess the current status of the survey, and, for example, confirm that the chosen observing strategy is successful.
- The unbiased modelling of the survey selection function. The selection function of the survey is both angular and radial. The inter-play between the angular and radial selection functions needs to be fully modelled.

As the construction of this mask is so important, and rather complicated, we describe its construction in detail. First, we investigate the problem of how to define the geometry of the mask. Second, we focus our attention on how to define the redshift completeness for a given mask geometry. Then we approach, via a simple criteria, the problem of defining a galaxy sample observed under good condition. Finally, we explain how to the three preliminary steps are put together and how one can make optimal use of the redshift completeness mask.

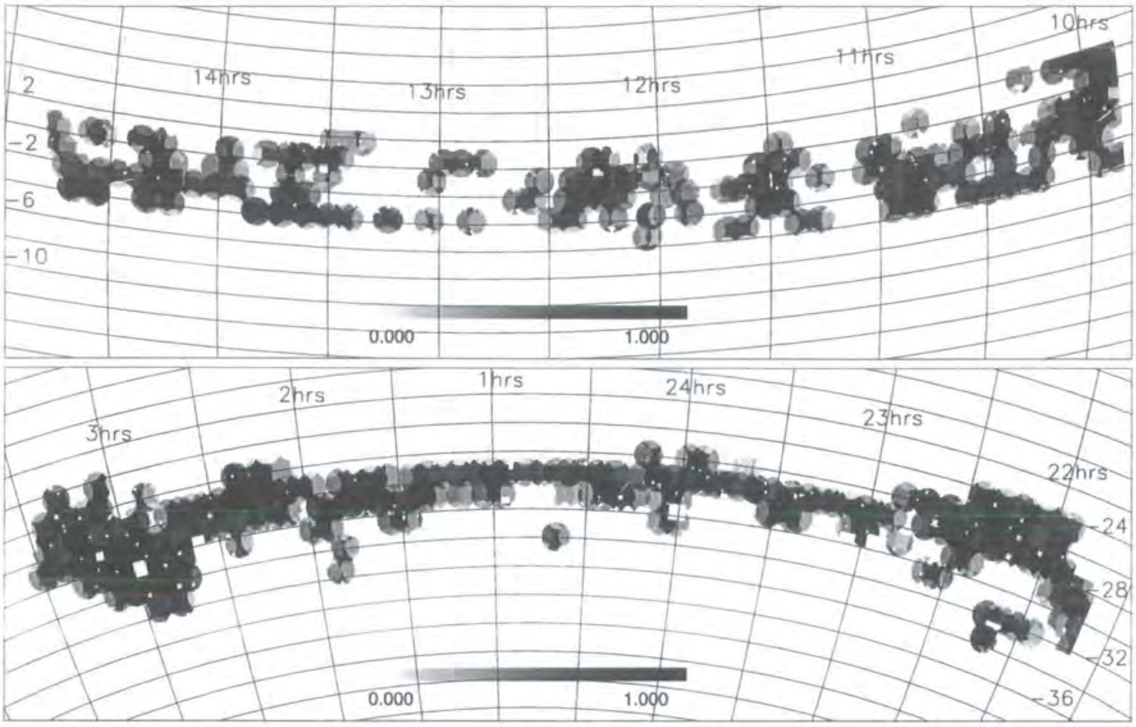


Figure 2.10: The redshift completeness, $R(\theta) \equiv N_z(\theta)/N_{\rm p}(\theta)$ as a function of position for the 100k public release dataset (2dFGRS data from October 1997 to February 2000). The top panel is for the NGP, bottom panel for the SGP, plotted in a zenithal equal area projection.

Optimal mask geometry

The best way to define a redshift completeness mask for the 2dFGRS is to make use of the geometry defined by the complete set of 2°[§] fields that were used to tile the survey region for spectroscopic observations. This way of proceeding takes into account the geometry imposed by the pattern of 2° fields and the way in which the galaxies were targeted for spectroscopic observation. Only regions which are within one or more 2° field can be observed. A large majority of the sky inside the survey boundaries is covered by at least one 2° field, but more often by several overlapping fields. A small fraction of the sky will not be covered by any 2° field: either it corresponds to a region cut out around bright stars or satellite tracks (and in which case there are no target objects in that region) or it corresponds to small isolated regions which were not successfully covered by the tiling (see §2.2.1). We refer to these two types of region as holes (or drill holes) and gap sectors respectively. Finally we define a sector as the region delimited by a unique set of

---

[§] For this section, it is worth knowing that each 2° field is in fact 2.1°in diameter, but the used area is 2.028°in diameter.

overlapping 2° fields. This is obviously the most natural way of partitioning the sky.

While these small sectors accurately define the geometry of the survey, some of them contain too few galaxies to accurately define the completeness. To suppress unwanted Poisson noise in our completeness estimates, it is preferable to consolidate many of these small sectors with their larger, more populated neighbours. The way we have chosen to do this is to append such small sectors to their most populated neighbouring sector.

The parameter N_{consol} indicates the minimum number of objects a consolidated sector has to contain. Hence all sectors with less than N_{consol} objects need to be incorporated into bigger sectors. We typically choose $N_{\text{consol}} = 10$, which is a reasonable compromise between suppressing Poisson noise and artificially erasing the pattern of 2dF fields through which the survey was defined.

By adopting a similar consolidation scheme to gap sectors as for normal sectors, we successfully partition the full sky into the following three types of region:

- consolidated sectors, each of which contain at least N_{consol} objects.
- consolidated gap sectors
- regions 'outside' the survey, comprising also the drill holes within the survey.

Once the mask geometry has been defined, it remains for us to define the redshift completeness for the three types of 2dFGRS regions described above.

Definition of Redshift Completeness

Within each sector[¶], which we can parametrise by a generalized position coordinate θ, we define the redshift completeness, $R(\theta)$, as the ratio of the number of objects for which redshifts have been obtained, $N_z(\theta)$, to the total number of objects contained in the parent catalogue, $N_{\text{p}}(\theta)$:

$$R(\theta) = N_z(\theta)/N_{\text{p}}(\theta)\,. \tag{2.6}$$

This redshift completeness definition does not automatically give the galaxy redshift completeness, $Rg(\theta)$. An unbiased estimator of $Rg(\theta)$ is given by:

$$Rg(\theta) = Nz_{\text{gal}}(\theta)/N_{\text{gal}}(\theta)\,, \tag{2.7}$$

where $N_{\text{gal}}(\theta)$ is the number of galaxies in the input catalogue. As $N_{\text{gal}}(\theta)$ is not known a priori, $Rg(\theta)$ is not trivially estimated. However, by using the approach outlined in §2.3.2, we can create an estimator for $N_{\text{gal}}(\theta)$. An unbiased estimator is given by

$$\widehat{N_{\text{gal}}}(\theta) = P_{\text{gal}}(\theta)\, N_{\text{p}}(\theta)\,, \tag{2.8}$$

[¶] From now on, when talking about (gap) sectors, we in fact mean consolidated (gap) sectors.

where $P_{\rm gal}(\theta)$ is the probability that an object in the input catalogue is a galaxy. Using the expression given in Eq. 2.5 and the relation between $P_{\rm gal}(\theta)$ and $P_{\rm star}(\theta)$, Eq. 2.7 becomes:

$$\begin{aligned} Rg(\theta) &= Nz_{\rm gal}(\theta)/((1 - P_{\rm star}(\theta))\, N_{\rm p}(\theta)) && (2.9)\\ &= Nz_{\rm gal}(\theta)/(Nz_{\rm gal}(\theta)\, N_{\rm p}(\theta)/N_z(\theta)) && (2.10)\\ &= N_z(\theta)/N_{\rm p}(\theta)\,. && (2.11) \end{aligned}$$

Hence, if we use the stellar contamination estimator as defined in §2.3.2, the galaxy redshift completeness, $Rg(\theta)$, is equal to the redshift completeness, $R(\theta)$, as defined in Eq. 2.6. This clearly demonstrates the usefulness of our redshift completeness definition. $R(\theta)$ intrinsically measures the current completeness of the survey with respect to the underlying density. Moreover, $R(\theta)$ is monotonically increasing with time, as clearly demonstrated by the set of completeness figures shown in section 1.3.2 (see Fig. 1.5, Fig. 1.6 and Fig. 1.7).

We must note, however, that our redshift completeness assumes that the true underlying galaxy density is well described by the input catalogue. In other words, if the input catalogue suffers from an unknown incompleteness, our adopted definition of redshift completeness will not be able to correct for it, and will simply inform us about the redshift success relative to the input catalogue.

This redshift completeness definition can be applied directly to sectors and gap sectors. However, it is not advisable to apply blindly this procedure to the full redshift catalogue, as it is then possible to bias the galaxy redshift completeness. Indeed fields observed under poor conditions will not present the same characteristics as fields observed in good conditions. The most obvious example is that poor observing conditions directly imply a lower redshift success rate, but also biased towards brighter objects. Therefore it is essential to characterize the quality of the observing conditions in an objective manner, which we do by defining the field completeness. We note here that it is always possible to go back to the observing reports in order to know the observing conditions, but for samples of hundred thousands of objects this is not feasible, nor is it trivial to make an objective selection based on observing reports.

Definition of Field Completeness

As already mentioned in §§2.3.2 and §§2.3.3, it is important to be able to determine in a statistical sense which galaxies were observed under good observing conditions, instead of

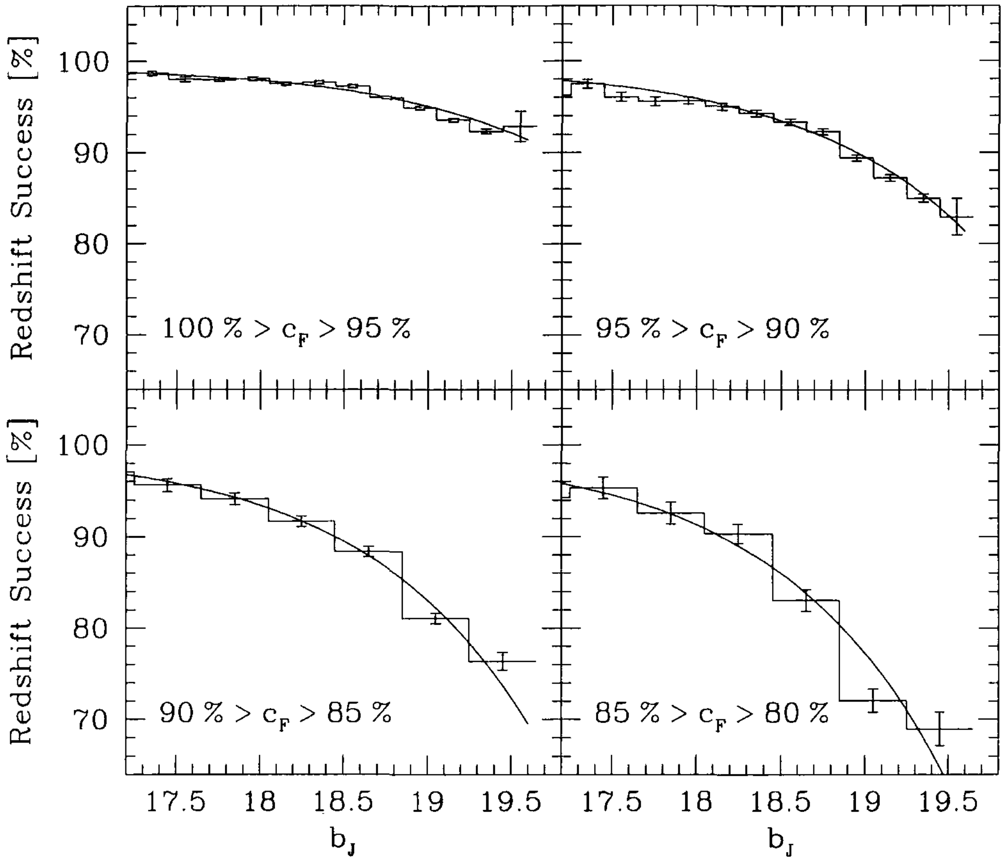


Figure 2.11: Redshift completeness as a function of apparent magnitude. The different panels are for fields of different overall field completeness, as indicate on each panel. Note that over 76% of the observed fields fall in the first two bins, with $c_F > 90\%$. The curves are one-parameter model fits (see §2.3.4 for further details).

automatically using all the galaxies observed to date. Good observing conditions can be defined in various ways, but one easily quantifiable definition is via the field completeness, c_F. For each observed 2dF field, which we can label by its field number n_f, we define the field completeness, $c_F(n_f)$, as the ratio of the number of targeted objects with redshift, $Nz_{tar}(n_f)$, to the total number of targeted objects in that field, $N_{tar}(n_f)$,

$$c_F(n_f) = Nz_{tar}(n_f)/N_{tar}(n_f). \qquad (2.12)$$

When this ratio is high (i.e. above 90%), it means that we have succeeded to obtain good quality spectra for most of the targeted objects, indicating that this observed field should not suffer from any strong bias. This definition of good observing conditions is well justified by the histograms shown in Fig. 2.11. There is a clear correlation between field completeness and good observing conditions. When $c_F(n_f)$ is above 95%, the success rate for measuring redshifts is nearly independent of apparent magnitude and tends to be between 92% and 99%. For fields with 90%<$c_F(n_f)$ <95%, the completeness shows a weak dependence on apparent magnitude, decreasing from 96% at b_J=17.0 to 83% at b_J=19.5. Finally, there is no doubt that fields with $c_F(n_f)$ <90% have a completeness that depends strongly on apparent magnitude, as shown by the bottom two panels of Fig. 2.11. All fields with completeness below 70% are scheduled to be reobserved, and currently 76% of the fields have a completeness greater than 90%. We discuss this magnitude completeness further in §2.3.4.

The redshift completeness of a given sector, $R(\theta)$, should be distinguished clearly from the redshift completeness of a given field, $c_F(n_f)$, as they measure two distinct quantities. First, $R(\theta)$ measures the current redshift completeness of the survey with respect to the input catalogue, whereas $c_F(n_f)$ measures the redshift success of an observed field. Secondly, even if $c_F(n_f)$ is maximal for every observed field, $R(\theta)$ will not necessarily also be maximal, as some objects will never be targeted (see §2.2.3): either as a result of the initial tiling, but also due to missing/dead fibres when observing. Thirdly, $R(\theta)$ contains information from different observing runs, as usually multiple overlapping fields contribute to a single sector, whereas the $c_F(n_f)$ only reflects the observing conditions for a particular run. Finally, $R(\theta)$ is monotonically increasing with time, whereas $c_F(n_f)$ is fixed forever[‖] once a field has been observed.

Using the redshift incompleteness mask

The steps described above are combined to define the redshift completeness mask as follows:

- Create the mask geometry using all the 2° fields planned for the entire survey. View the input catalogue through the mask. Identify and consolidate all small sectors. Identify all gap sectors and consolidate them if necessary.

- Select a good redshift sample by requiring all used fields to have a field completeness above some threshold, typically 75%. Ignore all data acquired in low completeness fields.

- For each sector, calculate the redshift completeness using Eq. 2.6. At the same time, the stellar contamination, given by Eq. 2.5, can be estimated.

[‖] Assuming that the field is not going to be reobserved, in which case the final value of $c_F(n_f)$ is the ratio of the sum of the targeted galaxies with redshift over the sum of the targeted objects within that field.

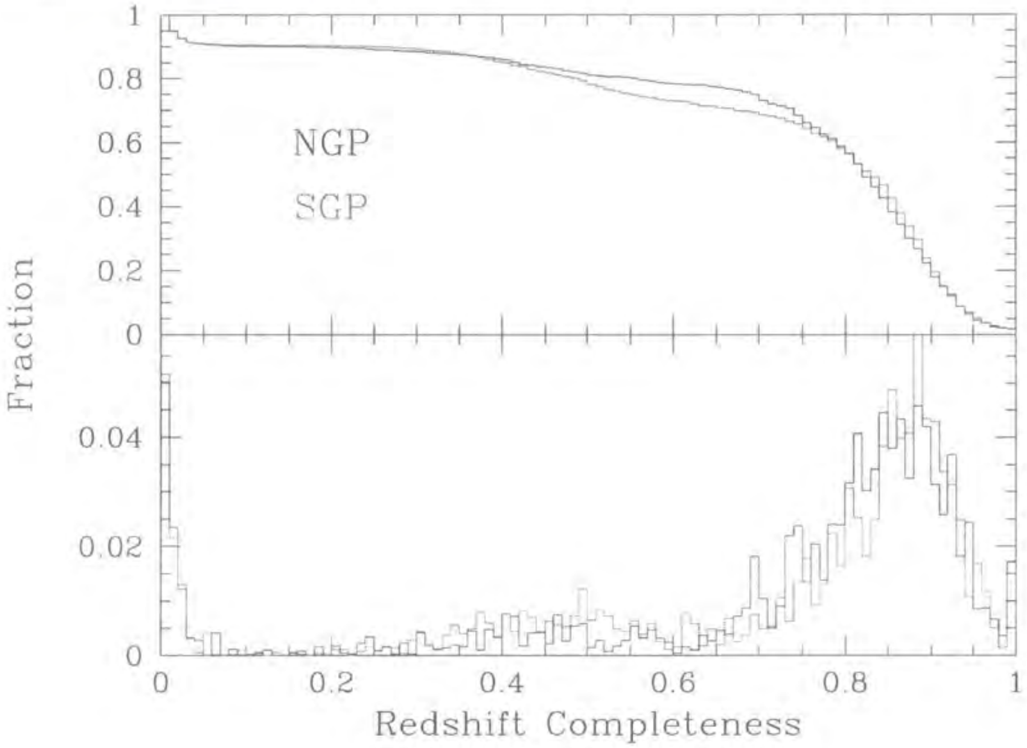


Figure 2.12: The upper panel shows the total fraction of the sky in the NGP and SGP survey strips where the redshift completeness is greater than a given threshold. The lower panel shows the explicit histogram of the redshift completeness.

As a result, a redshift completeness mask (see Fig. 2.10) and a stellar contamination mask (see Fig. 2.8) are obtained.

The redshift completeness mask, like the one shown in Fig. 2.10, can be used to locate regions in which the redshift completeness is high. It can also be used as a first step in either applying weights to statistically correct for incompleteness or to construct a catalogue of random/unclustered points that have the same angular pattern of incompleteness as the redshift sample for use in estimating correlation functions (e.g. see chapters 5, 6 and 7). For the latter purpose one should also take account of how the redshift completeness depends on position within a sector as a result of constraints on fibre positioning and other considerations. This is best done by using the parent catalogue to derive weights for each galaxy with a measured redshift (further details in section 2.5). Also, as discussed in the next section, one should take account of how the redshift completeness depends on apparent magnitude, as we know from Fig. 2.11 that the success of measuring a redshift is dependent upon the object's apparent magnitude.

2.3.4 Magnitude completeness mask or μ-mask

The success rate of measuring redshifts is generally very high, as currently over 76% of the observed fields have a completeness greater than 90%. As one approaches the magnitude limit of the survey it becomes more difficult to obtain good-quality spectra from which reliable redshifts can be measured. Hence the success of measuring redshifts is also a function of apparent magnitude.

In Fig. 2.11 we show the redshift completeness as a function of apparent magnitude for four different intervals of field completeness, c_F. We see that in all cases the completeness is a function of apparent magnitude, and that the magnitude at which the completeness begins to drop is brightest in the fields with lowest c_F, which are generally those taken in marginal observing conditions. Overall, the magnitude-dependent incompleteness is well modelled by a one parameter function of the form

$$c_z(m,\mu) = \gamma(1 - \exp(m - \mu)) \tag{2.13}$$

with $\gamma = 0.99$ and the parameter μ depending on c_F. It can be seen in Fig. 2.11 that this simple one parameter fit in μ describes rather well the magnitude incompleteness for different field completeness values. Our aim is therefore to construct a map which returns for any position on the sky the best estimate of μ and so describes the magnitude dependence of the incompleteness.

In the next two paragraphs we describe the construction of what we refer to as the μ-mask and how it should be used.

Construction of μ-mask

For each field we have chosen to fix μ by combining Eq. 2.13 with a simple power-law model for the galaxy number counts, $N(m) \propto \exp(\alpha m)$, so that c_F and μ are related by

$$c_F(\mu) = \frac{\int_{m_1}^{m_2} N(m) c_z(m,\mu)\, dm}{\int_{m_1}^{m_2} N(m)\, dm}, \tag{2.14}$$

where m_1 and m_2 are the bright and faint magnitude limits. This integral can be evaluated to give

$$c_F(\mu) = \gamma\,[1 - g(\alpha)\, \exp(m_1 - \mu)], \tag{2.15}$$

where

$$g(\alpha) = \frac{\alpha}{\alpha+1} \frac{\exp\left[(\alpha+1)(m_2 - m_1)\right] - 1}{\exp\left[\alpha(m_2 - m_1)\right] - 1} \tag{2.16}$$

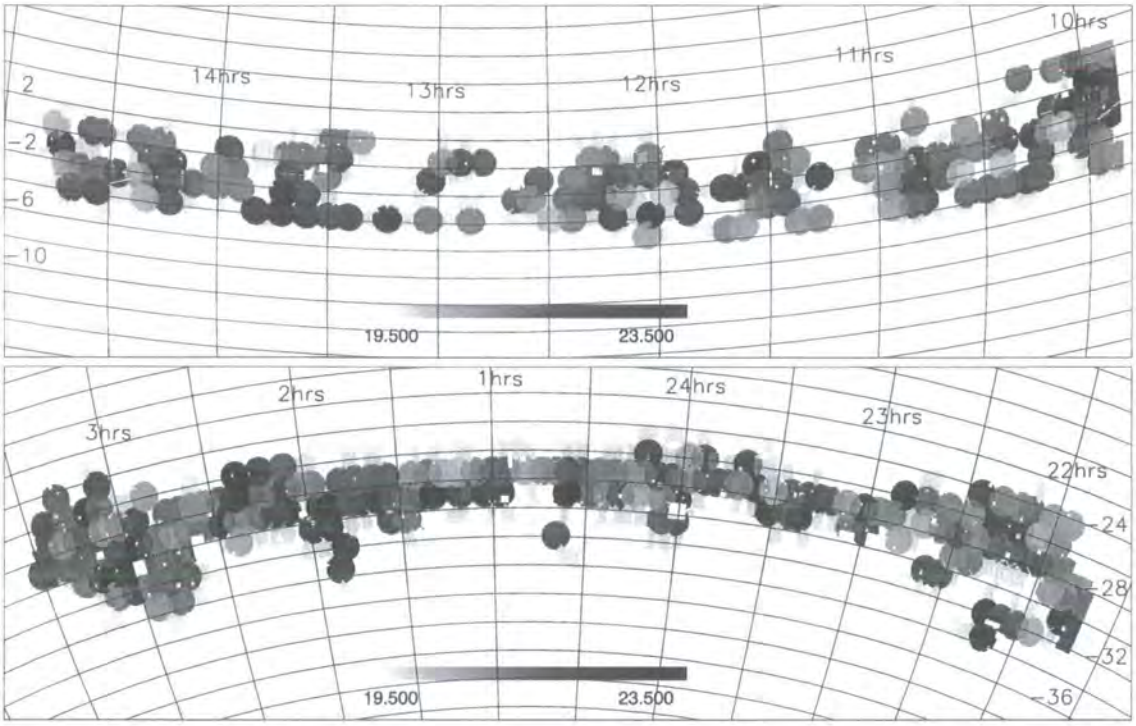


Figure 2.13: The magnitude completeness masks shows the spatial distribution of μ values for the 100k public release dataset (2dFGRS data from October 1997 to February 2000). The top panel is for the NGP, bottom panel for the SGP, plotted in a zenithal equal area projection. The smaller the value of μ, the stronger is the dependence of incompleteness on apparent magnitude. The difference between the geometry of the observed survey in this figure and Fig. 2.10 is due to sector consolidation at the edge of the current survey, which has been cut out from Fig. 2.10. See text for further details.

and we take α=0.55*ln(10.) as appropriate for the galaxy number counts around b_J=19. This equation can be inverted, yielding

$$\mu = \ln\left[\frac{g(\alpha)\ \exp(m_1)}{(1 - c_F/\gamma)}\right], \tag{2.17}$$

which enables μ to be computed for each observed field. This gives us a direct relation between c_F and μ. Note that the smaller the field completeness is, the smaller the value for μ is. Hence a large value for μ is equivalent to a high field completeness, which is equivalent, from Fig. 2.11, to a weak magnitude dependent redshift incompleteness.

Our goal is to define the value of μ characterising the magnitude-dependent completeness for each position in the sky. Since the 2° fields overlap, it again makes sense to define a mean value of μ for each sector, ie. to define μ as function of our generalized coordinate θ. The value of $\mu(\theta)$ is defined by an appropriately-weighted average of the μ values of the

$N_{\rm F}(\theta)$ overlapping observed fields comprising the sector. Specifically we assume that for each sector one can parametrise the magnitude-dependent redshift incompleteness with a function of the form given by Eq. 2.13. As γ is a assumed to be a fixed constant and as the magnitude-dependent redshift incompleteness for each sector, $c_z^s(m, \mu(\theta))$ is the weighted average of the magnitude-dependent redshift incompleteness of the fields which constitute that sector, we can write,

$$c_z^s(m, \mu(\theta)) \;=\; \sum_{i=1}^{N_{\rm F}(\theta)} f_i\, c_z(m, \mu_i)\,, \tag{2.18}$$

where f_i is the fraction of observed galaxies in this sector that were targeted in field i. This can be rearranged to give

$$\begin{aligned} c_z^s(m, \mu(\theta)) &= \gamma \sum_{i=1}^{N_{\rm F}(\theta)} f_i \;-\; \gamma \exp(m) \sum_{i=1}^{N_{\rm F}(\theta)} f_i \exp(-\mu_i) && (2.19) \\ &\equiv \gamma\,(1 \;-\; \exp(m \;-\; \mu(\theta)))\;, && (2.20) \end{aligned}$$

which by identification gives:

$$\mu(\theta) = -\ln\left[\sum_{i=1}^{N_{\rm F}(\theta)} f_i \exp(-\mu_i)\right]. \tag{2.21}$$

Various uses of the μ-mask

We show in Fig. 2.13 the spatially varying magnitude completeness mask or μ-mask. The overall picture is rather good, in the sense that most observed regions have similar μ values between 21.0 and 22.5. From Fig. 2.11, we see that this range corresponds to a field completeness between 93% and 98%. Overall it looks like the NGP has a tendency to be slightly better than the SGP. However, discussing the details of the μ-mask, when shown in this form, is not optimal as one can easily be visually influenced by some striking structure. Therefore, we show the histogram and cumulative distribution of μ values for both the NGP and SGP regions in Fig. 2.14. From this plot we conclude that ~40% (~30%) of the area surveyed so far in the NGP (SGP) has a μ value greater or equal to 22.1 which corresponds to the value used in the fit shown in the top left panel of Fig. 2.11. This tells us that over a third of each region has been targeted in nearly optimal conditions, ie. with a weaker dependence of redshift completeness on apparent magnitude than that shown in the top left panel of Fig. 2.11. Similarly we can see that more than 85% (80%) of the surveyed area is well fitted by a μ curve less steep than the one shown in the top right panel of Fig. 2.11. This tells us directly that the magnitude

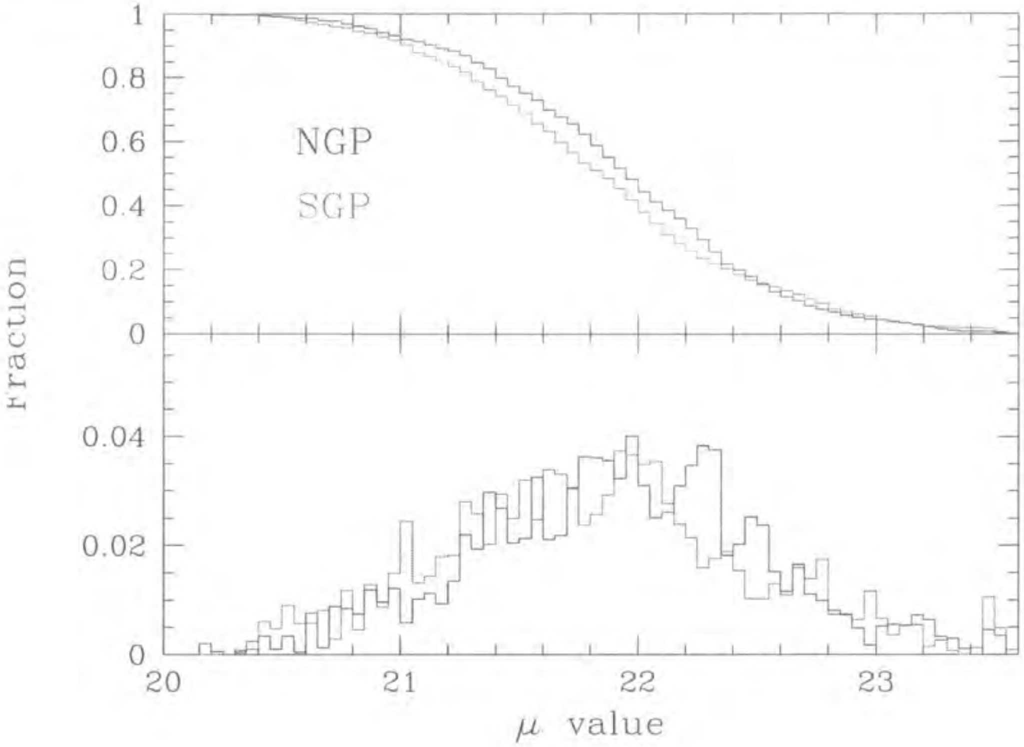


Figure 2.14: The upper panel shows the total fraction of the sky in the NGP and SGP survey strips where the μ value, which corresponds to a certain degree of magnitude incompleteness, is greater than a given threshold. The lower panel shows the histogram of the distribution of μ values. In general, the smaller the value of μ, the stronger the magnitude incompleteness. See text for further details.

dependent redshift completeness is a small concern for the majority of the survey. Indeed a redshift incompleteness at the magnitude limit of the survey less than 15% over more than 80% of the surveyed area is really a rather small incompleteness which can easily be corrected for by using the μ-mask.

Finally, it is now straightforward to combine the magnitude completeness given by $\mu(\theta)$ with the redshift completeness $R(\theta)$ to define an estimate of redshift completeness that depends on both position and magnitude:

$$S(\theta, m) = \frac{N_{\rm p}(\theta)}{N_{\rm e}(\theta)} R(\theta)\, c_z(m, \mu(\theta)). \tag{2.22}$$

In the first factor, $N_{\rm p}(\theta)$ is again the number of parent catalogue galaxies in this sector while

$$N_{\rm e}(\theta) = \sum_{i=1}^{N_p(\theta)} c_z(m_i, \mu(\theta)) \tag{2.23}$$

is an estimate of the number of galaxies one would expect to have measured redshifts for given the value of μ that has been assigned to this sector.

2.4 Redshift Cut dependent Masks

In this section we consider a special case where the redshift completeness mask, described in §2.3.3, needs to be modified in order to properly take into account the problem raised by applying an upper redshift cut on the data analyzed. In §2.4.1, we explain how this new implementation can become necessary, and we describe the modifications done to the redshift completeness mask in §2.4.2.

2.4.1 Motivation behind modified completeness masks

The redshift catalogue masks, developed in §2.3.3 and §2.3.4, assume that one is interested in the completeness analysis of the full survey. However there are cases where using the completeness analysis of the full survey is not appropriate. This is the case encountered when we analyze galaxy clustering as a function spectral type (see chapter 7 for the clustering analysis). In that particular case, it is shown that we can not use the full parent catalogue, as it is only possible to spectrally classify galaxies which are below a given $z_{\rm max}$ threshold, and only then if the spectrum is of high enough signal to noise. Hence the ideal parent catalogue for this case would be a catalogue containing all the sources with redshift less than $z_{\rm max}$. This theoretical complete input catalogue is however impossible to create, and the use of the full catalogue as an approximation is not correct, unless $z_{\rm max}$ is so large that the number of real sources beyond $z_{\rm max}$ is very small.

The way to proceed with this issue is, instead of creating a new parent catalogue, to modify the completeness mask accordingly to incorporate this new upper redshift cut, as explained in more detail in the following section.

2.4.2 Modifications to the redshift completeness mask

Within each sector, we define a $z_{\rm max}$-redshift completeness, $R_{z_{\rm max}}(\theta)$, as the ratio of the number of objects with very good quality spectra** below $z_{\rm max}$, $Nq_{z<z_{\rm max}}(\theta)$, to the total number of objects contained in the parent catalogue with redshift below $z_{\rm max}$,

---

[**] By this, we mean that any further measurement required needs a higher signal to noise spectrum than is necessary to simply measure a redshift. In chapter 7, a very good quality spectrum will be equivalent to being able to successfully spectrally classify a galaxy.

$N_{\rm p(<z_{\rm max})}(\theta)$:

$$R_{z_{\rm max}}(\theta) = Nq_{z<z_{\rm max}}(\theta)/N_{\rm p(<z_{\rm max})}(\theta). \tag{2.24}$$

It is clear that we need to define an unbiased estimator in order to define $N_{\rm p(<z_{\rm max})}(\theta)$. This is best done by assuming that the success of measuring a redshift is independent of the galaxy's intrinsic redshift. We can therefore write:

$$\begin{aligned} R_{z_{\rm max}}(\theta) &= \frac{Nq_{z<z_{\rm max}}(\theta)}{\frac{N_{z<z_{\rm max}}(\theta)}{N_z(\theta)} N_{\rm p}(\theta)} &(2.25)\\ &= \frac{Nq_{z<z_{\rm max}}(\theta)}{N_{z<z_{\rm max}}(\theta)} R(\theta)\,, &(2.26)\end{aligned}$$

where we introduced the standard redshift completeness, $R(\theta)$, as given by Eq. 2.6. In other words, the modified redshift completeness is simply the product of the redshift completeness of the full catalogue and the fraction of very good quality spectra below $z_{\rm max}$. From Eq. 2.26, it is clear that we have implemented a redshift completeness which is essentially independent of the redshift distribution of the objects within a given sector.

For this modified redshift completeness mask, we have explicitly made use of the assumption that this redshift completeness is independent of the redshift of the objects. This assumption is correct as long as the redshift range considered is not too small. In other words, the following relation is correct:

$$\frac{N_z(\theta)}{N_{\rm p}(\theta)} \simeq \frac{N_z(\theta)|_{z_1<z<z_2}}{N_{\rm p}(\theta)|_{z_1<z<z_2}} \tag{2.27}$$

as long as the number of objects between z_1 and z_2 is large, and sufficiently representative of the overall numbers of object along that line of sight. The reason this assumption is still valid as long as the redshift range covered by z_1 and z_2 is large enough relies on the fact that there will statistically be several structures within that redshift range. Hence it becomes more unlikely to overrepresent/underrepresent a single structure. At the same time if there are many objects in the chosen redshift range, the statistical noise should become small, and hence improve the reliability of the measurement.

2.5 Weighting Schemes

The aim of the redshift completeness mask developed in §2.3.3 is to characterize the angular selection function of the survey. This mask does not optimally correct for the incompleteness on the very smallest scales, as the smallest 'building block' of the mask is a set of unique sectors covering the whole sky. As one of the aims when constructing the mask is to have a smooth angular selection function, describing the survey well on larger scales, we have on purpose fine tuned these 'building blocks' to contain at least $N_{\rm consol}$ objects. This indirectly means that the redshift completeness mask is not able to correct for any incompleteness on the very smallest scales, and definitively not on scales containing less than $N_{\rm consol}$ objects. The mean size of a consolidated sector is $\sim 0.4\Box^{\circ}$ and with an average of $\sim$ 164 galaxies per $\Box^{\circ}$ in the input catalogue, this implies that on average each sector contain $\sim$ 65 galaxies. Hence, we can affirm that the redshift completeness mask will typically not be able to correct for incompleteness on scales below $\sim 0.2\Box^{\circ}$, which on average contain $\sim$ 30 galaxies. Therefore there is a real need to create weighting schemes which can compensate for any small scale incompleteness (mostly due to fibre constraints on close pairs). The major difference between these weighting schemes and the masks developed in section 2.3 is typically in their specific use. For example, when we want to carry out a clustering analysis of the survey (as done in chapters 5, 6 and 7), we need to construct a catalogue of random/unclustered points and select an optimal sample. Any random catalogue constructed for the 2dFGRS survey will have its selection function modulated through the use of the different masks in order to match the observed selection function, whereas the real data, in addition to a selection scheme based on the chosen mask thresholds, has to be weighted to account for any systematic incompleteness on the smallest scales, which is extremely difficult to model appropriately in a random catalogue. Indeed, any systematic small scale incompleteness is mostly tiling dependent, so it varies across the sky in a non-trivial fashion.

Therefore in association with the mask defined in §2.3.3, we have designed two weighting schemes in order to correct for any localized (small scale) completeness variations. The first one is based on the redistribution of the weights of all non-observed galaxies to all galaxies with redshift and is described in §2.5.1. The second weighting scheme, detailed in §2.5.2, compensates for the nearest missing galaxies around any galaxy with redshift. We conclude this section with a short description on how to use the stellar contamination mask, described in §2.3.2, in combination with the two other weighting schemes described in this section.

2.5.1 Nearest neighbour weighting in number

A logical way to proceed is to consider all galaxies without measured redshifts within the selected area of the survey and redistribute equally their weights to their nearest galaxies. By nearest we mean nearest on the sky, as this procedure only applies to galaxies for which we have not obtained a redshift, hence no 3-D spatial information is available.

The main parameter of this weighting scheme is N_w, which characterises the number of galaxies to which the weight of the galaxy without redshift is going to be redistributed. We choose to redistribute equally the weight to all neighbouring galaxies; hence each galaxy receives a weight equal to $1/N_w$ which has to be added to the galaxy's own weight. Mathematically this translates into:

$$w_i = 1 + \sum_{j=N_g+1}^{N_p} f_j(\theta_j - \theta_i)\,, \tag{2.28}$$

where N_g is the number of spectroscopically confirmed galaxies, N_p is the total number of objects in the input catalogue and $f_j(\delta\theta)$ is a function equal to $1/N_w$ if $\delta\theta < \theta_j^{lim}$ and zero otherwise. θ_j^{lim} is defined such as to contain exactly the N_w closest neighbours around galaxy j. We assume that all objects are ordered such that 1 to N_g contains only spectroscopically confirmed galaxies and $N_g + 1$ to N_p all other objects. This scheme, by definition, conserves the sum of the weights, which is obviously an attractive feature, especially when it is used in clustering analysis where the sum of the weights is an important parameter for estimating the mean number density of the sample. Naturally we only apply this weighting code to regions which are fairly complete, as otherwise all galaxies with redshifts nearby to any poorly observed region will be systematically overweighted.

We have tested this weighting scheme by changing the value of N_w from 2 to 20, and from these tests we conclude that the best range of values for N_w lies between 8 and 12. Indeed a too large value for N_w is not recommended, as not only is the weighting scheme going to correct for missing objects on too large a scale, for which we already have an adequate incompleteness mapping via the redshift completeness mask, but it will also fail to correct for incompleteness on the smallest scales, which is "la raison d'être" of implementing this additional weighting scheme. On the other hand a too small value for N_w is also undesirable, because the weight is then redistributed too locally and can therefore have the tendency of giving too large a weight to a small number of galaxies. Moreover, as we try here to correct mostly for incompleteness due to missing close pairs (or triplets), we need to bear in mind that the adaptive tiling used allows us very frequently to target close groups with two or more different tiles. Hence redistributing the weight to only 2, 3 or 4 of the nearest neighbours does not take into account this feature of the adaptive tiling, which quite often allows for several attempts to target close groups of galaxies. Finally, assigning a 'large' weight to a galaxy is only acceptable if one knows precisely that it should be that galaxy and not any other which is related to the missing one. As most of the time this information is not known, it is clearly statistically better to adopt a smoother weighting scheme.

As the results of this weighting scheme are fairly insensitive to the choice of values for N_w between 8 and 12 (see panel (b) of Fig. 2.16 for a direct comparison of the influence of the chosen N_w value), we opt for $N_w = 10$. This choice can be compared to that for $N_{\rm consol}$, which is the minimum numbers of galaxies within a consolidated sector, and is also equal to 10.

2.5.2 Nearest neighbour weighting in distance

Instead of opting for a fixed number of objects to which the weight of the missing galaxies should be redistributed, we choose here to weight all galaxies with redshift according to the inverse fraction of galaxies with redshift within a radius r_w to the total number of objects in the parent catalogue within r_w. We directly implemented this scheme with the possibility of choosing a radially dependent weight assignment. In other words, instead of automatically opting for an equal weight assignment within the radius r_w, the weight is assigned dependent on the neighbour's distance away from the galaxy with redshift. We first discuss the equal weight assignment within r_w, and then briefly introduce the more sophisticated radially dependent weighting assignment.

As is the case with the code developed in §2.5.1, the weighting scheme described here should only be applied to fairly complete regions, within which the expectation of finding galaxies with measured redshifts is large compared to the number of galaxies with unknown redshift. This is more important for this weighting scheme as we loop over all galaxies with a successfully measured redshift, and edge effects are definitely more important. Indeed, for a galaxy at the edge of the currently observed survey, typically half of the objects within a radius r_w will not yet have been targeted. Hence there is a tendency to overweight galaxies at the current edge of the survey[††]. The best way of taking that effect into account is to only apply this weighting scheme to fairly complete regions for which neighbouring regions still have a rather large number of galaxies with redshift.

Once the edge effect problem is solved, the algorithm loops over all spectroscopically

---

[††] This edge problem is not as serious for the weighting scheme described in §2.5.1 as in that case we loop over all galaxies with missing redshift, and redistribute their weight to the nearest galaxies with redshift only if they are close enough to the galaxy with missing redshift.

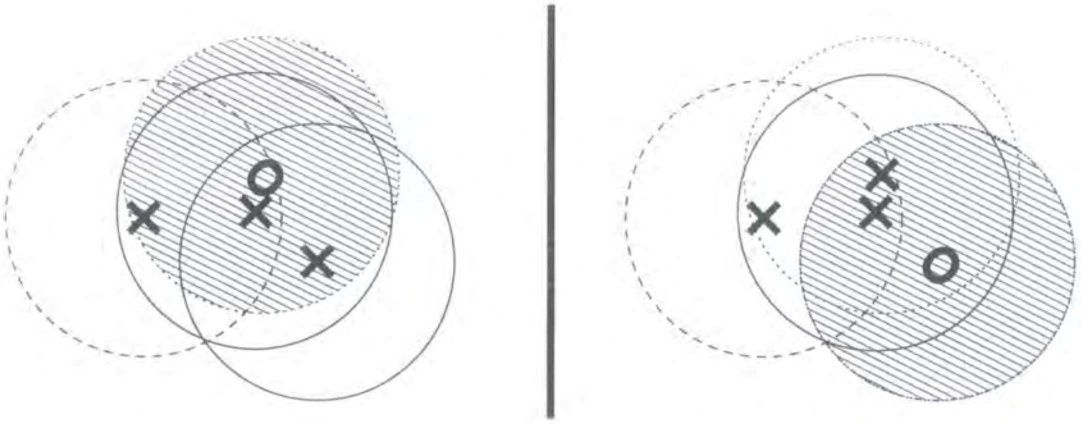
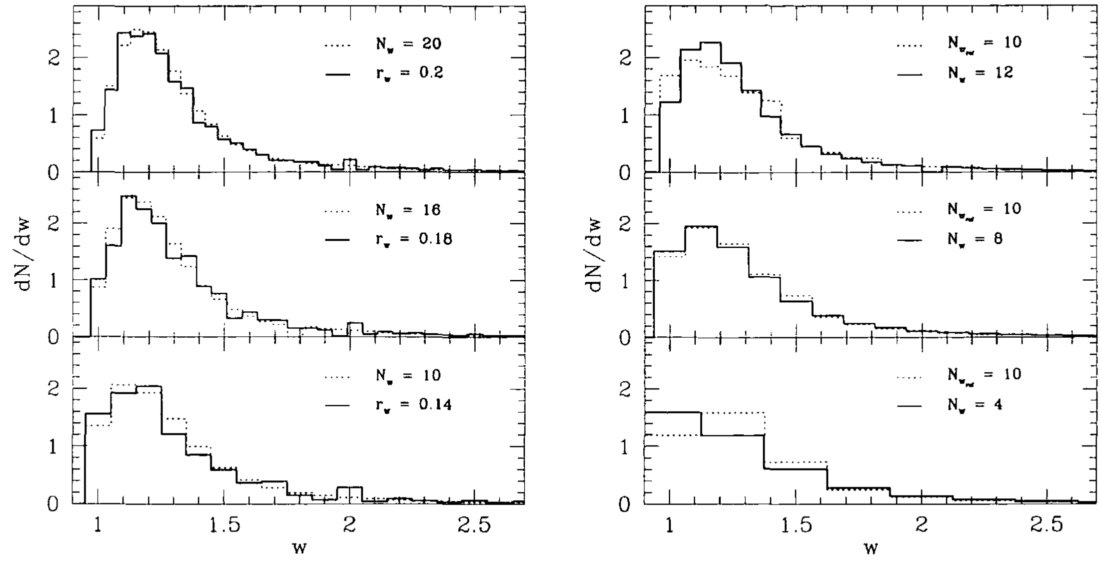


Figure 2.15: We show in the two diagrams how weights are assigned following the nearest neighbour weighting scheme discussed in this section. The three crosses and the circle represent three galaxies with redshift and one without redshift respectively. On each cross/circle, we have centred a circle of radius r_w, with different linetypes for each cross/circle. We highlight the problem which can arise when applying the weighting scheme with a too small search radius, r_w, in an underdense region. In both cases, the overall completeness is 75%, as given by the 3 galaxies with redshift and the one without redshift. The expected correction is such that the sum of the weights in both panels should be equal to 4.0. In the left panel the sum of the weights is 4.33, whereas it is 3.66 in the right hand panel. This problem only arises with the weighting scheme developed in §2.5.2, whereas the weighting scheme which redistributes the weights of unobserved galaxies (see §2.5.1) will clearly not be affected by this issue, as this weighting scheme by definition conserves the sum of the weights.



(a) Comparison of two weighting schemes

(b) Influence of N_w parameter

Figure 2.16: In panel (a), we compare the two weighting schemes outlined in this section, for various values of the weighting scheme parameters N_w and r_w, by showing the distribution of galaxies with a given weight. The values for r_w, in degrees, are tuned such that the mean number of galaxies on the sky within a radius r_w is equal to N_w. The agreement between the two very different weighting schemes presented is remarkable. The only noticeable difference is seen for $w = 2$, with an extra 'bump' in the distribution for the weighting scheme parametrised by r_w. In panel (b), we show the influence of the choice of the value for N_w, each time compared to our adopted value of $N_w = 10$.

confirmed galaxies and assigns a weight given by

$$w_i = w(\theta_i) = \frac{\text{Number of objects within } r_w \text{ of galaxy } i}{\text{Number of galaxies with redshift within } r_w \text{ of galaxy } i} . \tag{2.29}$$

This formula can easily be adapted to a more advanced scheme taking into account the spatial distribution of the surrounding galaxies. Assuming that $f(\Delta\theta)$ is a function describing the relative weight one gives to a neighbouring galaxy at distance $\Delta\theta$ away from a given galaxy (with $f(\Delta\theta) = 0$ when $\Delta\theta \geq r_w$), one can write:

$$w_i = \frac{\sum_{j=1}^{N_p} f(\theta_i - \theta_j)}{\sum_{j=1}^{N_g} f(\theta_i - \theta_j)} , \tag{2.30}$$

where N_g is the number of spectroscopically confirmed galaxies, and N_p the total number of objects in the input catalogue. As for Eq. 2.28, we assume that all objects are ordered.

It is clear already from Eq. 2.29 that this weighting scheme will not conserve the sum of the weights, simply because the weight of a missing galaxy will be taken account of differently depending on whether it is in a high-density or low-density environment. In a low density environment it will be very dependent on the distribution of the neighbouring galaxies and the chosen value for the search radius r_w, as shown explicitly by Fig. 2.15. This is simply due to the fact that this weighting scheme tries to compensate for galaxies with missing redshift by adding a local weight to the galaxies for which redshift have been measured and not trying to redistribute the weight of the galaxies without redshift. This inadequacy of not succeeding in keeping the sum of the weights constant is only readily apparent in regions where the weighting scheme only involves a few objects. Hence in regions of higher density or by choosing a larger value for the radius r_w (e.g. within which the number of galaxies with redshift is large compared to the number of missing galaxies), this apparent problem will not be serious. This is well demonstrated by panel (a) of Fig 2.16, where we compare the distribution of weights obtained from our two different weighting schemes. Each panel shows results for the two implemented weighting schemes, each parametrised respectively by r_w and N_w. We clearly see that both weighting schemes statistically do a similarly good job in assigning weights to neighbouring galaxies of the missed ones. There is just one small difference between the two schemes which can be visually spotted: the weighting scheme described in this section suffers from an overrepresentation of one specific fraction: 2/1. As this trend is there in all three panels (ie. it does not really depend upon the value of search radius r_w), it is most certainly due to the edge problem previously mentioned. Indeed, all these weighting scheme figures have been obtained without using any sector completeness threshold, and so illustrate why a sector completeness threshold is necessary. Typically 50% is an adequate value which barely changes the histograms presented.

Finally we show in Fig. 2.17 that our weighting scheme which redistributes the weights of the missing galaxies to its nereast neighbours correctly takes into account small scale incompleteness. Indeed down to 0.1°, the difference between the angular correlation function of a perfect catalogue and a corrected redshift catalogue is below 0.5%, and down to $1.5'$ this difference is on average below 3%. Similar results are obtained for the weighting scheme presented in this section.

2.5.3 Stellar contamination mask as part of the weighting schemes

In the two previous sections, we have not included the problem of stellar contamination of the input catalogue and how this can affect our proposed weighting schemes. In §2.3.2 we

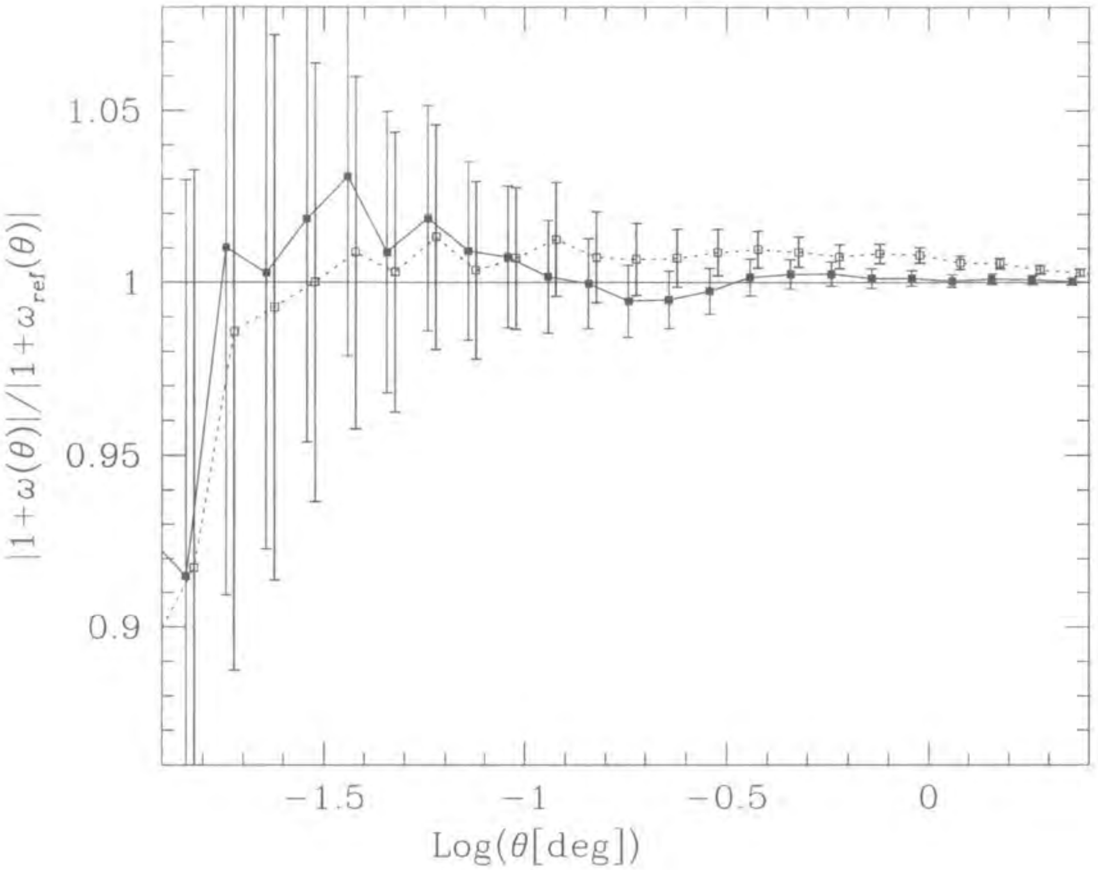


Figure 2.17: We plot the ratio of two angular correlation function estimates as function of the angular separation. The reference sample is, for both curves plotted, the angular correlation function of a region of ~ 80 □°, taken from the input catalogue. The dotted line shows the deviation from the 'true' angular correlation function if one estimates the angular correlation function of galaxies with redshift, without compensating for small scale incompleteness. The solid line shows the same ratio, but this time we used the weighting scheme described in §2.5.1 when estimating the angular correlation function of galaxies with redshift. This test clearly shows the need for a weighting scheme for the smaller scales and that we can now be confident that our results are very robust down to 0.1°, which corresponds, at the median redshift of the 2dFGRS, to a separation of ~ 0.5 h^{-1} Mpc. The errorbars plotted assume Poisson statistics.

pointed out that the parent catalogue of the 2dFGRS suffer to some extent from stellar contamination, which can be precisely assessed for all regions we have surveyed, as shown by the stellar completeness mask presented in Fig. 2.8. The purpose of this section is to include stellar contamination into the two weighting schemes we have designed.

Obviously, there are several ways of including the stellar contamination in our weighting schemes, but the one we propose has the advantage of clearly making the best use of our definition of stellar contamination (as given by Eq. 2.5). For every object of the parent catalogue which belongs to a region we have already targeted, we can define its probability of being a galaxy, denoted P_{gal}. For all spectroscopically confirmed objects, we know the value for this probability: $P_{\mathrm{gal}} = 1$ if it is a galaxy, zero otherwise. For all objects without spectroscopic confirmation, we use the estimator given by Eq. 2.5 and the relation $P_{\mathrm{gal}} = 1 - P_{\mathrm{star}}$ to get an estimate for P_{gal}. This method gives a weight to all objects, which we call the stellar contamination weight or simply the galaxy probability weight.

Once this weight is computed, one just needs to proceed as usual with both weighting schemes, but instead of redistributing/assigning a unit weight for all missed objects, we redistribute/assign a weight, P_{gal}, for them. In other words, instead of Eq. 2.28, the formula for the weighting scheme described in §2.5.1 becomes:

$$w_i = 1 + \sum_{j=N_g+1}^{N_p} f_j(\theta_j - \theta_i)\, P^j_{\mathrm{gal}} = 1 + \sum_{j=N_z+1}^{N} f_j(\theta_j - \theta_i)\, P^j_{\mathrm{gal}} \tag{2.31}$$

with N_g, N_p and $f_j(\Delta\theta)$ having the same definition as in Eq. 2.28, and N_z being the number of spectroscopically confirmed objects (galaxies and 'stars'). Similarly Eq. 2.30 describing the weighting scheme detailed in §2.5.2 becomes:

$$w_i = \frac{\sum_{j=1}^{N_p} f(\theta_i - \theta_j)\, P^j_{\mathrm{gal}}}{\sum_{j=1}^{N_g} f(\theta_i - \theta_j)\, P^j_{\mathrm{gal}}} = 1 + \frac{\sum_{j=N_g+1}^{N_p} f(\theta_i - \theta_j)\, P^j_{\mathrm{gal}}}{\sum_{j=1}^{N_g} f(\theta_i - \theta_j)}, \tag{2.32}$$

where $f(\Delta\theta)$, N_g, and N_p have the same definitions as in Eq. 2.30. The weights obtained from these more sophisticated schemes taking into account stellar contamination of the parent catalogue are extremely similar to the simpler ones, as the individual correction is on average around 6%, which is the mean stellar contamination of the 2dFGRS input catalogue.

2.6 Conclusion

This chapter has shown many very encouraging results of the 2dFGRS survey as well as describing in detail the full complexity of a survey of this size. As a result of the work invested in constructing the sophisticated masks and weighting schemes to take into account the smallest imperfections of the full survey, we are now ready to start quantifying important physical quantities such as the measure of the abundance of galaxies as function of absolute magnitude (luminosity functions, see chapters 3 and 4) or the measure of clustering of different galaxies (see chapters 5, 6 and 7).

Chapter 3

The b_J-band Luminosity Function and Survey Selection Function

We use the 2dF galaxy redshift survey to estimate the b_J-band galaxy luminosity function and present a complete model of the survey selection function, together with the angular selection function described in chapter 2. We estimate the 2dFGRS photometric catalogue to be ~88% complete, after a close comparision with deeper CCD photometry released by the Sloan Digital Sky Survey. Taking into account uncertainties in the photometric calibration, in the adopted $k+e$-correction and in the normalization, we obtain an estimate of the b_J-band galaxy luminosity function which is well fitted by a Schechter function with $M^{\star}_{b_J} - 5\log h = -19.67 \pm 0.07$, $\alpha = -1.21 \pm 0.03$ and $\phi^{\star} = (1.71 \pm 0.08) \times 10^{-2}\,h^3\,\mathrm{Mpc}^{-3}$. This b_J-band galaxy luminosity funcion is in very good agreement with the estimate of Blanton et al. (2001), as long as one corrects properly for colour selection and the same normalisation method is used and identical $k+e$-corrections are applied.

3.1 Introduction

The galaxy luminosity function, which describes the abundance of galaxies as a function of their luminosity, is one of the most fundamental properties of the galaxy distribution. It has come under great scrutiny and our knowledge of it has increased steadily as the size of the redshift surveys used to determine it has grown (e.g. Efstathiou, Ellis & Peterson 1988, Loveday et al. 1992, Zucca et al. 1997, Folkes et al. 1999, Blanton et al. 2001, Madgwick et al. 2001). Here we present an estimate of the b_J-band luminosity function from the 2dF Galaxy Redshift Survey (2dFGRS) which is currently the largest galaxy redshift survey. The luminosity function is an important statistic in its own right and understanding how it arises is a major goal of models of galaxy formation (e.g.

Kauffmann, White & Guiderdoni 1993, Cole et al. 1994, Cole et al. 2000, Somerville & Primack 1999). Also, to fully exploit the 2dFGRS it is important to have a model of the luminosity function so that the selection function of the survey can be computed. This is a vital ingredient in analysing all aspects of galaxy clustering within the survey.

This chapter presents an estimate of the overall b_J-band galaxy luminosity function. This estimate takes into account k-corrections (which result from redshifting of the measured wavelength range) and average evolutionary corrections for galaxies in the sample. We also take account of photometric errors and small corrections for incompleteness in the redshift survey. The analysis here is complementary to that in Madgwick et al. (2001) and the earlier work by Folkes et al. (1999), as there the analysis was only applied to a subset of the 2dFGRS data and focused on how the luminosity function depended on spectral type. Moreover, these papers did not apply evolutionary corrections as they did not attempt to model the full selection function of the survey. We compare and discuss our result in relation to these and other recent determinations of the luminosity function. We also compare estimates for different regions of the survey to test uniformity of the catalogue and our model assumptions. Throughout, we use mock galaxy catalogues constructed from large N-body simulations to check our methods and to assess the influence of large scale structure upon our results. We also use the estimated luminosity function and our modelling of the survey selection limits and completeness to produce a complete (angular, redshift and apparent magnitude) description of the 2dFGRS selection function. The predictions of this selection function are compared with various properties of the real catalogue including the galaxy number counts and redshift distributions.

The outline of this chapter is as follows. In section 3.2 we describe some relevant details of the 2dFGRS, which have not been reviewed so far. We discuss the accuracy of the photometry in §3.2.1, consider various aspects of the completeness of the photometric and redshift catalogues in §3.2.2 and review the accuracy of the redshifts in §3.2.3. Section 3.3 describes how we model the galaxy $k + e$-corrections. In section 3.4 we briefly describe a set of mock catalogues, which we use both to test our implementation of the luminosity function estimators and to assess the effects of large scale structure. We present a series of luminosity function estimates in section 3.5, where we compare results for different regions and subsets of the survey. In section 3.6 we examine the 2dFGRS number counts that we use to normalize our luminosity function estimates and compare them to counts from the SDSS. Our normalized estimate of the 2dFGRS luminosity function is presented in section 3.7. We compare our results with independent luminosity function estimates in section 3.8. In section 3.9 we use our best estimate of the 2dFGRS luminosity function together with the description of the survey magnitude limits and completeness to construct a model of the survey selection function. From this we extract the expected redshift distribution which we compare with those of the real survey and mock catalogues. We summarize our results in section 3.10.

3.2 The 2dF Galaxy Redshift Survey

In this chapter we use the 153 986 redshifts obtained prior to May 2001 in the main NGP and SGP strips. This sample covers a large fraction of the full 2dFGRS area, but as shown in Fig. 1.7, within this area the sampling rate varies with position on the sky, as explained in detail in chapter 2.

For accurate statistical analysis of the 2dFGRS it is essential to fully understand the criteria that define its parent photometric galaxy catalogue and also the spatial and magnitude dependent completeness of the redshift catalogue. Here we complement the description given in chapter 2 and in (Colless et al. 2001), by making a direct comparison of the 2dFGRS catalogue with the overlapping Early Data Release (EDR) of the Sloan Digital Sky Survey (SDSS). The two datasets have approximately 30 000 galaxies in common of which about 10 000 have had redshift measurements made in both surveys. In the following section we use this data to assess the accuracy of the 2dFGRS photometry, the completeness of the parent galaxy catalogue and the accuracy of the redshifts.

3.2.1 Photometric accuracy

The 2dFGRS magnitudes that we use here are the same as those made public in June 2001 "100k Release". They are based on APM scans of photographic plates from the UK Schmidt Telescope (UKST) Southern Sky Survey and their accuracy depends on the accuracy of the zeropoint and non-linearity corrections of each plate as well as the measurement errors within each plate. The plate zero-points and non-linearity corrections are set using a combination of plate overlaps and external CCD photometry and are believed to have an rms error of less than 0.1 magnitudes for galaxies with magnitudes in the range $17 < b_J < 19.5$. As described in Colless et al. (2001) the individual galaxy magnitudes have a larger scatter with a 68% spread of approximately ±0.15 magnitudes.

The top two panels of Fig. 3.1 compare 2dFGRS magnitudes with CCD magnitudes from the ESO Imaging Survey (EIS hereafter Arnouts et al. 2001). The EIS has both B

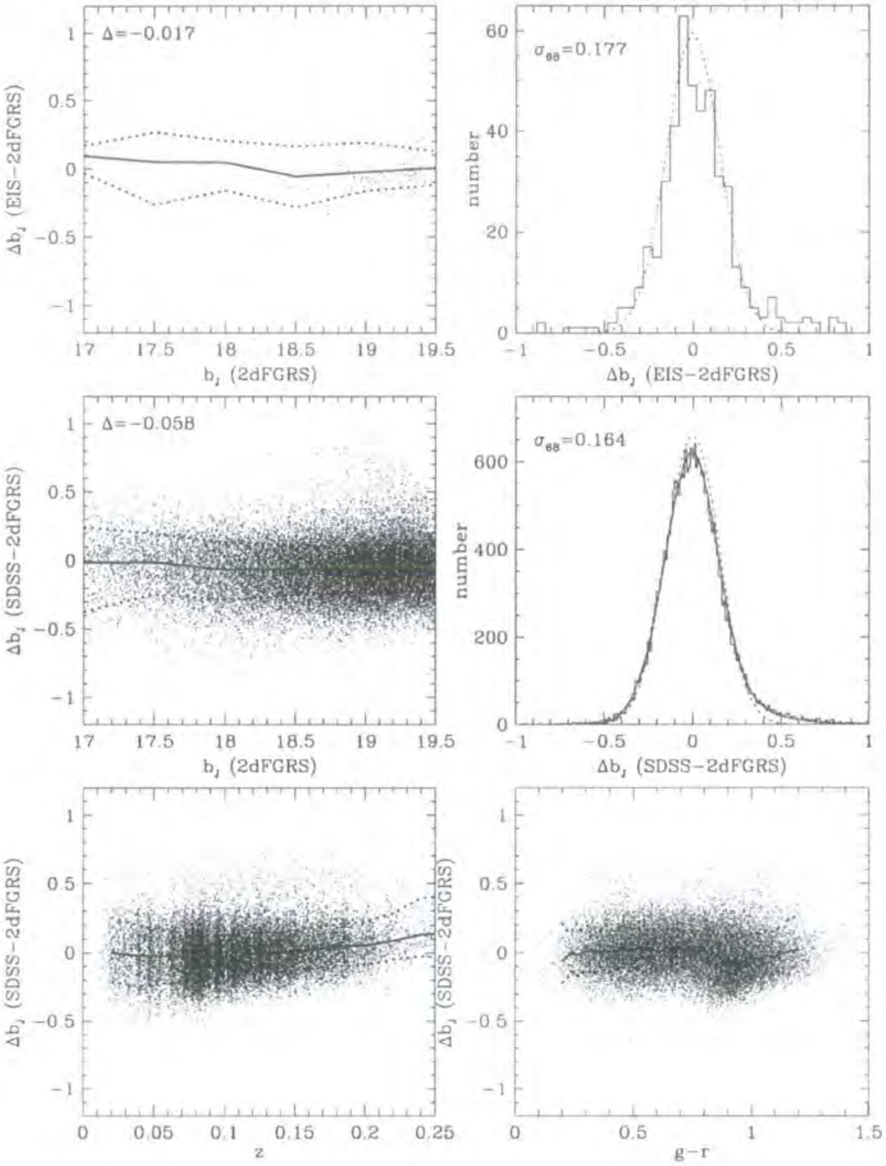


Figure 3.1: Comparison of 2dFGRS photographic b_J magnitudes and CCD magnitudes from EIS and SDSS. The upper panels compare the 2dFGRS with the EIS. The left hand panel is a scatter plot of the magnitude difference versus 2dFGRS magnitude and the solid and dotted lines show the magnitude dependence of the median, 16% and 84% quantiles of the distribution. The median magnitude difference, Δ, for all the galaxies in the range $17 < b_J < 19.5$ is indicated. The distribution of magnitude differences with respect to this median is shown as a histogram in the top right hand panel. The dotted curve is a Gaussian with $\sigma = 0.15$ magnitudes. A robust estimate of the width of this distribution σ_{68}, defined such that $2\sigma_{68}$ spans 68% of the distribution, is also indicated on the panel. The two panels in the middle row show the corresponding distributions for the comparison of the 2dFGRS and SDSS data. The bottom panels show the SDSS 2dFGRS magnitude differences versus redshift (left) and g−rcolour (right). Again the median, 16% and 84% quantiles are shown.

and V-band data and we synthesise b_J using the colour equation $b_J = B - 0.28(B - V)$ (Blair & Gilmore 1982). The coefficient in this colour equation is close to the -0.27 ± 0.02 that we find empirically from the EIS data. The EIS data, which falls on just one UKST plate in the SGP region of the 2dFGRS, forms part of the CCD data that was used to calibrate the 2dFGRS photometry (see §2.1.2). The left hand panel shows that the relation between 2dFGRS and EIS magnitudes is linear and has a small zero-point offset, the median 2dFGRS magnitude being fainter than EIS by just $|\Delta| = 0.017$ magnitudes. The histogram in the right hand panel shows the distribution of magnitude differences after the median offset has been subtracted. The dotted curve which describes the core of the distribution quite well is a Gaussian with $\sigma = 0.15$, but one can see that the measured distribution has small non-Gaussian tails.

The remaining panels of Fig. 3.1 compare 2dFGRS magnitudes with CCD magnitudes from the SDSS EDR (Stoughton et al. 2002). Here we have estimated b_J from the SDSS photometry using the transformation

$$b_J = g + 0.155 + 0.1524\,(g - r). \tag{3.1}$$

This equation comes from adopting the colour equations given for B and V in Fukugita et al. (1996) and combining this with $b_J = B - 0.28(B - V)$ (Blair & Gilmore 1982), as we did above for the EIS data. The bottom right hand panel of Fig. 3.1 is an empirical test of the colour term in our adopted transformation. The very weak dependence of the median magnitude difference on colour is consistent with the $0.1524(g - r)$ colour term and strongly rejects the colour term of $0.088(g - r)$ that was adopted in the comparison made of SDSS and 2dFGRS data made in Blanton et al. (2001).

The middle left hand panel of Fig. 3.1 shows that, in the range $17 < b_J < 19.5$, the relation between 2dFGRS and SDSS magnitude is linear and that the scatter between the two measurements is only very weakly dependent on magnitude. There is a zero-point offset with the median 2dFGRS magnitude being fainter than that of SDSS by $|\Delta| = 0.057$ magnitudes. This is not surprising as the zero-point in the SDSS EDR data is only claimed to be accurate to ± 0.03 magnitudes (Blanton et al. 2001) and the 2dFGRS zero-point has similar uncertainty. The SDSS EDR data spans 15 UKST plates in the NGP region of the 2dFGRS and there is some plate-to-plate variation in the median offset between 2dFGRS and SDSS magnitudes. We find an rms variation of 0.083 magnitudes which is consistent with the 0.1 magnitudes rms specification claimed in Colless et al. (2001) and adds little to the measurement error on an individual galaxy magnitude. We expect the variation in plate zero-points to be somewhat less in the SGP region of the 2dFGRS as this region was constructed from a more homogeneous set of high quality UKST plates than are available in the NGP. At present there is not enough public CCD data to verify this claim. In the other panels of Fig. 3.1 the median offset on each plate has been subtracted from the magnitude differences.

The histogram in the middle right hand panel of Fig. 3.1 shows the distribution of 2dFGRS-SDSS magnitude differences. Again the dotted curve which describes the core of the distribution quite well is a Gaussian with $\sigma = 0.15$. The tail, in excess of this Gaussian, of objects for which the 2dFGRS measures a fainter magnitude than SDSS is very small. There is a somewhat larger tail of objects in which 2dFGRS measures a brighter magnitude than SDSS. It is most likely that these correspond to close pairs of images which SDSS has resolved, but which are merged into a single object in the 2dFGRS catalogue. This distribution is well fitted by the model shown by the solid curve. This model is the sum of a Gaussian and a log-normal distribution. The Gaussian component has $\sigma = 0.14$ and accounts for 70% of the probability and the remaining 30% is distributed as a Gaussian in $\ln(1 + \Delta b_J)$ with $\sigma = 0.235$. We adopt this model as a description of the distribution of the 2dFGRS magnitude measurement errors. A good estimate of the width of the distribution is therefore $\sigma_{68} = 0.164$, defined such that $2\sigma_{68}$ spans 68% of the distribution.

The bottom left hand panel of Fig. 3.1 shows for the subset of galaxies for which redshifts have been measured the magnitude difference as a function of redshift. Below $z \approx 0.16$ there is very little variation in median magnitude difference. At higher redshift there is a weak trend with the 2dFGRS magnitude becoming systematically brighter than the SDSS magnitude. We note that, in contrast, the model of the APM magnitudes constructed in Blanton et al. (2001) falsely predicted that the 2dFGRS magnitude would monotonically become fainter than the SDSS magnitude with increasing redshift. The main reason for the inaccuracy of the Blanton et al. (2001) model is that it neglected to take account of the way in which APM and 2dFGRS magnitudes are calibrated. The calibration of the raw APM magnitudes involves both a zero-point and non-linearity correction so that, in principle, for galaxies in each interval of apparent magnitude the median calibrated 2dFGRS magnitude equals the median total magnitude as set by the calibrating CCD data (Maddox et al. 1990c).

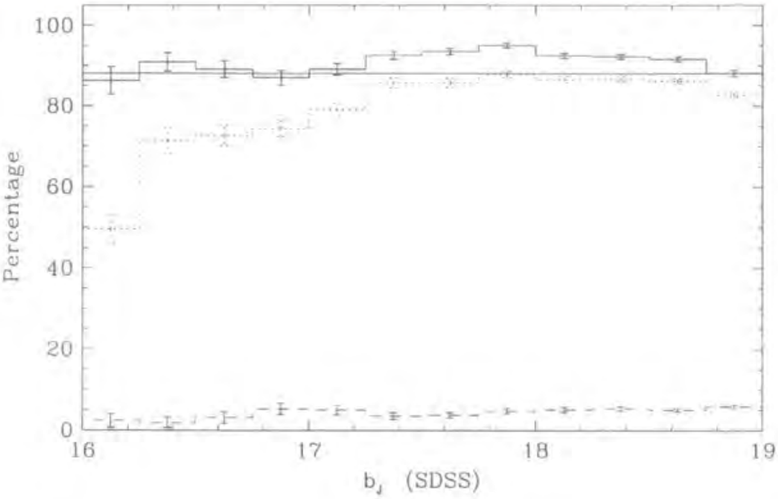


Figure 3.2: The solid histogram shows, as a function of apparent magnitude, the percentage of spectroscopically confirmed galaxies in the SDSS EDR that have 2dFGRS counterparts. The dotted histogram shows the percentage of objects photometrically classified as galaxies in the SDSS EDR. The horizontal line indicates our adopted 88% completeness. The dashed histogram shows the percentage of objects in the 2dFGRS parent catalogue in the same area that are spectroscopically identified as stars.

3.2.2 Completeness of the 2dF parent catalogue

In constructing the parent catalogue of the 2dFGRS, the same parameters and thresholds were used to perform star-galaxy separation as in the original APM galaxy survey (Maddox et al. 1990b). Thus the expectation is that the parent galaxy catalogue will be 90-95% complete and that contamination from stellar objects will be 5-10% (Maddox et al. 1990b). In fact, the spectroscopic identification of the 2dFGRS objects shows that the stellar contamination is 5.4% overall and is only very weakly dependent on apparent magnitude (see Fig. 3.2). The SDSS EDR allows us to make a useful test of the 2dFGRS galaxy completeness. In the SDSS commissioning data, the star-galaxy classification procedure is expected to be more than 99% complete and and therefore to have less than 1% stellar contamination (Blanton et al. 2001). In Fig. 3.2 we assess the completeness of the 2dFGRS parent catalogue both against the SDSS spectroscopic sample and against the SDSS photometric catalogue.

To compare to the SDSS spectroscopic sample, we selected all SDSS objects that are spectroscopically confirmed as galaxies and have magnitudes brighter than $r = 17.6$ and $b_J = 19.4$. The solid histogram in Fig. 3.2 shows, as a function of apparent magnitude, the percentage of these galaxies which have counterparts in the 2dFGRS. The completeness varies very little with magnitude over the entire range $16 < b_J < 19.0$*. The dip in the estimated completeness evident in the faintest bin is an artifact. Here, due to the random measurement errors in the APM/2dFGRS magnitudes, some of the selected SDSS galaxies have APM magnitudes that are too faint to be included in the 2dFGRS parent catalogue. Over the magnitude range $17.3 < b_J < 18.8$ the completeness is between 91% and 95%.

To compare to the SDSS photometric catalogue we selected all SDSS objects that satisfy the star-galaxy classification criteria used by Blanton et al. (2001), i.e. $r_{\rm PSF} - r_{\rm model} > 0.242$. This criterion, which compares an estimate of an object's magnitude made under the assumption that it is a point source, with the magnitude derived from fitting a model galaxy template, is very effective at rejecting faint stars from the sample. At bright magnitudes the sample becomes contaminated by stars. The dotted histogram in Fig. 3.2 shows, as a function of apparent magnitude, the percentage of these objects which have counterparts in the 2dFGRS. Brighter than about $b_J \lesssim 17$ this underestimates the completeness of the 2dFGRS as the SDSS sample has some stellar contamination. Also in the faintest bin the 2dFGRS magnitude limit again causes the completeness to be underestimated. In the intermediate magnitude range $17.3 < b_J < 18.8$, the completeness is between 85% and 88%. This is significantly less than the estimate made from the spectroscopic sample.

If the SDSS spectroscopic sample were a random selection of the photometric catalogue then one would expect the two estimates of incompleteness to agree. However, this is not the case as thresholds have been applied in r-band magnitude and in surface brightness. Also, close pairs of galaxies are under-represented in the SDSS spectroscopic sample due the mechanical limits on how close the optical fibres that feed the spectrograph can be placed. The r-band magnitude limit does not bias our incompleteness estimate as the incompleteness is not colour dependent. However, the other two selection effects will bias the estimate. Comparison of the 2dFGRS parent catalogue with deeper wide-area CCD photometry Pimbblet et al. (2001) and Cross & Driver (2001) has shown that the 2dFGRS can have a tendency to miss low surface brightness galaxies and mis-classifies a fraction of close galaxy pairs. This is a plausible explanation of the difference between the two completeness estimates. We therefore adopt 88 ± 2% as the 2dFGRS galaxy completeness, consistent with the estimate from the SDSS photometric catalogue. This

---

*In the overlap region between SDSS EDR and 2dFGRS NGP data, the magnitude limit of some regions is just fainter than $b_J = 19.1$.

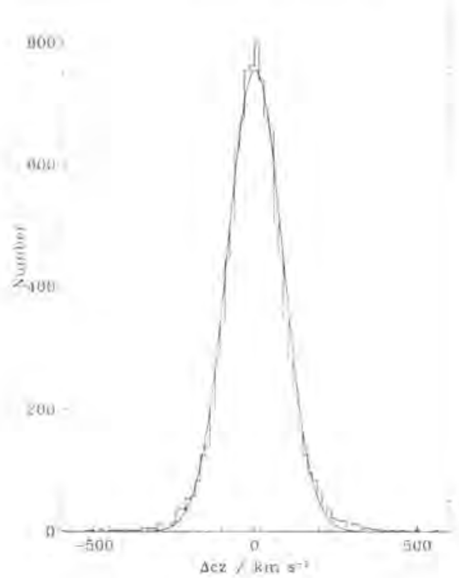


Figure 3.3: A histogram of the 2dFGRS-SDSS redshift differences for a sample of 10 763 galaxies for which both surveys have measured redshifts with $z > 0.003$. The smooth curve is a Gaussian with $\sigma = 85.0$ km s^{-1}.

value is indicated by the horizontal line in Fig. 3.2.

3.2.3 Accuracy and reliability of redshift measurements

The 2dFGRS redshift measurements are all assigned a quality flag Q (Colless et al. 2001). For most purposes only $Q \geq 3$ redshifts are used. From a comparison of repeat observations, Colless et al. (2001) estimated that these have a reliability (percentage of galaxies whose redshifts are within a 600 km s^{-1} tolerance) of 98.4% and and rms accuracy of 85 km s^{-1}. A comparison of the 2dFGRS redshifts with the 10 790 galaxies which also have redshifts measurements in the SDSS EDR provides a useful check of these numbers. The fraction of objects for which the redshifts differ by more than 600 km s^{-1} is only 1.0%. The redshift differences for the remainder are shown in Fig. 3.3. This distribution has a width of $\sigma_{68} = 85.0$ km s^{-1}, and is well fitted by a Gaussian with similar width. The rms error is in excellent agreement with the internal estimate of Colless et al. (2001), while the reliability of 99% is significantly greater. Part of the reason for this difference is that we have only compared redshifts when both the measured SDSS and 2dFGRS redshift are greater than 0.003. This excludes a small number of 2dFGRS redshifts that are very small due either to contamination of the spectra by moonlight or light from a nearby star. If we further reduce the sample to 8 059 objects by excluding objects whose SDSS and

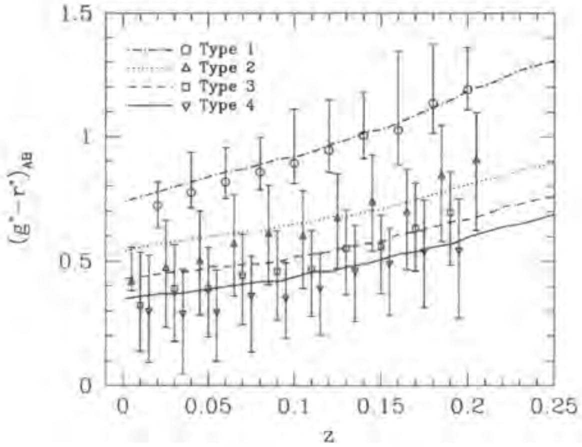


Figure 3.4: Galaxy g − r colours as a function of redshift. The symbols and error bars show for each 2dFGRS spectral type the median, 10 and 90 percentiles of the g − r colour distribution as a function of redshift. The curves are the predictions for model galaxies computed using the Bruzual & Charlot stellar population synthesis code, whose star formation histories have been selected to reproduce the median colour as a function of redshift in each class.

2dFGRS positions differ by more than 1.5 arc second then the reliability increases slightly to 99.22%. This could indicate that some of the discrepant redshifts arise from very close galaxy pairs that are unresolved in the 2dFGRS parent catalogue.

3.3 k+e–corrections

The final ingredient that is required to characterise the selection function of the 2dFGRS is a model describing the change in galaxy magnitude due to redshifting of the b_J-filter bandpass (k-correction) and as a result of galaxy evolution (e-correction). These corrections depend on the galaxy spectra and their star formation histories. As these are correlated, one can parameterise the $k+e$-corrections as functions of the observed spectra.

The 2dFGRS spectra have been classified using a method based on Principal Component Analysis (PCA). A continuous parameter, η, has been defined as a linear combination of the first two principal components (Madgwick et al. 2001). The definition of η is such that it correlates with the strength of absorption/emission features. Galaxies with old stellar populations and strong absorption features have negative values of η, while those with young stellar populations and strong emission lines have positive values. Therefore, we expect the value of η to correlate with the galaxy's k and $k+e$-corrections. In Madg-

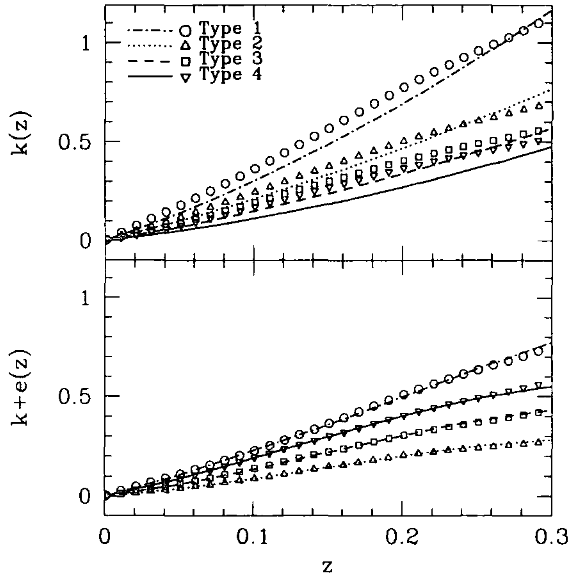


Figure 3.5: Model k and $k+e$-corrections for each 2dFGRS spectral class. The symbols in the top panel show the k-corrections for four models selected to match the g − r colours as a function of redshift shown in Fig. 3.4. The curves show the corresponding k-corrections adopted in Madgwick et al. (2001). The symbols in the lower panel show our model $k + e$-corrections. The smooth curves in this panel are simple analytic fits [Type 1 : $k+e = (2z + 2.8z^2)/(1 + 3.8z^3)$, Type 2: $k+e = (0.6z + 2.8z^2)/(1 + 19.6z^3)$, Type 3: $k+e = (z + 3.6z^2)/(1 + 16.6z^3)$, Type 4: $k+e = (1.6z + 3.2z^2)/(1 + 14.6z^3)$].

wick et al. (2001), the continuous η distribution was divided into four spectral classes (Type 1: $\eta < -1.4$, Type 2: $-1.4 \leq \eta < 1.1$, Type 3: $1.1 \leq \eta < 3.5$, Type 4: $3.5 \leq \eta$) and the mean k-correction for each type estimated from the average spectrum of galaxies in each class. A current weakness of this approach is that the overall system response of the 2dF instrument is not well calibrated. This implies that the resulting k-corrections have a systematic uncertainty of around 10% (Madgwick et al. 2001). Due to this problem and also because we wish to estimate $k + e$-corrections and not just k-corrections, we have taken a complementary approach.

In Fig. 3.4 we have used the SDSS EDR data to plot the median observed g − r colour as a function of redshift for each spectral class. As expected, we see that galaxy colour and its dependence on redshift correlates with the spectral class. Type 1 galaxies with the most negative value of η and oldest stellar populations are reddest and Type 4 bluest. The curves plotted in Fig. 3.4 are models constructed using the Bruzual & Charlot (1993) stellar population synthesis code. In a manner very similar to that described in chapter 4, we ran a grid of models each with the same fixed metallicity and with a star formation history of the form $\psi(t) \propto \exp(-[t(z) - t(z_{\rm f})]/\tau)$ for a range of time scales, τ. Here, $t(z)$ is the age of the universe at redshift z and the galaxy is assumed to start forming stars at $z_{\rm f} = 20$. To relate redshift and time, we assumed a cosmological model with $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ and a Hubble constant of $H_0 = 70\ \mathrm{km\,s^{-1}\ Mpc^{-1}}$. The k and $k+e$-corrections that we derive are only very weakly dependent on these choices. The models plotted Fig. 3.4 are the four which best reproduce the observed dependence of the g − r colours with redshift for the four spectral types. The models provide us with a complete description of the galaxy spectral energy distribution and its evolution with redshift and so can be used to define k or $k+e$-corrections for each spectral Type. These are shown by the symbols in Fig. 3.5. The Madgwick et al. (2001) k-corrections, shown by the curves in top panel, are similar but systematically smaller than those we have derived. This systematic difference is comparable to the systematic difference that is to be expected given the current uncertainty in the calibration of the 2dF instrument, which the Madgwick et al. (2001) k-corrections rely on. The bottom panel of Fig. 3.5 shows our $k+e$-corrections. Simple analytic fits to the $k+e$-correction for each spectral class are given in the figure caption and are shown by the smooth curves. Note that the ordering of the k and $k+e$-corrections are not the same. This is because there are competing effects that contribute to the evolutionary correction. As one increases the redshift, the age of the stellar population one is viewing decreases. This makes galaxies increase in luminosity with increasing redshift, as younger stellar populations have smaller mass-to-light ratios; also the shape of the galaxy spectrum changes. In addition, there are fewer stars present at earlier times and this tends to produces a decrease in luminosity with redshift. For galaxies with ongoing star formation, these effects can all be significant in determining the overall $k+e$-correction.

It is not possible to assign values of η to all galaxies in the 2dFGRS. In fact, only galaxies with $z < 0.2$ can be classified in this way because of the spectral coverage. In practice the classification is only reliable below $z = 0.15$, due to interference of sky-lines with the Hα line. Approximately 10% of those galaxies whose redshift is below 0.15 have spectra with insufficient signal-to-noise to define η. Thus, for some purposes, it is necessary to adopt a mean k or $k+e$-correction that can be applied to all galaxies in the 2dFGRS. In Fig. 3.6, we show k and $k+e$-corrections averaged over the varying mix of galaxies at each redshift and give simple fitting formulae.

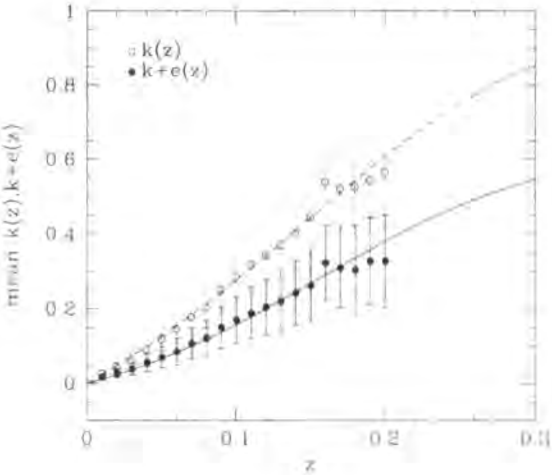


Figure 3.6: The curves show the fits $k(z) = (2.2z + 6z^2)/(1 + 15z^3)$ and $k(z) + e(z) = (z + 6z^2)/(1 + 20z^3)$ to the mean k and $k + e$-correction as a function of redshift. The mean corrections at each redshift, shown by the symbols, have been computed as a function of redshift from the known fractions of each spectroscopic class. The error bars show the rms scatter about the mean of these distributions.

3.4 Mock and Random Catalogues

One of the main purposes of having a quantitative description of the survey selection function is so that one can construct random (unclustered) and mock (clustered) galaxy catalogues. The random catalogues provide a very flexible description of the survey selection function and are most often employed when making estimates of galaxy clustering, as in chapters 6 and 7. The mock catalogues, where the galaxy positions are determined from cosmological N-body simulations, are even more useful. The underlying galaxy clustering and galaxy luminosity function are known for mock catalogues and so they are very useful for testing and developing codes designed to estimate these quantities. They also provide a means of assessing the statistical errors due to realistic large scale structure on any quantity that can be estimated from the genuine redshift survey (e.g. Percival et al. 2001, but also chapters 6 and 7). Here we briefly describe the steps involved in producing the mock catalogues which we use below in sections 3.6 and 3.9 and that are also employed in chapters 4, 6 and 7. These have been created from the very large Hubble Volume simulations of the Virgo consortium (Evrard et al. 2001). For more details of their construction than are given below see Baugh et al. (2001).

The approach we have taken to generate mock and random catalogues that match the selection and sampling of the 2dFGRS has two stages. In the first stage, we generate idealized mock catalogues, which have a uniform magnitude limit (somewhat fainter than that of the true survey) and have no errors in the redshift or magnitude measurements. In the second stage, we have the option of introducing redshift and magnitude measurement errors and we sample the catalogue to take account of both its slightly varying magnitude limit and also of how the completeness of the redshift catalogue depends upon position and apparent magnitude. The steps involved in these two stages are outlined below. In practice, in order to have a fast and efficient algorithm, some steps are combined, but the result is entirely equivalent to this simplified description.

1. To generate the mock catalogues the first step is to sample the mass distribution in the N-body simulation so as to produce a galaxy catalogue with the required clustering. We do this using one of the simple, ad hoc, biasing schemes described in Cole et al. (1998). To be precise we use their Method 2, but with the final density field smoothed with a Gaussian with smoothing length $R_S = 2h^{-1}$Mpc and with the parameters α and β chosen to match the observed galaxy power spectrum. For this we took the APM galaxy power spectrum (Baugh & Efstathiou 1993) scaled up in amplitude by 20% to match the amplitude of clustering measured in the 2dFGRS at its median redshift. This results in a fractional rms fluctuation in the density of galaxies in spheres of 8h^{-1}Mpc of $\sigma_8 = 0.87$.

2. We must also choose the location and orientation of the observer within the simulation. In the mock catalogues used here this is done by applying certain constraints so that the local environment of the observer resembles that of the Local Group (for details see Baugh et al. 2001).

3. We then adopt a Schechter function with $M^{\star}_{\mathrm{b_J}} - 5\log h = -19.67$, $\alpha = -1.21$ and $\phi^{\star} = 1.71 \times 10^{-2} h^3 \mathrm{Mpc}^{-3}$ as a description of the present day galaxy luminosity function. We combine this with the model of the average $k + e$-correction shown in Fig. 3.6 and the adopted faint survey magnitude limit to calculate the expected mean comoving space density of galaxies, $\bar{n}(z)$, as a function of redshift.

4. We now loop over all the galaxies in the simulation cube that fall within the angular boundaries of the survey and randomly select or reject them so as to produce the required mean $\bar{n}(z)$. In the case of random catalogues, we simply generate randomly positioned points within the boundaries of the survey with spatial number density given by $\bar{n}(z)$.

5. For each selected galaxy we generate an apparent magnitude consistent with its redshift, the assumed luminosity function and faint magnitude limit of the survey.

To degrade these ideal mock catalogues to match the current completeness and sampling of the 2dFGRS requires four more steps, which includes the use of the masks developed in section 2.3.

1. We perturb the galaxy redshifts by drawing random velocities from a Gaussian with $\sigma = 85\ \mathrm{km\,s^{-1}}$, which we have seen is a good description of the random measurement errors (see Fig. 3.3).
2. We perturb the galaxy apparent magnitudes, to account for measurement errors, by drawing random magnitude errors from a distribution that accurately fits the histogram of SDSS-2dFGRS magnitude differences shown in Fig. 3.1.
3. We make use of the map of the survey magnitude limit as a function of position to throw out galaxies that are too faint to be included in the 2dFGRS parent catalogue.
4. The final step takes account of the current level of completeness of the 2dFGRS redshift catalogue. Here we make use of the maps $R(\theta)$ and $S(\theta, \mathrm{b_J})$, that are defined in section 2.3. These maps quantify the completeness of the 2dFGRS redshift catalogue. At each angular position, θ, one retains only a fraction, $R(\theta)$, of the redshifts, or taking account of the slight dependence of completeness upon the apparent magnitude one instead retains a fraction $S(\theta, \mathrm{b_J})$, which depends upon apparent magnitude, $\mathrm{b_J}$, as well as position.

3.5 The 2dFGRS Luminosity Function for Different Sub-Samples

The luminosity functions (LF) we present below are estimated using fairly standard implementations of the STY estimator (Sandage, Tammann & Yahil 1979) and the stepwise maximum likelihood estimator (SWML hereafter; Efstathiou, Ellis & Peterson 1988). The only modifications we have made to the methods described in these papers are:

1. We use the map, $\mathrm{b_J}^{\mathrm{lim}}(\theta)$, of the survey magnitude limit to define the apparent magnitude limit for each individual galaxy.

2. We use the map of $\mu(\theta)$ to define a weight, $1/c_z(\mathrm{b_J}, \mu(\theta))$, for each galaxy (see Eq. 2.20) to compensate for the magnitude dependent incompleteness.

A brief outline of these estimators is given in appendix B. Provided that the most incomplete 2dF fields are excluded from our sample, then the variation in these weights is small. The dataset used here only includes fields whose overall redshift completeness is greater than 70%[†]. For this sample the mean weight is 1.06 and the rms variation about this only 0.06. Furthermore one can significantly reduce the influence of the weight by applying an additional magnitude cut and discarding galaxies fainter than, for example, $\mathrm{b_J} = 19.2$.

We have applied both our STY and SWML LF estimators to galaxy samples extracted from the mock galaxy catalogues. In the case of the idealized mock catalogues, not only do the mean estimated luminosity functions agree precisely with the input luminosity function, but also the error estimates agree well with the scatter between the estimates from the different mock catalogues. For the degraded mocks, the estimated luminosity functions reproduce well the input luminosity functions convolved with the assumed magnitude errors. It is perhaps also worth noting that we checked that the independently written STY code used in Madgwick et al. (2001) gave identical results when applied to the same sample and assuming the same k-corrections.

Due to the large size of the 2dFGRS, the statistical errors on our estimated luminosity functions are extremely small. It is therefore important to verify that systematic errors are well controlled. This is partially demonstrated in Fig. 3.7, where we compare luminosity function estimates for various sub-samples of the 2dFGRS.

For all the samples shown in Fig. 3.7 we have applied a bright magnitude cut of $\mathrm{b_J} > 17$ and assumed an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology. In addition, we have applied various other cuts for the different samples. The smooth curve in each panel of Fig. 3.7 is a Schechter function with $M^{\star}_{\mathrm{b_J}} - 5\log h = -19.67$, $\alpha = -1.21$ and $\phi^{\star} = 1.71 \times 10^{-2}\,h^3\,\mathrm{Mpc}^{-3}$. This is the STY estimate for the sample selected by $17 < \mathrm{b_J} < 19.2$ and $z < 0.25$. In both the STY and SWML LF estimates the normalization of the luminosity function is arbitrary. To aid in the comparisons of the estimates shown in Fig. 3.7, we have normalized each of them to produce 152 galaxies per square degree brighter than $\mathrm{b_J} = 19.2$. It can be seen by comparing the SWML estimates in each panel to the STY estimate that the Schechter function is not a good fit at the very bright end. However, it should

---

[†] 76% of the observed fields have an overall field completeness greater than 90%.

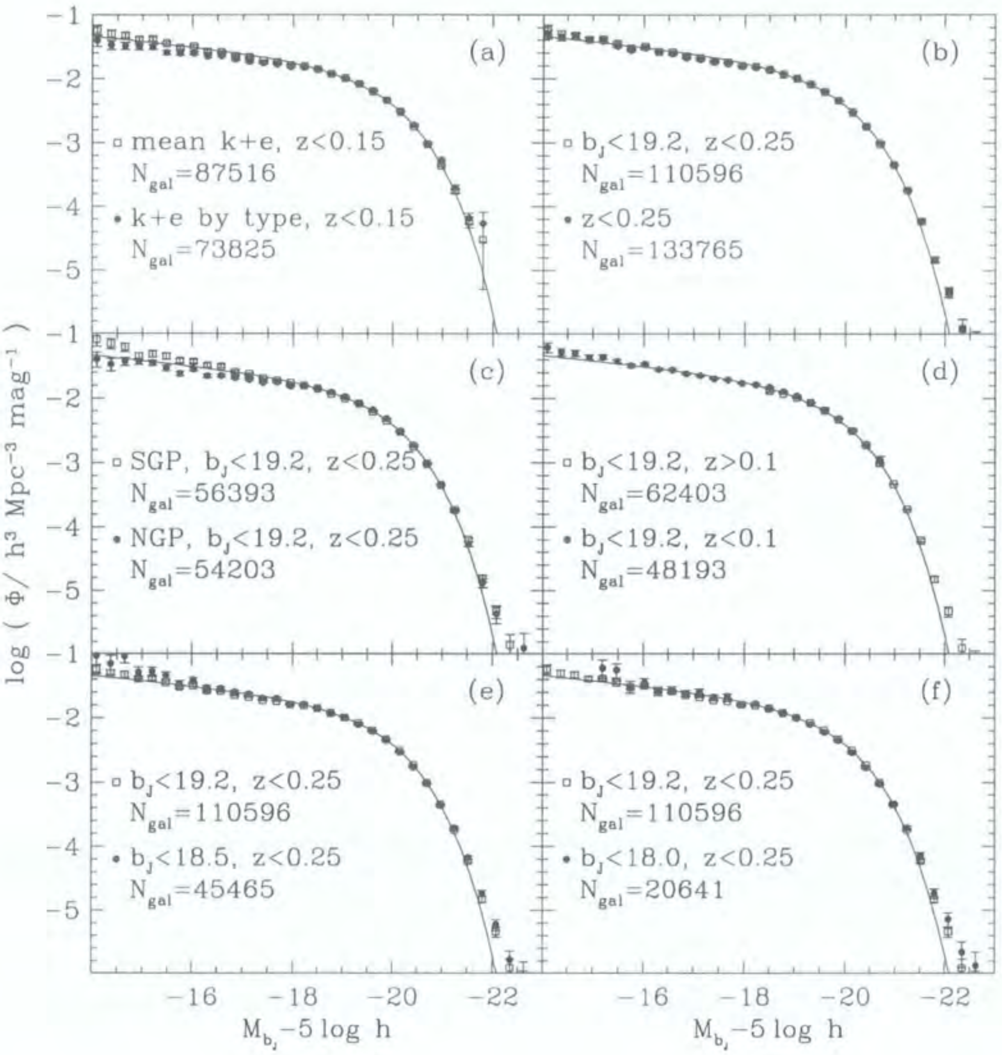


Figure 3.7: Luminosity functions for different sub-samples of the 2dFGRS data. The smooth curve in each panel is a Schechter function with $M^{\star}_{\rm b_J} - 5\log h = -19.67$, $\alpha = -1.21$ and $\phi^{\star} = 1.71 \times 10^{-2}\,h^3\,{\rm Mpc}^{-3}$. This is the STY estimate for the sample selected by $17 < {\rm b_J} < 19.2$ and $z < 0.25$ and computed using the average $k + e$-correction shown in Fig. 3.6. This curve is reproduced in each panel as fiducial reference. In each panel the points and error bars show SWML luminosity function estimates for two different subsets of the 2dFGRS as indicated by the selection criteria given in each legend (see text for details). Also indicated on each panel is the number of galaxies in each sample. An $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology is assumed and the luminosity functions have been normalized to produce 152 galaxies per square degree brighter than ${\rm b_J} = 19.2$.

be borne in mind that in these estimates we have made no attempt to correct for the magnitude measurement errors. Thus these luminosity functions all represent the true luminosity function convolved with the magnitude measurement errors. As we show in section 3.7, this has a more complicated effect on the Schechter function fit, which is mostly determined by galaxies around $M^{\star}_{\mathrm{b_J}}$.

The influence of the assumed $k + e$-correction is investigated in Fig. 3.7(a). Both samples are defined by the limits $17 < \mathrm{b_J} < 19.2$ and $z < 0.15$. For one sample we use the average k+e correction shown in Fig. 3.6, while for the other we adopt the spectral class dependent $k + e$-corrections of Fig. 3.5. The sample for which we apply the class dependent $k + e$-corrections is slightly smaller as a small fraction (5%) of the spectra have an insufficient signal-to-noise ratio to enable them to be spectrally classified and in addition, at this stage, not all the spectra have been processed. Note that the upper redshift limit is imposed to avoid the interval where contamination by sky lines causes the spectral classification to be unreliable (Madgwick et al. 2001). We see that from 2.5 magnitudes fainter than $M^{\star}_{\mathrm{b_J}}$ to the brightest magnitudes probed there is essentially no difference between the two luminosity function estimates. Only for magnitudes fainter than $M_{\mathrm{b_J}} - 5\log h \simeq -17$ does the class dependent estimate fall slightly below the estimate that assumes a global $k + e$-correction and then only by an amount comparable to the statistical errors. As this systematic error is so small we adopt for all other estimates the global k+e correction which then allows us to use the full redshift sample.

Fig. 3.7(b) shows SWML estimates for samples including galaxies with redshifts up to $z = 0.25$. The two estimates shown compare the results for the sample limited by $\mathrm{b_J} < 19.2$ and the sample to the full depth of the 2dFGRS, which has a spatially varying magnitude limit of 19.4 ± 0.1, as shown by Fig. 2.6. The close agreement of these two estimates indicates that no significant bias or error has been introduced when taking account of the varying magnitude limit and the corrections for the magnitude dependent incompleteness.

Most of the remaining panels of Fig. 3.7 use samples limited by $\mathrm{b_J} < 19.2$, but essentially identical results are found if the samples are extended to the full depth of the survey. Fig. 3.7(c) compares the luminosity function estimates from the spatially separated SGP and NGP regions of the 2dFGRS. Brighter than $M_{\mathrm{b_J}} - 5\log h \simeq -17.5$ the two regions yield luminosity functions with identical shapes. Note that both luminosity functions have been normalized to produce 152 galaxies per square degree brighter than $\mathrm{b_J} = 19.2$ rather than to the actual galaxy number counts in each region. This good agreement suggests that any systematic offset in zero point of the magnitude scale in the two disjoint regions is very small. If one allows an offset between the zero points of the NGP and SGP magnitudes scales then comparing these two luminosity functions constrains this offset to the rather small value of 0.02 ± 0.007. Fainter than $M_{\rm b_J} - 5\log h \simeq -17.5$ the two estimates differ systematically by an amount comparable to the estimated statistical errors. We return to this briefly in Section 3.7.

Fig. 3.7(d) compares results from samples split by redshift. Here the combined effect of the redshift and apparent magnitude limits result in estimates that only span a limited range in absolute magnitude. To normalize these luminosity functions we extrapolated the estimates using their corresponding STY Schechter function estimates. The estimates from the two samples agree well in the overlapping magnitude range and also agree well with the full samples shown the other panels.

The final two panels in Fig. 3.7 look at luminosity function estimates from bright subsamples of the 2dFGRS. Fig. 3.7(e) shows an estimate for galaxies brighter than $\rm b_J = 18.5$ and Fig. 3.7(f) for galaxies brighter than $\rm b_J = 18.0$. The statistical errors in the estimates from these smaller samples are significantly larger. Nevertheless, the estimates on average agree well with those from the deeper samples.

3.6 Galaxy Number Counts

In the previous section we have demonstrated the that shape of the 2dF galaxy luminosity function is robust to variations in the sample selection and assumed $k+e$-corrections. We have not addressed the issue of uncertainty in its normalization, as we simply normalized all the estimates to produce 152 galaxies per square degree brighter than $\rm b_J = 19.2$. We now investigate the uncertainty in this normalization due both to large scale structure and the uncertainty in systematic corrections.

The upper panel in Fig. 3.8 shows the 2dFGRS galaxy $\rm b_J$-band number counts in both the NGP and SGP. In this figure we have subtracted a Euclidean model from the counts to expand the ordinate so that small differences in the counts are visible. These counts are counts of objects in the 2dFGRS parent catalogue (after the removal of the merged images that did not form part of the 2dFGRS target list) multiplied by a factor of 1.078 to take account of the stellar contamination (5.4%) and incompleteness (12%), discussed in §3.2.2. While these numbers are derived from comparison with the SDSS EDR we note that they are very comparable to the original estimates given in Maddox et al. (1990b).

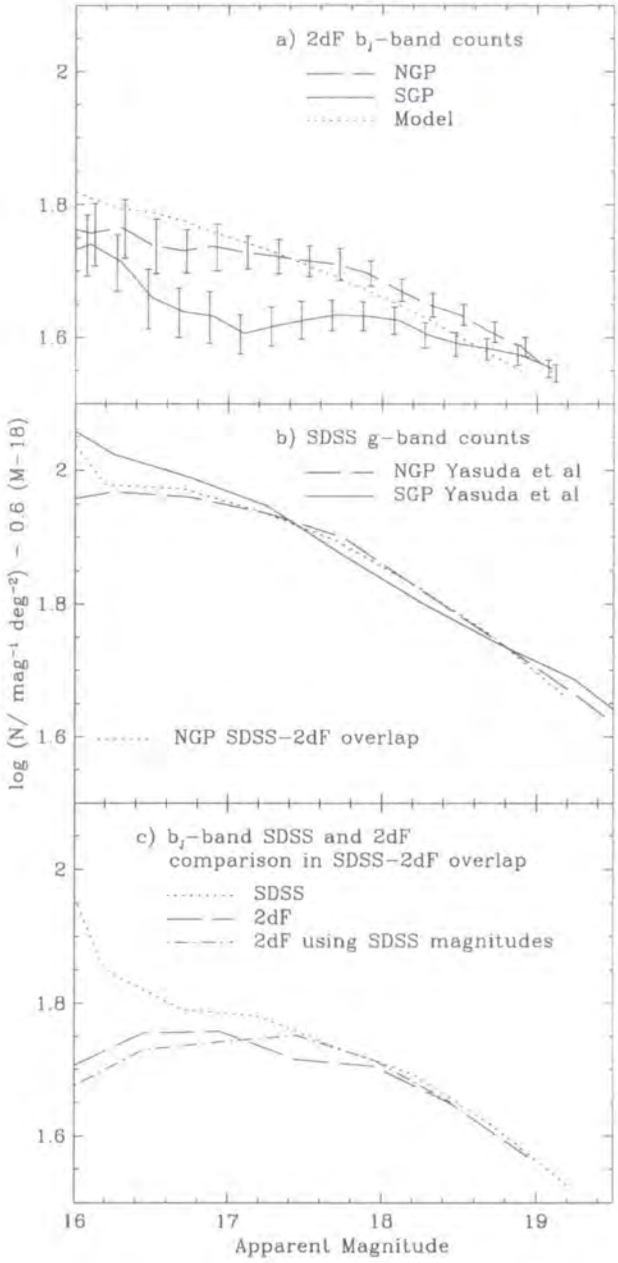


Figure 3.8: This figure shows 2dFGRS and SDSS galaxy number counts in the b_J and g-bands. In each panel we plot the logarithm of the number of galaxies per unit apparent magnitude after subtraction of a Euclidean model. The upper panel shows the 2dFGRS b_J-band counts separately in the NGP and SGP regions. The error bars show the rms variation we expect due to large scale structure, estimated from our 22 mock catalogues. The middle panel compares the published SDSS g-band counts of Yasuda et al. (2001) and our own estimate of the SDSS counts in the area which overlaps with the 2dFGRS NGP region. The bottom panel compares, in the overlap region, SDSS and 2dFGRS b_J-band counts.

The error bars placed on the measured counts are the rms scatter seen in our 22 mock catalogues and give an estimate of the variation one expects in the counts due to large scale structure. The dotted curve is mean number counts in the mocks and is the the expectation for a homogeneous universe.

It has long been know that the galaxy counts in the APM catalogue are steeper than model predictions for a homogeneous universe (Maddox et al. 1990d). As we have subtracted the Euclidean slope this manifests itself in Fig. 3.8 by the SGP curve having a shallower slope than model prediction shown by the dotted curve. The NGP counts are greater than those of the SGP throughout the range $16 < b_J < 19$ and are also steeper[‡] than the model prediction, though are not as extreme as the SGP. The 1-σ error bars determined from the mock catalogues show that variations from the homogeneous model prediction such as those shown in the NGP should be common. The SGP counts are harder to reconcile with the model, but it should be borne in mind that even on quite large scales the galaxy density field is non-Gaussian and so 1-σ error bars do not fully quantify the expected variation.

To normalize our estimates of the galaxy luminosity function we use the cumulative count of galaxies per square degree brighter than $b_J = 19.2$. In the 752 □° of the NGP strip this is 158.6 ± 6.3, where the error is again the rms found in the mock catalogues. The corresponding numbers for the 1089 □° SGP strip are 147.9 ± 6.3 and for the combined 1841□°, it corresponds to 152 ± 4.5. We note that the NGP and SGP number counts differ by 7% but this is reasonably common in the mock catalogues.

The middle panel in Fig. 3.8 shows SDSS g-band counts[§]. We show both the published SDSS counts from Yasuda et al. (2001) and our own estimate from the SDSS EDR in the region in which it overlaps with the 2dFGRS NGP strip. The very accurate agreement between the published northern counts and our estimate from the EDR data demonstrates that the simple star-galaxy classification we used in §3.2.2 works well fainter than $g > 16.5$ and that we have correctly estimated the area of the overlap between the SDSS EDR and the NGP region of the 2dFGRS. The Yasuda et al. (2001) counts are accurate to brighter magnitudes as they utilise a more sophisticated star-galaxy separation algorithm and are supplemented by visually classification.

The lower panel of Fig. 3.8 compares SDSS and 2dFGRS counts within the approximately 173 □° area of overlap of the two datasets. Here we have estimated b_J from

---

[‡] Shallower in Fig. 3.8, due to subtraction of the Euclidean slope.

[§] This being the SDSS band closest to b_J.

the SDSS Petrosian magnitudes using equation 3.1, but also included explicitly the 0.058 zero-point offset we measured in §3.2.1. We see that between $18 < b_J < 19$, 2dFGRS and SDSS number counts agree very accurately. In this area the cumulative count of galaxies per square degree brighter than $b_J = 19.2$ is 160, 5% higher than that averaged over the ten times larger area covered by the combined NGP+SGP 2dFGRS strips. Between $17 < b_J < 18$ the 2dFGRS counts are approximately 8% below the SDSS counts. Brighter than $b_J = 17$ this difference increases rapidly, but this is due to stellar contamination in our SDSS sample as can be seen by reference to the middle panel of Fig. 3.8. Interestingly if we compute the counts for the 2dFGRS objects, but using the magnitudes derived from the SDSS data then there is slightly better agreement between 2dFGRS and SDSS at $b_J \approx 17.5$.

We conclude from this comparison that in the 173 $\square^\circ$ region of overlap the 2dFGRS counts (corrected using the standard estimates of stellar contamination and incompleteness) are in good agreement with the SDSS counts fainter than $b_J > 17$, but are 5% higher than those averaged over the full area of the 2dFGRS. Over the full area we find 152 galaxies per square degree brighter than $b_J = 19.2$ with a 1-σ statistical error, estimated from mock catalogues, of just 3%.

3.7 The Normalized 2dFGRS Luminosity Function

We now use these number counts to normalize our luminosity function estimates. In the upper panel of Fig. 3.9 we present two independent estimates of the galaxy luminosity function from the NGP and SGP regions. Here the LF estimate in each region is normalized by its own galaxy number counts. Thus the two estimates are independent and the differences between them can be used as an estimate of the statistical error. These can be compared with the plotted SWML errors, but note should be taken that the SWML errors do not take account of the uncertainty in the normalization of the LF. For these two estimates the mock catalogues indicate that the contribution to the uncertainty from large scale structure is about 4%. Also of importance is the uncertainty in the incompleteness corrections. We have corrected assuming a global 12% incompleteness in the 2dFGRS photometric catalogue and the uncertainty in this adds, in quadrature, approximately 2% to the normalization uncertainty (see §3.2.2). An indication of this uncertainty is given by vertical 3σ range error bar plotted in the upper right of each panel of Fig.3.9. If this is added in quadrature to the SWML errors then one finds that the differences between the

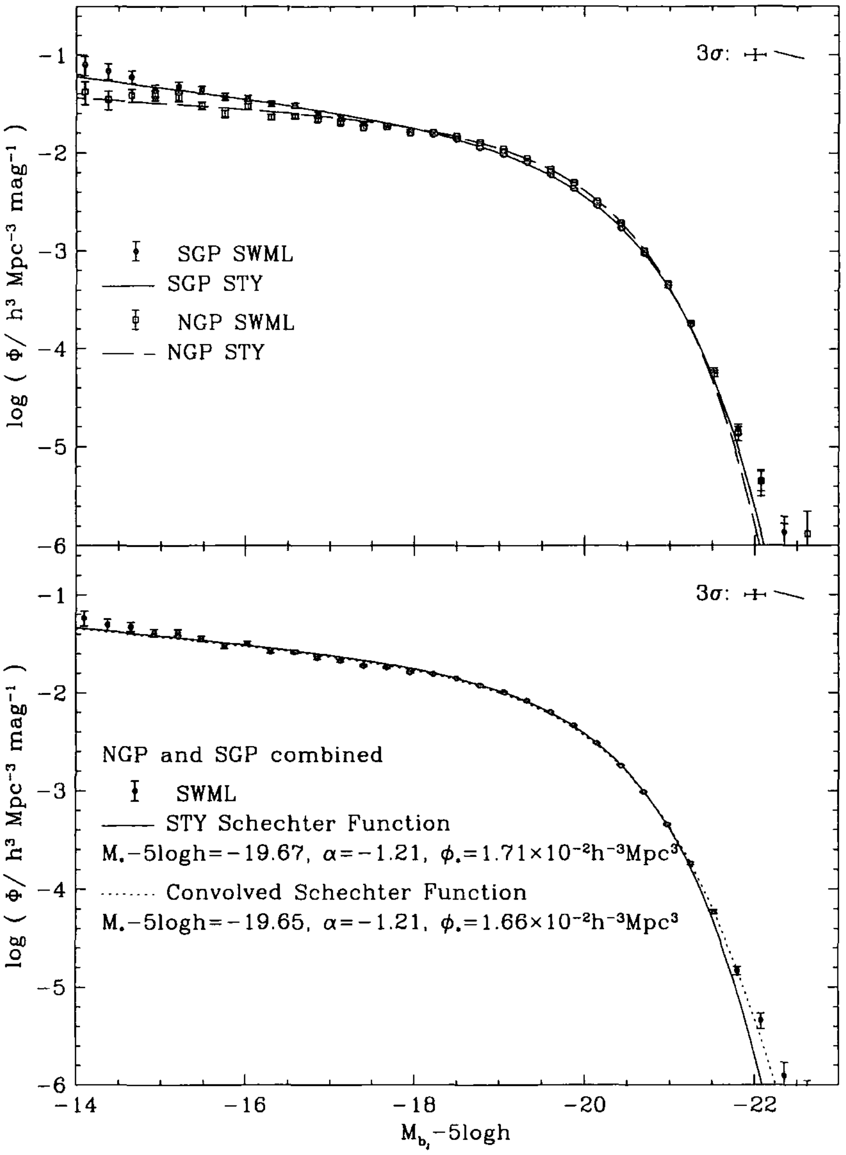


Figure 3.9: The upper panel shows the independent estimates of the $z = 0$ galaxy luminosity function from the NGP and SGP regions. The lower panel shows the combined NGP+SGP estimate normalized to the mean NGP+SGP number counts. The symbols show SWML estimates and the smooth solid curves STY Schechter function estimates. The dotted curve in the lower panel is the fit obtained using a Schechter function convolved with the distribution of magnitude measurement errors. The parameters of the Schechter functions are given in the legend. The error bars shown in the upper right of each panel are 3σ (for clarity) errors showing the additional uncertainty in the normalization (vertical), in the photometric zero point (horizontal) and in the $k + e$-corrections (slanted). These three sources of error are all independent but effect each data point in the LF coherently.

Table 3.1: The stepwise maximum likelihood (SWML) estimates of the 2dFGRS $z = 0$ galaxy luminosity function for three different cosmological models. The quoted errors do not take account of uncertainty in the normalization, the photometric zero-point or uncertainty in the appropriate evolutionary correction (see Section 3.6). Also these estimates are not deconvolved for the effect of random magnitude measurement errors.

	$\Omega_0 = 0.3$, $\Lambda_0 = 0.7$	$\Omega_0 = 1.0$, $\Lambda_0 = 0.0$	$\Omega_0 = 0.3$, $\Lambda_0 = 0.0$
$M_{b_J} - 5\log h$	$\phi/\mathrm{h}^3\mathrm{Mpc}^{-3}$	$\phi/\mathrm{h}^3\mathrm{Mpc}^{-3}$	$\phi/\mathrm{h}^3\mathrm{Mpc}^{-3}$
-15.200	$(3.949 \pm 0.287) \times 10^{-2}$	$(4.542 \pm 0.323) \times 10^{-2}$	$(4.303 \pm 0.310) \times 10^{-2}$
-15.475	$(3.425 \pm 0.221) \times 10^{-2}$	$(3.876 \pm 0.245) \times 10^{-2}$	$(3.734 \pm 0.238) \times 10^{-2}$
-15.750	$(2.896 \pm 0.165) \times 10^{-2}$	$(3.293 \pm 0.184) \times 10^{-2}$	$(3.175 \pm 0.178) \times 10^{-2}$
-16.025	$(3.085 \pm 0.147) \times 10^{-2}$	$(3.505 \pm 0.163) \times 10^{-2}$	$(3.371 \pm 0.158) \times 10^{-2}$
-16.300	$(2.546 \pm 0.111) \times 10^{-2}$	$(2.929 \pm 0.124) \times 10^{-2}$	$(2.791 \pm 0.119) \times 10^{-2}$
-16.575	$(2.513 \pm 0.097) \times 10^{-2}$	$(2.861 \pm 0.107) \times 10^{-2}$	$(2.726 \pm 0.103) \times 10^{-2}$
-16.850	$(2.190 \pm 0.075) \times 10^{-2}$	$(2.523 \pm 0.082) \times 10^{-2}$	$(2.421 \pm 0.080) \times 10^{-2}$
-17.125	$(2.048 \pm 0.059) \times 10^{-2}$	$(2.357 \pm 0.064) \times 10^{-2}$	$(2.276 \pm 0.062) \times 10^{-2}$
-17.400	$(1.820 \pm 0.043) \times 10^{-2}$	$(2.036 \pm 0.046) \times 10^{-2}$	$(1.955 \pm 0.045) \times 10^{-2}$
-17.675	$(1.751 \pm 0.035) \times 10^{-2}$	$(1.990 \pm 0.038) \times 10^{-2}$	$(1.904 \pm 0.037) \times 10^{-2}$
-17.950	$(1.555 \pm 0.027) \times 10^{-2}$	$(1.806 \pm 0.029) \times 10^{-2}$	$(1.730 \pm 0.029) \times 10^{-2}$
-18.225	$(1.491 \pm 0.022) \times 10^{-2}$	$(1.700 \pm 0.024) \times 10^{-2}$	$(1.621 \pm 0.023) \times 10^{-2}$
-18.500	$(1.353 \pm 0.018) \times 10^{-2}$	$(1.514 \pm 0.019) \times 10^{-2}$	$(1.460 \pm 0.018) \times 10^{-2}$
-18.775	$(1.147 \pm 0.013) \times 10^{-2}$	$(1.277 \pm 0.014) \times 10^{-2}$	$(1.234 \pm 0.014) \times 10^{-2}$
-19.050	$(9.779 \pm 0.102) \times 10^{-3}$	$(1.089 \pm 0.011) \times 10^{-2}$	$(1.046 \pm 0.011) \times 10^{-2}$
-19.325	$(7.968 \pm 0.077) \times 10^{-3}$	$(8.588 \pm 0.080) \times 10^{-3}$	$(8.430 \pm 0.079) \times 10^{-3}$
-19.600	$(6.108 \pm 0.057) \times 10^{-3}$	$(6.567 \pm 0.060) \times 10^{-3}$	$(6.385 \pm 0.059) \times 10^{-3}$
-19.875	$(4.428 \pm 0.042) \times 10^{-3}$	$(4.533 \pm 0.044) \times 10^{-3}$	$(4.517 \pm 0.043) \times 10^{-3}$
-20.150	$(2.927 \pm 0.030) \times 10^{-3}$	$(2.840 \pm 0.030) \times 10^{-3}$	$(2.877 \pm 0.030) \times 10^{-3}$
-20.425	$(1.747 \pm 0.021) \times 10^{-3}$	$(1.605 \pm 0.020) \times 10^{-3}$	$(1.695 \pm 0.021) \times 10^{-3}$
-20.700	$(9.309 \pm 0.132) \times 10^{-4}$	$(8.022 \pm 0.129) \times 10^{-4}$	$(8.380 \pm 0.129) \times 10^{-4}$
-20.975	$(4.343 \pm 0.081) \times 10^{-4}$	$(3.369 \pm 0.079) \times 10^{-4}$	$(3.821 \pm 0.080) \times 10^{-4}$
-21.250	$(1.746 \pm 0.048) \times 10^{-4}$	$(1.102 \pm 0.044) \times 10^{-4}$	$(1.289 \pm 0.045) \times 10^{-4}$
-21.525	$(5.668 \pm 0.268) \times 10^{-5}$	$(3.305 \pm 0.247) \times 10^{-5}$	$(3.962 \pm 0.250) \times 10^{-5}$
-21.800	$(1.413 \pm 0.137) \times 10^{-5}$	$(8.686 \pm 1.326) \times 10^{-6}$	$(1.091 \pm 0.136) \times 10^{-5}$
-22.075	$(4.404 \pm 0.797) \times 10^{-6}$	$(2.852 \pm 0.818) \times 10^{-6}$	$(3.481 \pm 0.815) \times 10^{-6}$
-22.350	$(1.188 \pm 0.446) \times 10^{-6}$	$(6.212 \pm 4.351) \times 10^{-7}$	$(4.776 \pm 3.346) \times 10^{-7}$

Table 3.2: Schechter function fits to the underlying 2dFGRS galaxy luminosity function for three different cosmologies. The parameters listed are such that when the corresponding Schechter function is convolved with the apparent magnitude measurement errors they accurately reproduce the observed 2dFGRS galaxy luminosity function. The last column list the integrated luminosity density in solar units ($M^{\odot}_{b_J} = 5.3$). The various contributions to the quoted errors are explained in the text. If one is interested in the absolute error in the LF these errors should be added in quadrature.

Ω_0	Λ_0	$M^{\star}_{b_J} - 5\log h$	α	$\Phi^{\star}/\mathrm{h^3Mpc^{-3}}$	$\rho_L/\mathrm{h\ L_{\odot}\ Mpc^{-3}}$
0.3	0.7	$-19.66 \pm 0.006 \pm 0.06 \pm 0.04$	$-1.21 \pm 0.01 \pm 0.02$	$(1.66 \pm 0.015 \pm 0.05 \pm 0.06) \times 10^{-2}$	$(1.88 \pm 0.18) \times 10^{-8}$
1.0	0.0	$-19.48 \pm 0.006 \pm 0.06 \pm 0.04$	$-1.18 \pm 0.01 \pm 0.02$	$(2.13 \pm 0.020 \pm 0.06 \pm 0.08) \times 10^{-2}$	$(1.99 \pm 0.19) \times 10^{-8}$
0.3	0.0	$-19.54 \pm 0.006 \pm 0.06 \pm 0.04$	$-1.19 \pm 0.01 \pm 0.02$	$(1.96 \pm 0.019 \pm 0.06 \pm 0.07) \times 10^{-2}$	$(1.93 \pm 0.19) \times 10^{-8}$

NGP and SGP estimates are entirely consistent within these errors except fainter than $M_{\mathrm{b_J}} - 5\log h \simeq -17.5$.

At the faint end the SGP LF is slightly steeper than that estimated from the NGP. This may reflect genuine spatial variations in the galaxy luminosity function as this faint portion of the luminosity function is determined from a very local volume. Such variations are perhaps to be expected given the results of chapters 6 and 7, where we show that galaxies of different luminosity have systematically different clustering properties. The faint end of the luminosity function may also be affected by incompleteness in the 2dFGRS. We have corrected the LF assuming that the incompleteness is independent of absolute magnitude. However, from the joint analysis of the 2dFGRS and the much deeper Millennium Galaxy Catalogue (MGC) by Cross & Driver (2001) we know part of the incompleteness is due to the 2dFGRS preferentially missing low surface brightness galaxies. The correlation between absolute magnitude and surface brightness then implies that low luminosity galaxies are under represented. The work of Cross & Driver (2001) suggests this only becomes important fainter than $M_{\mathrm{b_J}} - 5\log h \simeq -17.5$.

There are also two other significant contributions to the uncertainty of the galaxy luminosity on an absolute scale. The first of these is the zero-point of the photometry which has an accuracy of ± 0.04 magnitudes. An indication of this uncertainty is given by horizontal 3σ range error bar plotted in the upper right of each panel of Fig.3.9. The second important contribution is the uncertainty in the appropriate evolutionary correction. Our estimates of the galaxy luminosity function are at redshift $z = 0$ and so rely on an accurate model of the $k+e$-corrections to transform the measured luminosities, which have a median redshift of $z_{\mathrm{med}} \approx 0.1$, to present day values. The k+e-correction we use is accurately constrained by the SDSS (g−r) colours, but are nevertheless model dependent at some level. To gauge the error in the luminosity function estimates due to this uncertainty, we made SWML LF estimates using $k+e$-corrections that were increased or decreased by some factor compared to our standard model. We then constrained this factor by comparing luminosity function estimates made separately for the data above and below $z = 0.1$. The results of this test for the standard $k + e$-correction model were shown in Fig. 3.7(d), where it can be seen that the two luminosity functions match accurately. We find that if the $k + e$-corrections are increased or decreased by 18% then the position of the break in the luminosity function between the high and low redshift samples differs by 1σ (as determined using the SWML errors). Taking this as an estimate of the uncertainty in the $k+e$-correction, we find that the corresponding uncertainty in the luminosity function parameters are $\Delta\alpha = 0.02$, $\Delta M^{\star}_{\mathrm{b_J}} = 0.06$, and $\Delta\Phi^{\star}/\Phi^{\star} = 3\%$. The variations in $M^{\star}_{\mathrm{b_J}}$ and $\Phi^{\star}$ are strongly correlated. This contribution to the uncertainty in the LF estimates is indicated by the slanted 3σ range error bar plotted in the upper right of each panel of Fig.3.9.

The lower panel of Fig. 3.9 combines the SGP and NGP data to give our best estimate of the $\mathrm{b_J}$-band galaxy luminosity function assuming an $\Omega_0 = 0.3$ and $\Lambda_0 = 0.7$ cosmology. The points with error bars show the SWML estimate. Also shown are two Schechter function estimates, whose parameter values are indicated in the legend. The first is a simple STY estimate of the 2dFGRS luminosity function, while in the case of the second what has been fitted to the data is a Schechter function convolved with the distribution of magnitude measurement errors estimated from Fig. 3.1. We see that deconvolving the effect of the magnitude errors causes only a small reduction in $M^{\star}_{\mathrm{b_J}}$ and $\Phi^{\star}$. We also note that this function convolved with the errors (shown by the dotted curve) produces a good match to the SWML estimate. Thus there is little evidence for the underlying galaxy luminosity function differing significantly from the Schechter function form.

The numerical values of these estimates are listed in Tables 3.1 and 3.2 along with estimates for alternate cosmologies. Note that the SWML estimates are of the observed 2dFGRS luminosity function, which is distorted due to random apparent magnitude measurement errors. In contrast the Schechter function parameters we list in Table 3.2 refer to the underlying galaxy luminosity function deconvolved for the effect of magnitude measurement errors. In Table 3.2 we have broken down the errors on the Schechter function parameters into three components. The first is the statistical error returned by the STY maximum likelihood method. The large number of galaxies used in our estimates means that this statistical error is very small and is never the dominant contribution to the overall error. The second error is our estimate of the uncertainty induced through the uncertainty in the $k+e$-corrections. This is the dominant contribution to the error in α and also a significant contributor to both the errors in $M^{\star}_{\mathrm{b_J}}$ and $\Phi^{\star}$. The third error given for $M^{\star}_{\mathrm{b_J}}$ in Table 3.2 is that due to the current uncertainty in the 2dFGRS photometric zero point. This will be reduced when more calibrating CCD photometry is available. The third error given for $\Phi^{\star}$ is that due to the uncertainty in the galaxy number counts and has contributions from large scale structure (3%) and the uncertainty in the incompleteness corrections (2%). To determine the overall errors on an absolute scale these contributions should all be added in quadrature. The final column in Table 3.2 lists the implied $z = 0$ luminosity density in solar units. The error quoted on this quantity was computed by propagating all the previously mentioned errors.

The Schechter function parameters listed in Table 3.2 for the $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology differ slightly from those in Madgwick et al. (2001). This is to be expected as the Madgwick et al. luminosity functions are not corrected for evolution. Their paper focuses on the dependence of the luminosity function on spectral type. Adopting the average k-correction of Madgwick et al. and using this in place of our $k + e$-correction on our largest sample[¶], we find luminosity function parameters very close to those of Madgwick et al. (2001). The remaining very small differences are accounted for by slightly differing models for the magnitude errors and the adopted normalizations.

3.8 Comparison with other Luminosity Function Estimates

In Fig. 3.10 we compare the STY and SWML estimates of the b_J-band luminosity function from the combined NGP+SGP 2dFGRS sample defined by $17 < b_J < 19.2$ and $z < 0.25$ (shown in Fig 3.9) with estimates from other surveys. The upper panel compares 2dFGRS with various estimates made from SDSS. In this comparison we again assume an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology. Blanton et al. (2001) presented an estimate of the b_J-band LF for the case of $\Omega_0 = 1.0$. We do not use this, but instead estimate the b_J-band LF for our adopted cosmology using the g-band LF computed by Blanton et al. (2001) for an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology and the typical $B - V$ galaxy colour. Using the colour equations of Fukugita et al. (1996) and assuming $b_J = B - \beta(B - V)$, one finds $b_J = g + 0.12 + (0.44 - \beta)(B - V)$. Blanton et al. (2001) assumed a value of 0.35 for β, to which we return in the next paragraph. Thus an estimate of the b_J-band LF can be made by simply taking the g-band estimate and shifting the magnitudes using this equation with $B - V = 0.94$, which is the mean colour measured from the SDSS sample of galaxies fainter than $b_J = 19$. This procedure can been seen to work quite accurately, as when applied to the $\Omega_0 = 1$ g-band LF parameters given in Table 2 of Blanton et al. (2001), it reproduces the corresponding b_J parameters given in their Figure 23. Taking $\beta = 0.35$ and applying this procedure for the $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology gives $M^{\star}_{b_J} - 5\log h = -19.82$, $\alpha = -1.26$ and $\Phi^{\star} = 2.06 \times 10^{-2}\,h^3\,\mathrm{Mpc}^{-3}$. This Schechter function is shown by the long dashed curve in the upper panel of Fig. 3.10. As discussed in Blanton et al. (2001) this estimate is incompatible with the 2dFGRS estimate and predicts a significantly higher luminosity density than we find.

[¶] The Madgwick et al. (2001) sample is truncated at z=0.15, as briefly mentioned in section 3.3.

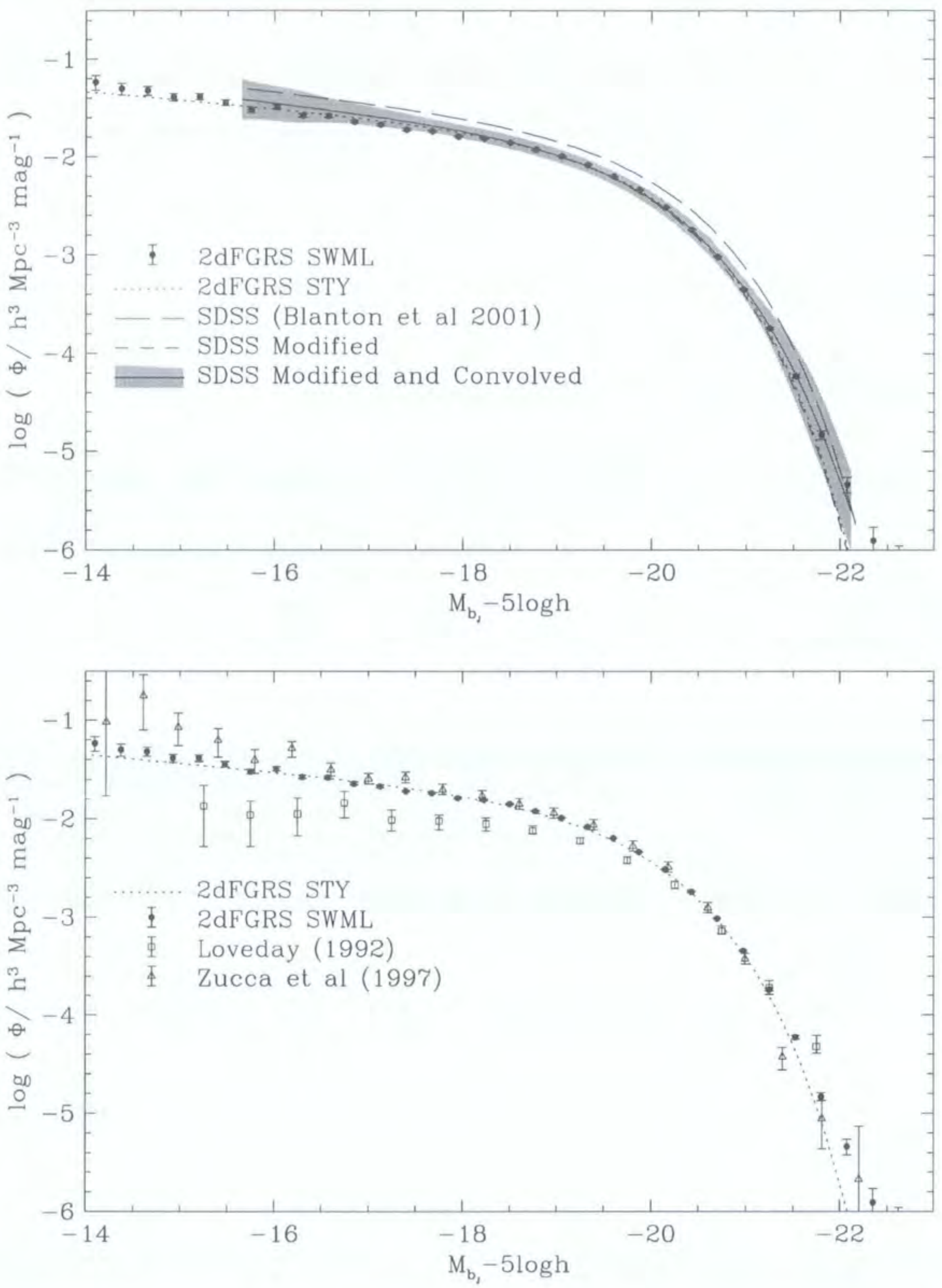


Figure 3.10: Comparison of the 2dFGRS b$_J$-band galaxy luminosity function with estimates from SDSS (Blanton et al. 2001, shown in the upper panel) and earlier estimates (Loveday et al. 1992 and Zucca et al. 1997, shown in the bottom panel). See text for further details.

The short dashed luminosity function in the upper panel of Fig. 3.10, a Schechter function with $M^\star_{\mathrm{b_J}} - 5\log h = -19.679$, $\alpha = -1.26$ and $\Phi^\star = 1.63 \times 10^{-2}\,\mathrm{h^3\,Mpc^{-3}}$, is the result of making three modifications to the Blanton et al. (2001) curve. First we have shifted $M^\star_{\mathrm{b_J}}$ by 0.08 magnitudes as is appropriate if one adopts colour equation $\mathrm{b_J} = B - 0.28(B - V)$ rather than $\mathrm{b_J} = B - 0.35(B - V)$, as used by Blanton et al. (2001). The latter is ruled out by the empirical relations found by matching the 2dFGRS catalogue with either the EIS or SDSS which are instead consistent with the former colour equation which is from Blair & Gilmore (1982). Second we have shifted $M^\star_{\mathrm{b_J}}$ by a further 0.058 to take account of the zero point offset between the SDSS and 2dFGRS photometry that we found in Fig. 3.1. Finally we have reduced $\Phi^\star$ by 21%, which is the reduction required for this luminosity function to reproduce the mean 2dFGRS number counts at $b_J = 19.2$ assuming our standard $k + e$-correction model. We note that Yasuda et al. (2001) also found a value of $\Phi^\star$ significantly lower than that of Blanton et al. when they normalized the SDSS g-band luminosity function using the SDSS galaxy counts. The Yasuda et al. estimate is still higher than our value as, while the SDSS counts agree with 2dFGRS in the area of overlap, this area is small (173 $\square^\circ$) and has a 5% higher density of galaxies than the full area (1880 $\square^\circ$) covered by the 2dFGRS survey, as discussed in section 3.6. This modified SDSS Schechter function is in near perfect agreement the Schechter function estimated from the 2dFGRS.

At the brightest magnitudes the 2dFGRS SWML estimate is above both the 2dFGRS STY estimate and the modified SDSS Schechter function estimate. The main reason for this is that magnitude measurement errors in the 2dFGRS have a significant effect of the bright end of the luminosity function, but little effect around $M^\star_{\mathrm{b_J}}$ and fainter. The solid curve surrounding by the shaded region shows the result of convolving the modified SDSS estimate with the model of the 2dFGRS magnitude errors shown in Fig. 3.1. The shaded region indicates the statistical error on the SDSS estimate and was read off figure 6 of Blanton et al. (2001). Comparing this with the 2dFGRS SWML estimate we see that the two are perfectly consistent with the larger 2dFGRS sample having significantly smaller statistical errors.

We have seen that after taking account of the zero point photometric offset and error in colour equation the only significant difference between the LF estimate of Blanton et al. (2001) and the 2dFGRS estimate is a difference in $\Phi^\star$. This difference arose not because the density of galaxies is higher in SDSS than 2dFGRS, since the counts agree to 5%, but because of the method used to constrain $\Phi^\star$. Blanton et al. (2001) used the method of Davis & Huchra (1982b) which weights galaxies as a function of redshift in order to get a minimum variance estimate of the galaxy density. This method gives more weight to galaxies at high redshift than normalizing to the counts. This results in a smaller statistical error in the normalization, but at the same time makes the result more dependent on the accuracy of the evolutionary correction. We have seen, in Section 3.7, that even with the low redshift constraint provided by the galaxy counts, the uncertainty in $\Phi^\star$ due to the uncertainty in the $k+e$-correction is significant. With the Davis & Huchra weighting this uncertainty becomes dominant. The Blanton et al. analysis did not take account of evolution (only k-corrections were applied) and this appears to have given rise to an artificially high estimate of value of $\Phi^\star$ in the g-band. We conclude that when normalized in the same way there is excellent agreement between SDSS and 2dFGRS luminosity function and that the dominant remaining uncertainty in the present day b_J-band luminosity function is due to residual uncertainties in evolutionary corrections.

The lower panel of Fig. 3.10 compares the 2dFGRS result with earlier estimates of Loveday et al. (1992) and Zucca et al. (1997). We see that the Zucca et al. estimate agrees well with 2dFGRS although it has statistical errors that are much larger. The difference between the luminosity function estimate of Loveday et al. (1992) and that of the 2dFGRS is that it has not only a lower $\Phi^\star$, but a flater faint end slope. The 2dFGRS and Loveday et al. estimate are both based on catalogues extracted from the APM survey. However the Loveday et al. sample is much brighter and almost disjoint from the sample analyzed in this chapter. In the SGP the bright galaxy number counts drop below model predictions extrapolated from fainter magnitudes, as shown in section 3.6 and in Maddox et al. (1990d). Thus it is not surprising that Loveday et al. (1992) found a lower value of $\Phi^\star$. The difference in the faint end slope on the other hand is likely to be related to the depth of the Stromlo-APM survey, and potentially also the sparse sampling strategy used, which might be more subject to 'miss-represent' the very faintest galaxies.

3.9 The 2dFGRS Selection Function

The luminosity function that we have derived, when combined with the maps defining the survey magnitude limit (see Fig. 2.6), redshift completeness (see Fig. 1.6) and μ-parameter (see Fig. 2.13 for an example of a μ-mask), defines a complete 3-dimensional description of the 2dFGRS selection function. The only significant feature of the 2dFGRS selection function ignored in this description is the under sampling of close galaxy pairs

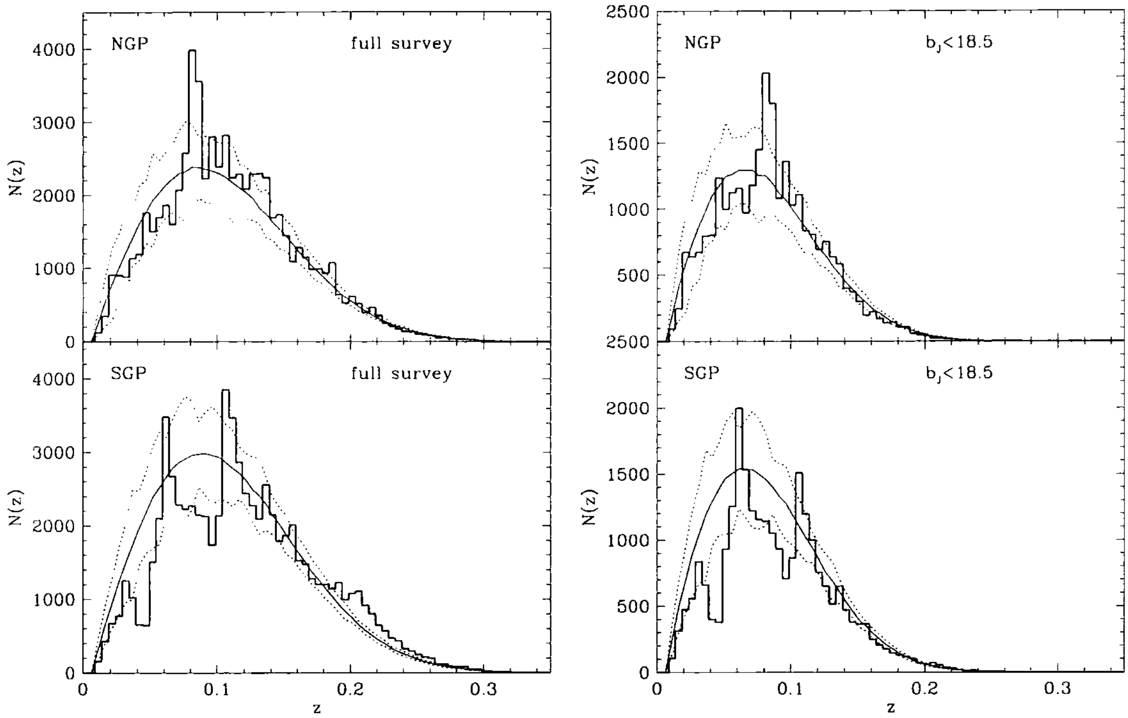


Figure 3.11: The histograms in these four panels show the observed redshift distribution in the NGP and SGP regions of the 2dFGRS. The left hand panels are to the full depth of the survey while the right hand panels only include galaxies brighter than $b_J = 18.5$. The smooth solid curves show the predicted redshift distributions based on our Schechter function estimate of the galaxy luminosity function, including the magnitude measurement errors, the variation in the survey magnitude limit and the dependence of completeness on apparent magnitude. The dashed lines indicate the rms variation in the redshift histograms we find within our ensemble of 22 mock catalogues.

induced by the mechanical limits on the positioning of the optical fibres that feed the 2dF spectrograph. Note that as the 2dF fields overlap not all close galaxy pairs are missed, just a small fraction. We have found that when making estimates of galaxy clustering an accurate way of dealing with this incompleteness is to assign the weight of the missed galaxies to neighbouring galaxies with redshifts, as explained in section 2.5.

It is interesting to compare the redshift distribution predicted by this selection function with the measured distribution. Note that the LF estimators we employed are independent of clustering and so the information contained in the redshift distribution of the galaxies has not been used in determining our model of the selection function.

In Fig. 3.11, we compare the smooth redshift distribution predicted by our model of the 2dFGRS selection function with the observed distribution. The left hand panels show

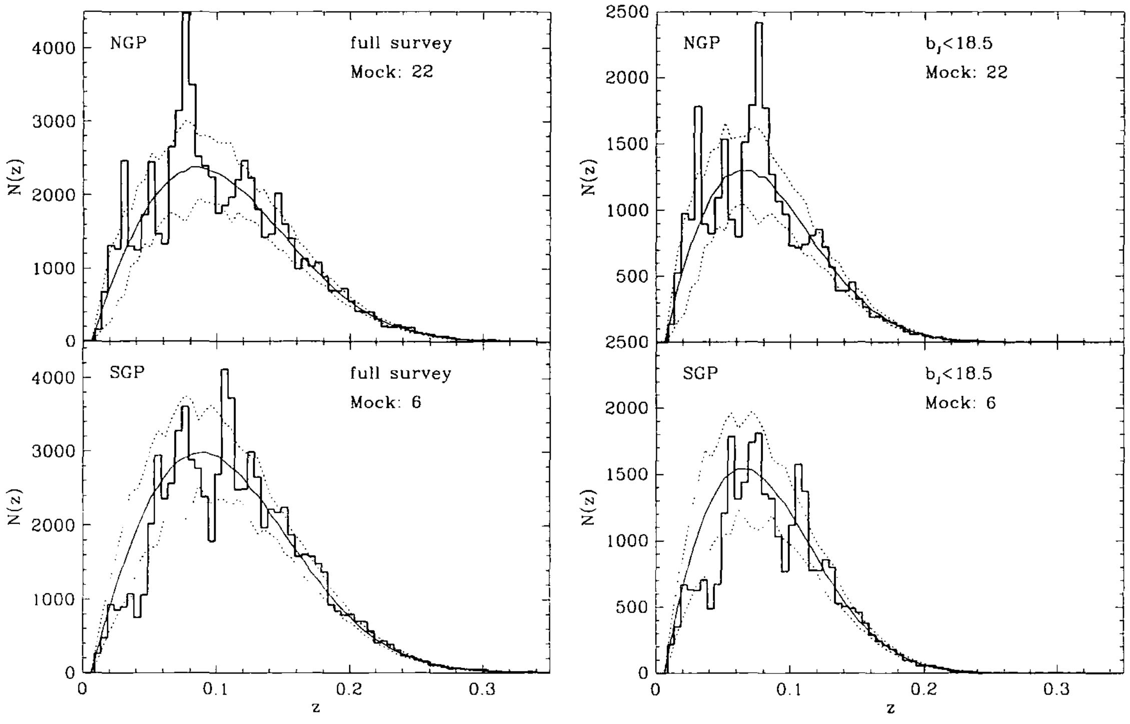


Figure 3.12: Same as Fig. 3.11, but in place of the genuine 2dFGRS data we have plotted the redshift histogram from two selected mock catalogues.

the redshift distributions for the full 2dFGRS survey split into the SGP and NGP regions. The right hand panels show the distributions only for galaxies brighter than b$_J$ = 18.5. The dotted lines on these plots show the rms variation in the redshift histograms that we find in our 22 mock catalogues. As gravitational clustering produces a pattern of galaxy clustering that is non-Gaussian and composed of voids, walls and filaments (e.g. see figures 8 to 15 of Cole et al. 1998, for mock 2dFGRS and SDSS cone plots), the rms variation in the $N(z)$ distribution does not give an adequate description of the variation one sees in the mock catalogues. For this reason we show in Fig. 3.12 corresponding redshift distributions that we find in two examples drawn from our ensemble of mock catalogues. From these we see that the few large spikes we see in the $N(z)$ of the 2dFGRS data are common features in the mock catalogue redshift distributions.

The redshift distribution in the 2dFGRS NGP has one spike close to the peak of the selection function and otherwise lies within 1-σ of our smooth selection function. Thus the density field in the NGP strip looks in no way unusual when compared to the expectation in our standard CDM $\Omega_0 = 0.3$ $\Lambda_0 = 0.7$ universe. In contrast, the density field in the SGP appears more extreme. Focussing first on the redshift distribution below $z < 0.2$, we see that the observed galaxy density is nearly always below the mean density predicted by the selection function. This observation is consistent with the steep APM galaxy number counts first discussed in Maddox et al. (1990d) and which we discussed in section 3.6. The galaxy density being below average over such a large range of redshift is certainly an unlikely occurrence. However, as illustrated by the example plotted in Fig. 3.12, which is many respects is quite similar to the observed 2dFGRS SGP, comparable variations do occur in the mock ΛCDM catalogues. Note that the two examples plotted in Fig. 3.12 were not chosen at random, but as we only have 22 mocks to choose from they do not represent extreme possibilities.

The 2dFGRS SGP strip also shows a strong overdensity, compared to the selection function model, in the redshift range $0.2 < z < 0.25$. As the volume contributing to this redshift interval is very large, a variation as large as this is very unlikely. It therefore seems unlikely that this perturbation in the $N(z)$ is due to large scale structure. This conclusion is also consistent with the feature in $N(z)$ not being localized on the sky. It therefore appears that this apparent over density must be due to a feature in the 2dFGRS selection function which we have not adequately modelled. Investigation into this problem is ongoing. At $z > 0.2$ the only galaxies which make it into the 2dFGRS are several magnitudes brighter than $M^*_{\rm b_J}$, where the galaxy luminosity function is very steep. Thus a small error in the measured magnitude can result in large change in the number of galaxies brighter than the survey magnitude limit. We have ruled out a simple offset in the magnitudes of the high redshift objects but a possibility that needs further investigation is that the random magnitude measurement become large for faint objects at high z. For now one should be careful, as we have been, to ensure that large scale clustering results are not strongly influenced by this feature. One way to assess the influence of this overdensity on large scale structure results, is to compare clustering results from the SGP region with those from the NGP region, which is very well described by our selection function, as shown by Fig. 3.11.

3.10 Conclusions

In this chapter we have looked into various aspects which have an influence on the estimate of the b$_J$-band galaxy luminosity function, which we aim to determine accurately.

We first investigated the overall completeness of the 2dFGRS photometric input catalogue, by a detailed comparison, using an overlapping region, with the SDSS EDR. This allowed us to estimate the completeness of the catalogue to be 88% ± 2%. The stellar contamination, as estimated internally with 2dFGRS redshifts in chapter 2, is in very good agreement with estimate obtained with external data from SDSS EDR. The photometric accuracy of the 2dFGRS was estimated by comparing to multi-band CCD data from SDSS EDR. We concluded that the error distribution for 2dFGRS magnitude is well fitted by a sum of a Gaussian (70%) and a log-normal distribution (30%), with $\sigma = 0.14$ and $\sigma = 0.235$ respectively. Corrections for band shifting and galaxy evolution over the redshift range probed were made by adopting a set of $k+e$-corrections using the Bruzual & Charlot (1993) stellar population synthesis code. These $k+e$ correction have been tuned such as to reproduce the g−rcolours from SDSS EDR as function of redshift and spectral type. A close look at the galaxy number counts in 2dFGRS and SDSS allowed us to estimate the mean number of galaxies per square degree brighter than $\mathrm{b_J} = 19.2$ to be 152 ± 4.5, where the error is estimated from a set of 22 mock catalogues.

After intensive reliability and consistency checks, which consisted of splitting the galaxy samples by region, by redshift and by magnitude limit, we obtained a robust estimate of the $\mathrm{b_J}$-band galaxy luminosity function. It is very well fit by a Schechter function with $M^{\star}_{\mathrm{b_J}} - 5\log h = -19.67 \pm 0.07$, $\alpha = -1.21 \pm 0.03$ and $\phi^{\star} = (1.71 \pm 0.08) \times 10^{-2}\,\mathrm{h^3\,Mpc^{-3}}$. We did a detailed investigation of the errors on these parameters, arising from the uncertainty in the normalisation, the photometric zero-point and the assumed $k+e$ correction. Together with the statistical uncertainty returned by the STY method, we list them all separately for each parameter in Table 3.2.

We find that our estimate luminosity function is in very good agreement with the one of Blanton et al. (2001), as long as one uses the right colour equation to transform the g-band into $\mathrm{b_J}$, together with a similar $k+e$-correction and a consistent method to normalize the luminosity function. From this comparison, we conclude that the dominant remaining uncertainty in the present day $\mathrm{b_J}$-band luminosity function is due to residual uncertainties in evolutionary corrections, and, at some level, the way in which one normalizes the galaxy luminosity function.

Chapter 4

Near Infra-Red Luminosity Functions

We combine the 2MASS[*] extended source catalogue and the 2dF galaxy redshift survey to produce an infrared-selected galaxy catalogue with 17,173 measured redshifts. We use this extensive dataset to estimate the galaxy luminosity functions in the J- and $\mathrm{K_S}$-bands. The luminosity functions are fairly well fit by Schechter functions with parameters $M^\star_{\mathrm{J}} - 5\log h = -22.36 \pm 0.02$, $\alpha_{\mathrm{J}} = -0.93 \pm 0.04$, $\Phi^\star_{\mathrm{J}} = 0.0104 \pm 0.0016 h^3$ Mpc^{-3} in the J-band and $M^\star_{\mathrm{K_S}} - 5\log h = -23.44 \pm 0.03$, $\alpha_{\mathrm{K_S}} = -0.96 \pm 0.05$, $\Phi^\star_{\mathrm{K_S}} = 0.0108 \pm 0.0016 h^3$ Mpc^{-3} in the $\mathrm{K_S}$-band (2MASS Kron magnitudes). These parameters are derived assuming a cosmological model with $\Omega_0 = 0.3$ and $\Lambda_0 = 0.7$. With datasets of this size, systematic rather than random errors are the dominant source of uncertainty in the determination of the luminosity function. We carry out a careful investigation of possible systematic effects in our data. The surface brightness distribution of the sample shows no evidence that significant numbers of low surface brightness or compact galaxies are missed by the survey. We estimate the present-day distributions of $\mathrm{b_J}-\mathrm{K_S}$ and $\mathrm{J}-\mathrm{K_S}$ colours as a function of absolute magnitude and use models of the galaxy stellar populations, constrained by the observed optical and infrared colours, to infer the galaxy stellar mass function. Integrated over all galaxy masses, this yields a total mass fraction in stars (in units of the critical mass density) of $\Omega_{\mathrm{stars}} h = (1.6 \pm 0.24) \times 10^{-3}$ for a Kennicutt IMF and $\Omega_{\mathrm{stars}} h = (2.9 \pm 0.43) \times 10^{-3}$ for a Salpeter IMF. These values are consistent with those inferred from observational estimates of the total star formation history of the universe provided that dust extinction corrections are modest.

---

[*] This chapter makes use of data products from the Two Micron All Sky Survey (2MASS), which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

4.1 Introduction

The near-infrared galaxy luminosity function is an important characteristic of the local galaxy population. It is a much better tracer of evolved stars, and hence of the total stellar content of galaxies, than optical luminosity functions which can be dominated by young stellar populations and are also strongly affected by dust extinction. Hence, infrared luminosities can be much more directly related to the underlying stellar mass of galaxies and so knowledge of the present form and evolution of the infrared galaxy luminosity function places strong constraints on the history of star formation in the universe and on galaxy formation models (e.g. Cole et al. (2000) and references therein).

The local K-band luminosity function has been estimated from optically selected samples by Mobasher et al. (1993), Szokoly et al. (1998) and Loveday (2000) and from K-band surveys by Glazebrook et al. (1995), and Gardner et al. (1997). The existing K-band surveys are small. The largest, by Gardner et al., covers only 4 deg^2 and contains only 510 galaxies. The recent survey of Loveday covers a much larger solid angle. In this survey the redshifts were known in advance of measuring the K-band magnitudes and this was exploited by targeting bright and faint galaxies resulting in an effective sample size much larger than the 345 galaxies actually measured. However, like all optically selected samples, it suffers from the potential problem that galaxies with extremely red infrared to optical colours could be missed. In this chapter we combine the 2-Micron All Sky Survey (2MASS) with the 2dF galaxy redshift survey (2dFGRS) to create an infrared selected redshift survey subtending 2151.6 deg^2. Currently the sky coverage of both surveys is incomplete, but already the overlap has an effective area of 619 deg^2. Within this area the redshift survey is complete to the magnitude limit of the 2MASS catalogue and so constitutes a complete survey which is 50 times larger than the previous largest published infrared selected redshift survey. A new catalogue of a similarly large area, also based on 2MASS, has very recently been analysed by Kochanek et al. (2001). They adopt isophotal rather than total magnitudes and concentrate on the dependence of the luminosity function on galaxy morphology.

This chapter has the following structure. In section 4.2.1 we briefly describe the relevant properties of the 2dFGRS and 2MASS catalogues. Section 4.2.2 is a detailed examination of the degree to which the matched 2MASS–2dFGRS galaxies are a complete and representative subset of the 2MASS catalogue. Section 4.2.3 examines the calibration of the 2MASS total magnitudes and section 4.2.4 demonstrates that the 2MASS catalogue and the inferred luminosity functions are not affected by surface brightness selection effects. In section 4.3 we present the method by which we compute k-corrections and evolutionary corrections and relate the observed luminosities to the underlying stellar mass. The estimation methods and normalization of the luminosity functions are described briefly in section 4.4. Our main results are presented and discussed in section 4.5. These include estimates of the J and K_S (K-short) luminosity functions, the $b_J - K_S$ and $J - K_S$ colour distributions as a function of absolute magnitude and the distribution of spectral type. We also estimate the stellar mass function of galaxies, which can be integrated to infer the fraction of baryons in the universe which are in the form of stars. We conclude in section 4.6.

4.2 The Dataset

The data that we analyze are the extended source catalogue from the second incremental release of 2MASS (http://pegasus.phast.umass.edu) and the galaxy catalogue of the 2dFGRS (http://www.mso.anu.edu.au/2dFGRS). Here, we present the relevant properties of these two catalogues and investigate their selection characteristics and level of completeness.

4.2.1 Selection criteria

The 2MASS is a ground-based, all-sky imaging survey in the J, H and K_S bands. Details of how extended sources are identified and their photometric properties measured are given by Jarrett et al. (2000). The detection sensitivity (10σ) for extended sources is quoted as 14.7, 13.9 and 13.1 magnitudes in J, H and K_S respectively. The complete survey is expected to contain 1 million galaxies of which approximately 580,000 are contained in the second incremental data release made public in March 2000. Concerning details about the 2dFGRS, we refer here to chapter 2 and mention that we use the 130,000 redshifts obtained prior to September 2000.

The overlap of the two surveys is very good. There are some gaps in the sky coverage due to strips of the sky that were not included in the 2MASS second incremental release, but overall a substantial fraction of the 2151.6 deg^2 of the 2dFGRS is covered by 2MASS. The homogeneity and extensive sky coverage of the combined dataset make it ideal for studies of the statistical properties of the galaxy population.

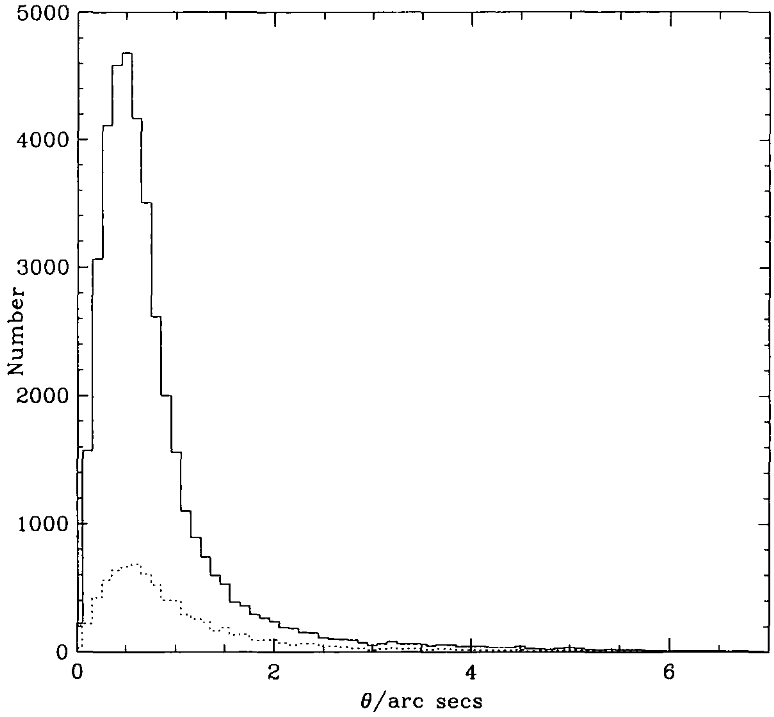


Figure 4.1: The distribution of angular separation, θ, for matched 2MASS–2dFGRS galaxies. The solid histogram is the distribution for the whole catalogue and the dotted histogram for the subset of 2MASS galaxies with semi-major axes larger than 12 arcsec.

4.2.2 Completeness of the matched 2MASS–2dFGRS catalogue

Here we consider whether all the 2MASS galaxies within the 2dFGRS survey region have 2dFGRS counterparts and assess the extent to which the fraction of galaxies with measured redshifts represents an unbiased sub-sample.

The astrometry in both 2MASS and 2dFGRS is, in general, very good and it is an easy matter to match objects in the two catalogues. We choose to find the closest pairs within a search radius equal to three quarters of the semi-major axis of the J-band image (denoted j_r_e in the 2MASS database). Scaling the search radius in this way helps with the matching of large extended objects. This procedure results in the identification of 2dFGRS counterparts for 40,121 of the 2MASS objects, when at random one would only expect to find a handful of such close pairs. Moreover, the distribution of separations shown in Fig. 4.1 peaks at 0.5 arcsec, with only 3% having separations greater than 3 arcsec. A significant part of this tail comes from the most extended objects as is evident from the dotted histogram in Fig. 4.1 which shows objects with semi-major axes larger than 12 arcsec. Thus, we can be very confident in these identifications.

The 40,121 2MASS objects for which we have found secure 2dFGRS counterparts amount to 88.6% of the 2MASS extended sources that fall within the boundary of the 2dFGRS. As discussed below, a more restrictive criterion that includes only sources fainter than J=12 that are confidently classified as galaxies by 2MASS, increases the fraction with 2dFGRS matches to 90.7%. The remaining 9.3% are missed for well understood reasons (star-galaxy classification: 4.6%; merged or close images: 4.4%; miscellaneous: 0.27%), none of which ought to introduce a bias. This is confirmed explicitly, in the middle row of Fig. 4.2, by the close correspondence between the photometric properties of the missed 9.3% and those of the larger matched sample. Hence, in estimating luminosity functions no significant bias will be introduced by assuming the matched sample to be representative of the full population. Furthermore, the distribution shown in the bottom row of Fig. 4.2 shows that the subset of 17,173 galaxies for which we have measured redshifts is a random sample of the full matched 2MASS-2dFGRS catalogue. This summary is the result of a thorough investigation, which we describe in the remainder of this section, into the reasons why 11.4% of the 2MASS sources are missed and whether their omission introduces a bias in the properties of the matched sample.

We first consider objects in the 2MASS catalogue which based on their images and colours are not confidently classified as galaxies. In the 2MASS database a high e_score or g_score indicates a high probability that the object is either not an extended source or not a galaxy. A cc_flag$\neq$0 indicates an artifact or contaminated and/or confused source. For detailed definitions of these parameters we refer the reader to Jarrett et al. (2000). Rejecting all objects which have either e_score>1.4, g_score>1.4 or cc_flag$\neq$0 removes just 6.7% of the total. However, removing these reduces significantly the fraction of the 2MASS sample that does not match with the 2dFGRS catalogue, from 11.4% to 9.6%. Thus, it is likely that about 30% of the 2MASS objects which have e_score>1.4, g_score>1.4 or cc_flag$\neq$0 are not galaxies.

The 2MASS may contain a tail of very red objects that are too faint in the b_J-band to be included in the $b_J<19.45$ 2dFGRS sample. Fig. 4.3 shows the distribution of b_J-J colours for the matched objects with J<14.7. (Here, the J-band magnitude we are using is the default magnitude denoted j_m in the 2MASS database. In section 4.2.3 we will consider the issue of what magnitude definition is most appropriate for estimating the luminosity function.) The vertical dashed line indicates the colour at which this sample starts to become incomplete due to the $b_J<19.45$ magnitude limit of 2dFGRS. The colour distribution cuts off sharply well before this limit, suggesting that any tail of missed very red objects is extremely small. In other words the 2dFGRS is sufficiently deep that even

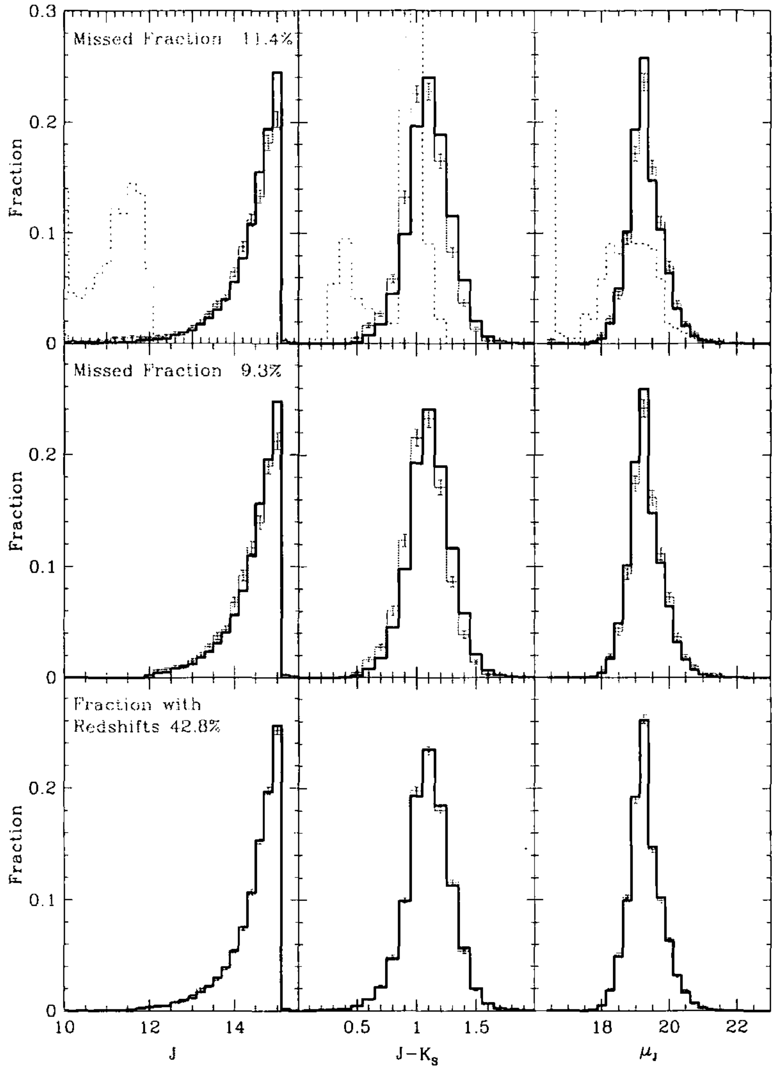


Figure 4.2: The distribution of J-band apparent magnitude, J$-$K$_S$ colour and J-band surface brightness, μ_J, for various sub-samples of the 2MASS catalogue. Here, the measure of surface brightness used is simply $\mu_J \equiv J - 5\log_{10} r$, where J is the Kron magnitude and r the Kron semi-major axis in arc seconds (j_m_e and j_r_e in the 2MASS database). In all three rows, the thick solid histograms are the distributions for 2MASS objects that are matched with 2dFGRS galaxies. The light solid histograms in the top row are the 11.4% of 2MASS galaxies that are not matched with 2dFGRS galaxies. Poisson errorbars are shown on these histograms. The dashed histograms are for the bright sub-sample with J<12. In the middle row, the light histograms show the distributions for the 9.3% of 2MASS galaxies fainter than J=12 and satisfying the additional image classification constraints discussed in the text that are not matched with 2dFGRS galaxies. In the bottom row, the light histograms show the distributions of the 42.8% of the matched 2MASS-2dFGRS galaxies for which redshifts have been measured. The values in each histogram are the fraction of the corresponding sample that falls in each bin.

the reddest objects detected at the faintest limits of 2MASS ought to be detected in 2dFGRS.

In the top row of Fig. 4.2 we compare the distributions of magnitude, colour and surface brightness for the matched and missed 2MASS objects. In general, the properties of the missed subset overlap well with those of the much larger matched subset. However, we do see that the distributions for missed objects contain tails of bright and blue objects. It is quite likely that this is due to the 2MASS extended source catalogue being contaminated by a small population of saturated or multiple stars. The dotted histograms in the top row of Fig. 4.2 show the distributions of magnitude, colour and surface brightness for the bright subset of the missed objects with $J<12$. Here we clearly see bimodal colour and surface brightness distributions. The blue peak of the colour distribution is consistent with that expected for stars (see Jarrett et al. (2000)). Excluding these bright, $J<12$, objects which are clearly contaminated by stars reduces the fraction of missed 2MASS objects from 9.6% to 9.3%. The magnitude, colour and surface brightness distributions for this remaining 9.3% are shown in the middle panel of Fig. 4.2. We see that the missed objects are slightly under-represented at the faintest magnitudes and also slightly bluer on average than the matched sample, while the distribution of surface brightness is almost indistinguishable for the two sets of objects. These differences are small and so will introduce no significant bias in our luminosity function estimates.

To elucidate the reasons for the remaining missed 9.3% of 2MASS objects we downloaded 100 1×1 arcmin images from the STScI Digitized Sky Survey (DSS) centred on the positions of a random sample of the missed 2MASS objects. In each image we plotted a symbol to indicate the position of any 2dFGRS galaxies within the 1×1 arcmin field. We also plotted symbols to indicate the positions and classifications of all images identified in the APM scans from which the 2dFGRS catalogue was drawn, down to a magnitude limit of $b_J \approx 20.5$. These images are classified as galaxies, stars, merged images (galaxy+galaxy, galaxy+star or star+star) or noise. This set of plots allows us to perform a census of the reasons why some 2MASS objects are not present in the 2dFGRS survey.

The main cause for the absence of 2MASS objects in the 2dFGRS is that the APM has classified these objects as stars. These amount to 49.5% of the missed sample (4.6% of the full 2MASS sample). In some cases, the DSS image shows clearly that these are stars and in others that they are galaxies. However, the majority of these objects cannot easily be classified from the DSS images. Thus, they could be galaxies that the APM has falsely classified as stars or stars that 2MASS has falsely classified as galaxies. The first possibility

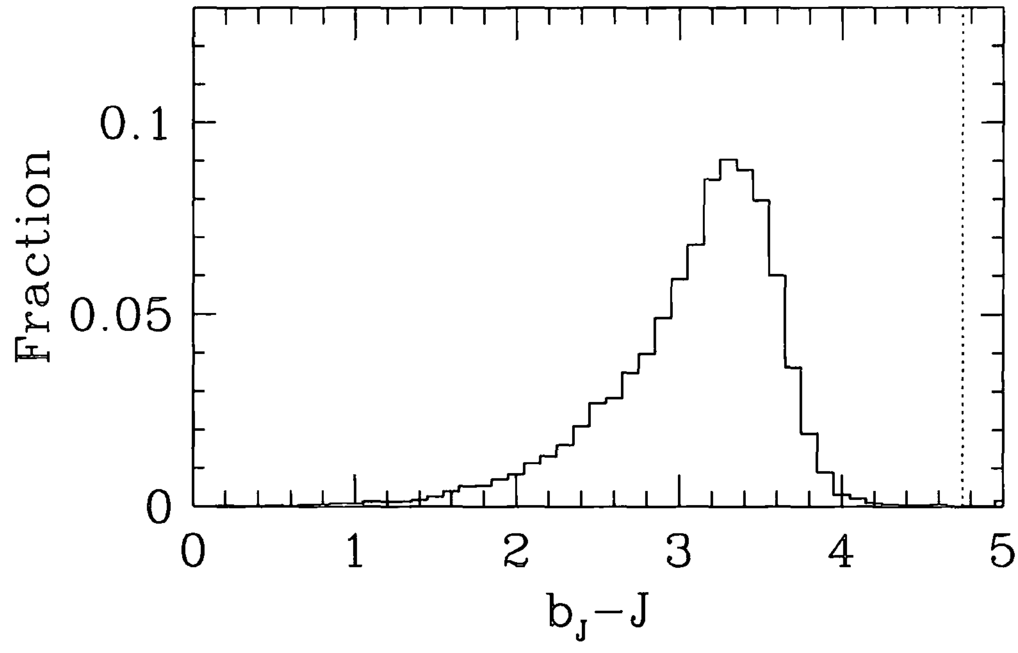


Figure 4.3: The solid histogram shows the distribution of b_J-J colours for 2MASS galaxies selected to have J<14.7. (Here, we use the 2MASS default magnitude, denoted j_m in the 2MASS database.) The vertical dashed line indicates the colour at which this sample starts to become incomplete due to the b_J<19.45 magnitude limit of 2dFGRS.

is not unexpected since the parameters used in the APM star-galaxy separation algorithm were chosen as a compromise between high completeness and low contamination such that the expected completeness is around 95% with 5% stellar contamination (Maddox et al. 1990b). It is hard to rule out the possibility that this class of object does not include a substantial fraction of stars, but if so, their presence appears not to distort the distribution of colours shown in Fig 4.3. Another 47.6% of the random sample (4.4% of the full 2MASS sample) are classified by the APM as mergers or else consist of two close images in the DSS but are classified by the APM as a single galaxy offset from the 2MASS position. The remaining 2.9% of the random sample (0.27% of the full 2MASS sample) are missed for a variety of reasons including proximity to the diffraction spikes of very bright stars and poor astrometry caused by the presence of a neighbouring unclassified image.

4.2.3 2MASS magnitude definitions and calibration

The 2MASS extended source database provides a large selection of different magnitude measurements. In the previous section we used the default magnitudes (denoted j_m and

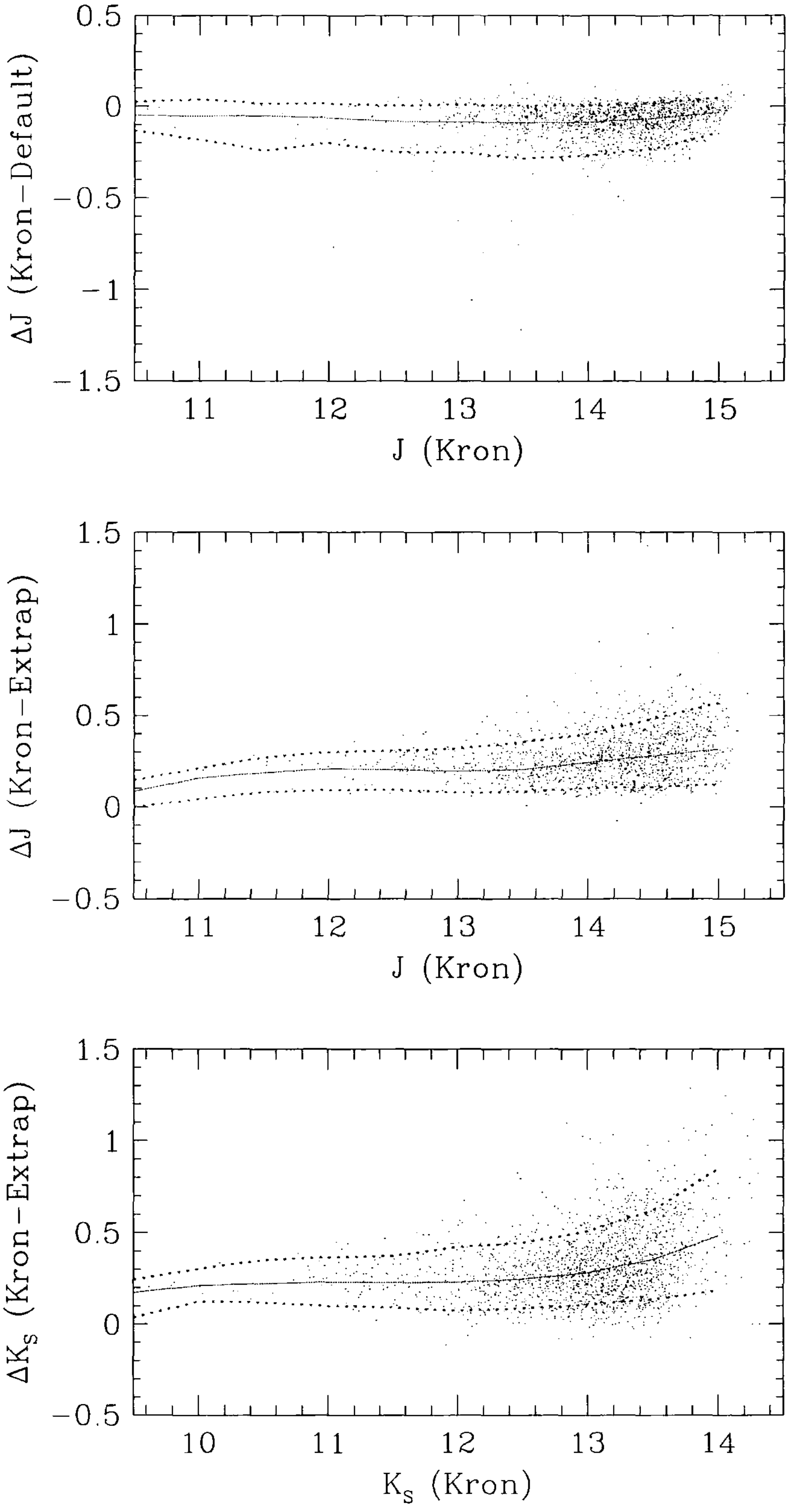


Figure 4.4: A comparison of the 2MASS default, Kron and extrapolated magnitudes in the J and K_S bands. The dots are the measured values for each of the galaxies in the matched 2MASS-2dFGRS catalogue. The solid and dotted lines indicate the median, 10 and 90 percentiles of the distribution.

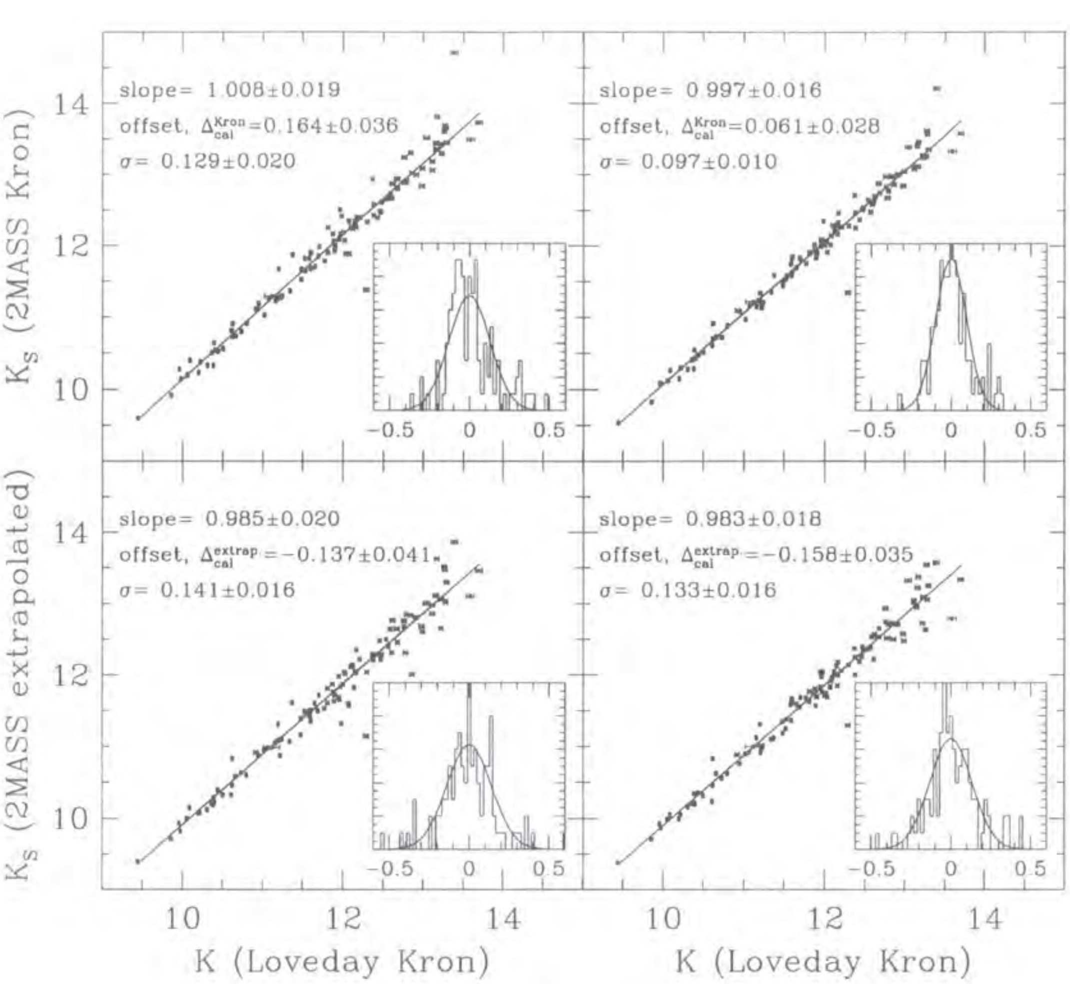


Figure 4.5: Comparison of 2MASS Kron and extrapolated magnitudes with the independent measurements of Loveday (2000). The left hand panels are for the K_S-band Kron and extrapolated magnitudes (k_m_e and k_m_ext in the 2MASS database). The right hand panels show Kron and extrapolated magnitudes inferred from the 2MASS J-band Kron and extrapolated magnitudes and the measured default aperture $J-K_S$ colours (j_m_e−j_m+k_m and j_m_ext−j_m+k_m in the 2MASS database variables). The horizontal errorbars show the measurement errors quoted by Loveday (2000). The solid lines show simple least squares fits. The slopes and zero-point offsets of these fits and the rms residuals about the fits are indicated on each panel. The inset plots show the distribution of residual magnitude differences.

k_m in the 2MASS database). These are magnitudes defined within the same circular aperture in each waveband. For galaxies brighter than K_S=14, the aperture is the circular K_S-band isophote of 20 mag arcsec^{-2} and for galaxies fainter than K_S=14 it is the circular J-band isophote of 21 mag arcsec^{-2}. These are not the most useful definitions of magnitude for determining the galaxy luminosity function. Since we are interested in measuring the total luminosity and ultimately the total stellar mass of each galaxy, we require a magnitude definition that better represents the total flux emitted by each galaxy. We consider Kron magnitudes (Kron 1980) and extrapolated magnitudes. Kron magnitudes (denoted j_m_e and k_m_e in the 2MASS database) are measured within an aperture, the Kron radius, defined as 2.5 times the intensity-weighted radius of the image. The extrapolated magnitudes (denoted j_m_ext and k_m_ext in the 2MASS database) are defined by first fitting a modified exponential profile, $f(r) = f_0 \exp[-(\alpha r)^{1/\beta}]$, to the image from 10 arcsec to the 20 mag/arcsec^2 isophotal radius, and extrapolating this from the Kron radius to 4 times this radius or 80 arcsec if this is smaller (Jarrett private communication). Note that improvements are being made to the extended source photometry algorithms developed and employed 2MASS team and so in the final 2MASS data release the definitions of the Kron and extrapolated magnitudes may be slightly modified (Jarrett private communication).

Fig. 4.4 compares the default, Kron and extrapolated magnitudes in the J and K_S bands for the matched 2MASS–2dFGRS catalogue. The upper panel shows that while the median offset between the J-band isophotal default magnitudes and the pseudo-total Kron magnitudes is small there is a large spread with some galaxies having default magnitudes more than 0.5 magnitudes fainter than the Kron magnitude. The Kron magnitudes are systematically fainter than the extrapolated magnitudes by between approximately 0.1 and 0.3 magnitudes. This offset is rather larger than expected: if the Kron radius is computed using a faint isophote to define the extent of the image from which the intensity weighted radius is measured, then the Kron magnitudes should be very close to total. For an exponential light profile ($\beta = 1$), the Kron radius should capture 96% of the flux, while for an $r^{1/4}$ law ($\beta = 4$), 90% of the flux should be enclosed. In other words, the Kron magnitude should differ from the total magnitude by only 0.044 and 0.11 magnitudes in these two cases. However, the choice of isophote is a compromise between depth and statistical robustness. In the case of the 2MASS second incremental release, an isophote of 21.7(20.0) mag arcsec^{-2} in J(K_S) was adopted (Jarrett private communication). These relatively bright isophotes, particularly the K_S-band isophote, could lead to underestimates of the Kron radii and fluxes for lower surface brightness objects and plausibly accounts for much of the median offset of 0.3 magnitudes seen in Fig. 4.4 between the K_S-band Kron and extrapolated magnitudes. This line of reasoning favours adopting the extrapolated magnitudes as the best estimate of the total magnitudes, but, on the other hand, the extrapolated magnitudes are model-dependent and have larger measurement errors.

To understand better the offset and scatter in the 2MASS magnitudes we have compared a subset of the 2MASS data with the independent K-band photometry of Loveday (2000). The pointed observations of Loveday have better resolution than the 2MASS images and good signal-to-noise to a much deeper isophote. This enables accurate, unbiased Kron magnitudes to be measured. Note that the offset between the 2MASS K_S-band and the standard K-band used by Loveday is expected to be almost completely negligible (Carpenter 2001). The left hand panels of Fig. 4.5 compare these measurements with the corresponding 2MASS Kron and extrapolated magnitudes. The right hand panels show K_S-band Kron and extrapolated magnitudes computed by taking the 2MASS J-band Kron and extrapolated magnitudes and subtracting the $J-K_S$ colour measured within the default aperture. These indirect estimates are interesting to consider as they combine the profile information from the deeper J-band image with the $J-K_S$ colour measured within the largest aperture in which there is good signal-to-noise. The straight lines plotted in Fig. 4.5 show simple least squares fits and the slope and zero-point offset of these fits are indicated on each panel along with bootstrap error estimates. Also shown in the inset panels is the distribution of residual magnitude differences about each of the fits and a gaussian fit to this distribution. The rms of these residuals and a bootstrap error estimate is also given in each panel.

From these comparisons we first see that all the fits have slopes entirely consistent with unity, but that their zeropoints and scatters vary. The zero-point offsets, $\Delta_{\rm cal}^{\rm Kron}$, between both the 2MASS Kron magnitude measurements and those of Loveday confirm that the 2MASS Kron magnitudes systematically underestimate the galaxy luminosities. In the case of the direct K_S-band 2MASS magnitudes the offset is $\Delta_{\rm cal}^{\rm Kron} = 0.164$ magnitudes. In the case of the Kron magnitudes inferred from the deeper J-band image profiles, the offset is reduced to $\Delta_{\rm cal}^{\rm Kron} = 0.061$ magnitudes. Conversely the 2MASS extrapolated magnitudes are systematically brighter than the Loveday Kron magnitudes by $-\Delta_{\rm cal}^{\rm extrap.} = 0.137$ and 0.158 magnitudes, where one would expect an offset of only $\Delta_{\rm Kron} = 0.044$ to 0.11 due to the difference in definition between ideal Kron and true total magnitudes. For both estimates of the extrapolated magnitude and for the directly estimated Kron magnitude the scatter about the correlation is approximately 0.14 magnitudes and we note a slight tendency for the scatter to increase at faint magnitudes. The magnitude estimate that best correlates with the Loveday measurements is the Kron magnitude estimated from the 2MASS J-band Kron magnitude and the default aperture $J-K_S$ colour. Here the distribution of residuals has a much reduced scatter of only 0.1 magnitudes and has very few outliers.

Our conclusion from the comparison of Kron magnitudes is that it is preferable to adopt the K_S-band magnitude inferred from the J-band Kron or extrapolated magnitude by converting to the K_S-band using default aperture colour, rather than to use the noisier and more biased direct K_S-band estimates. With this definition, we find that the 2MASS Kron magnitudes slightly underestimate the galaxy luminosities while the extrapolated magnitudes slightly overestimate the luminosities, particularly at faint fluxes. We will present results for both magnitude definitions, but we note that to convert to total magnitudes we estimate that the 2MASS Kron magnitudes should be brightened by $\Delta_{\rm cal}^{\rm Kron} + \Delta_{\rm Kron} = 0.1$–$0.17$ magnitudes and the extrapolated magnitudes dimmed by $-\Delta_{\rm cal}^{\rm extrap.} - \Delta_{\rm Kron} = 0.05$–$0.11$ magnitudes.

4.2.4 Completeness of the 2MASS catalogue

Here we define the magnitude limited samples which we will analyze in section 4.4 and test them for possible incompleteness in both magnitude and surface brightness. For the Kron and extrapolated magnitudes, the 2MASS catalogue has high completeness to the nominal limits of $J<14.7$ and $K_S<13.9$. However, to ensure very high completeness and avoid any bias in our luminosity function estimates, we made the following more conservative cuts. For the Kron magnitudes, we limited our sample to either $J<14.45$ or $K_S<13.2$, and for the extrapolated magnitudes to either $J<14.15$ or $K_S<12.9$. These choices are motivated by plots such as the top panel of Fig. 4.4. Here the isophotal default magnitude limit of $J<14.7$ is responsible for the right hand edge to the distribution of data points. One sees that this limit begins to remove objects from the distribution of Kron magnitudes for $J\gtrsim14.5$. An indication that the survey is complete to our adopted limits is given by the number counts shown in Fig 4.6, which only begin to roll over at fainter magnitudes.

More rigorously, we have verified that the samples are complete to these limits by examining their $V(z_i)/V(z_{{\rm max},i})$ distributions. Here, z_i is the redshift of a galaxy in the sample, $z_{{\rm max},i}$ is the maximum redshift at which this galaxy would satisfy the sample

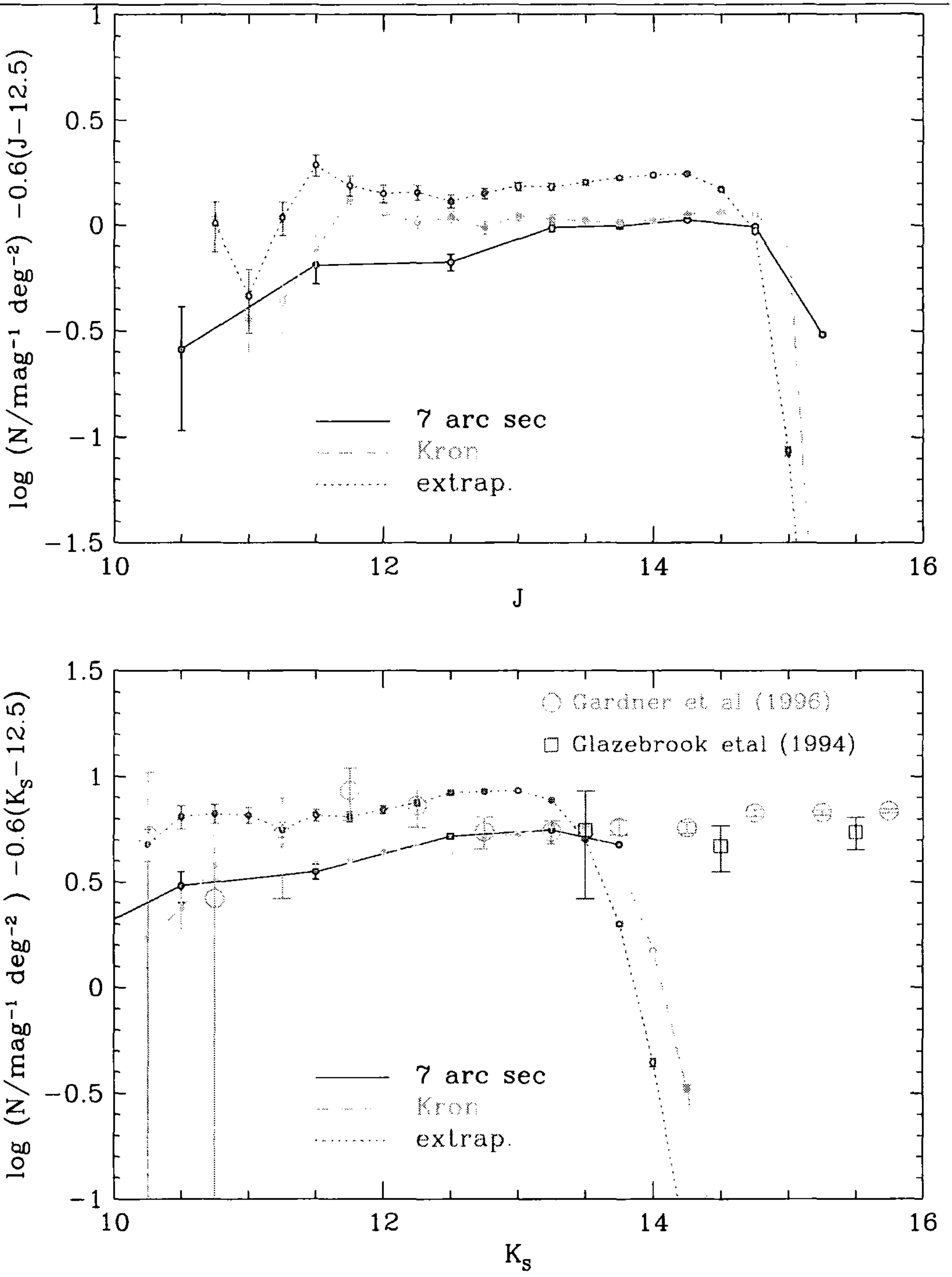


Figure 4.6: Differential galaxy number counts in the J and K_S bands, all with Poisson errorbars and with a Euclidean slope subtracted so as to expand the scale of the ordinate. The J and K_S counts linked by the solid line are the 2MASS 7 arcsec aperture counts of Jarrett et al. (in preparation). The counts linked by the dashed and dotted lines are those of the 2MASS–2dFGRS redshift catalogue for Kron and extrapolated magnitudes respectively. The K_S-band magnitudes are those inferred from the J-band magnitudes and aperture colours. In the K_S-band these are compared with the counts of Gardner et al. (1996) and Glazebrook et al. (1994) as indicated in the figure legend.

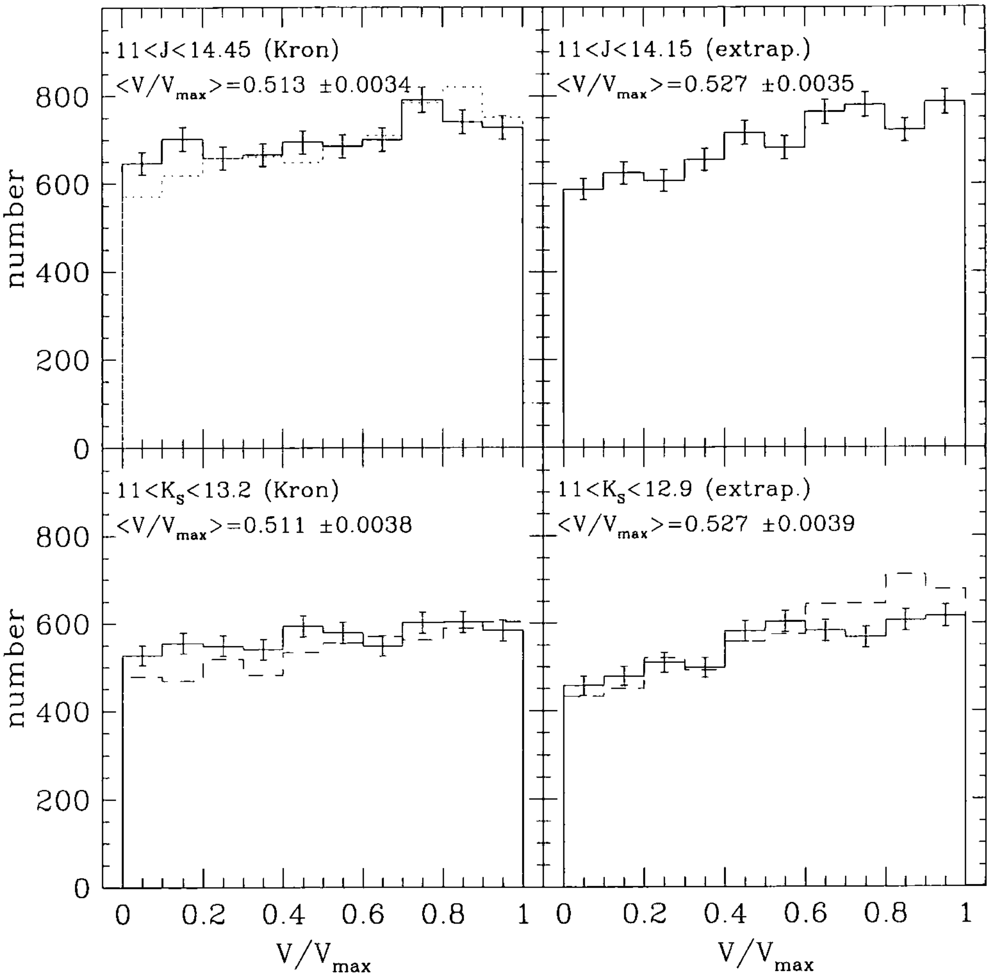


Figure 4.7: The distributions of $V/V_{\rm max}$ for our magnitude limited samples. The solid histograms in the four panels show the $V/V_{\rm max}$ distributions for our J and K$_{\rm S}$ Kron and extrapolated magnitude limited samples. The mean values of $\langle V/V_{\rm max}\rangle$ are indicated on each panel. The K$_{\rm S}$-band magnitudes are those inferred from the J-band magnitudes and aperture colours. The distributions for the directly measured K$_{\rm S}$-band Kron and extrapolated magnitudes are shown by the dashed histograms in the lower panels. The dotted histogram in the top-left panel shows the $V/V_{\rm max}$ distribution we obtain when attempting to take account of the 2MASS isophotal diameter and isophotal magnitude limits in estimating the $V_{\rm max}$ values.

selection criteria, and $V(z)$ is the survey volume that lies at redshift less than z. If the sample is complete and of uniform density, $V(z_i)/V(z_{\mathrm{max},i})$ is uniformly distributed within the interval 0 to 1. To evaluate z_{max} we made use of the default k+e corrections described in the following section, but the results are not sensitive to reasonable variations in the assumed corrections or in the cosmology. The solid histograms in Fig. 4.7 show these distributions for each of our four magnitude limited samples. Note that the $\mathrm{K_S}$-band magnitudes are those inferred from the J-band magnitudes and aperture colours. The dashed histograms in the lower panels show the corresponding distributions for the directly measured $\mathrm{K_S}$-band magnitudes. In all these cases we have computed V_{max} simply from the imposed apparent magnitude limits and have ignored any possible dependence of the catalogue completeness on surface brightness.

If the samples were incomplete the symptom one would expect to see is a deficit in the V/V_{max} distributions at large V/V_{max} and hence a mean $\langle V/V_{\mathrm{max}}\rangle<0.5$. There is no evidence for such a deficit in these distributions. In fact each has a mean $\langle V/V_{\mathrm{max}}\rangle$ slightly greater than 0.5. The slight gradient in the V/V_{max} distribution is directly related to the galaxy number counts shown in Fig 4.6, which are slightly steeper than expected for a homogeneous, non-evolving galaxy distribution. A similar result has been found in the bright $\mathrm{b_J}$-band counts (Maddox et al. 1990d). The $\mathrm{b_J}$-band result has variously been interpreted as evidence for rapid evolution, systematic errors in the magnitude calibration, or a local hole or underdensity in the galaxy distribution (Maddox et al. 1990d, Metcalfe, Fong & Shanks 1995, Shanks 1990). Here we note that the gradient in the V/V_{max} distributions (and also in the galaxy counts) becomes steeper both as one switches from Kron to the less reliable extrapolated magnitudes and as one switches from the J-band data to the lower signal-to-noise $\mathrm{K_S}$-band data. This gives strong support to our decision to adopt the $\mathrm{K_S}$-band magnitudes derived from the J-band Kron and extrapolated magnitudes and aperture $\mathrm{J}-\mathrm{K_S}$ colours. It also cautions that the mean $\langle V/V_{\mathrm{max}}\rangle>0.5$ cannot necessarily be taken as a sign of evolution or a local underdensity, but may instead be related to the accuracy of the magnitude measurements. The comparison to the observations of Loveday (2000) shows no evidence for systematic errors in the magnitudes, but does not constrain the possibility that the distribution of magnitude measurement errors may become broader or skewed at fainter magnitudes. Such variations would affect the V/V_{max} distributions and could produce the observed behaviour. We conclude by noting that while the shift in the mean $\langle V/V_{\mathrm{max}}\rangle$ is statistically significant, it is nevertheless quite small for the samples we analyze and has little effect on the resulting luminosity

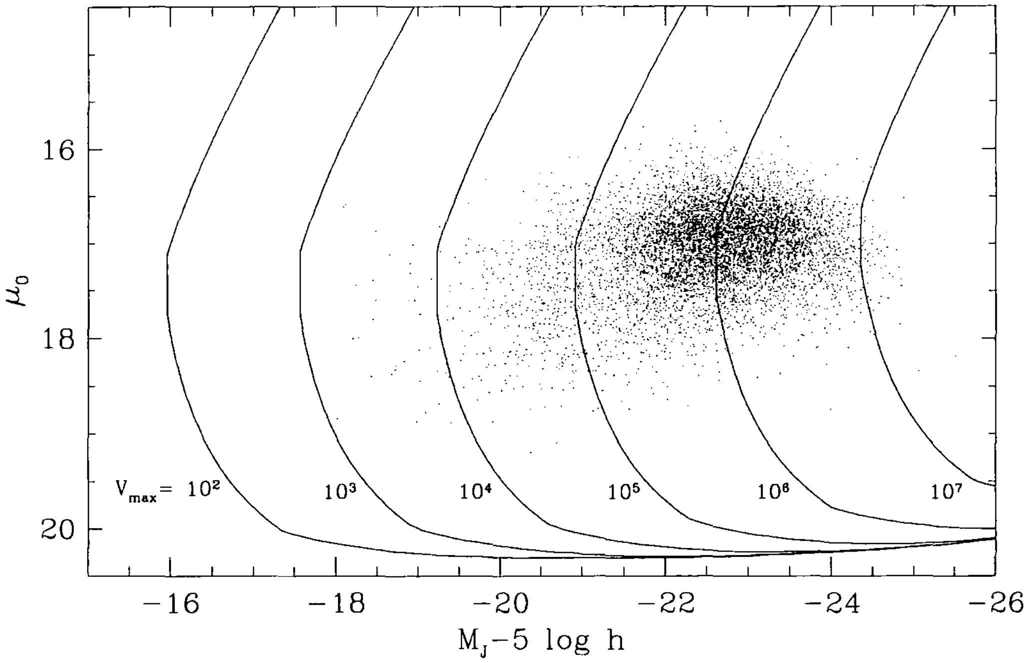


Figure 4.8: The points show the distribution of estimated central surface brightness, μ_0, and absolute magnitude, M_J, for our J< 14.45 (Kron) sample. The contours show visibility theory estimates of $V_{\rm max}$ as a function of μ_0 and M_J. The contours are labelled by their $V_{\rm max}$ values in units of $({\rm Mpc/h})^3$.

function estimates.

We now investigate explicitly the degree to which the completeness of the 2MASS catalogue depends on surface brightness by estimating $V_{\rm max}$ as a function of both absolute magnitude and surface brightness. This is an important issue: if the catalogue is missing low-surface brightness galaxies our estimates of the luminosity function will be biased. The approach we have taken follows that developed in Cross et al. (2001) for the 2dFGRS. We estimate an effective central surface brightness, μ_0^z, for each observed galaxy assuming an exponential light distribution, that the Kron magnitudes are total and that the Kron radii are exactly five exponential scale-lengths. This is then corrected to redshift $z = 0$ using

$$\mu_0 = \mu_0^z - 10\log(1+z) - k(z) - e(z) \tag{4.1}$$

to account for redshift dimming[†] and k+e corrections (c.f section 4.3). Note that in the

---

[†] The observed surface brightness is predicted to vary with redshift as $\mu \sim (1+z)^{-4}$, where two powers of $1+z$ come from distortion, that opens the beam solid angle, one from the dilation in the rate of reception of photons, and one from the loss of energy per photon (Peebles 1993).

2MASS catalogue, galaxies with estimated Kron radii less than 7 arcsec, have their Kron radii set to 7 arcsec. This will lead us to underestimate the central surface brightnesses of these galaxies, but this will only affect high surface brightness objects and will not affect whether a galaxy can or cannot be seen. The distribution in the M_J-μ_0 plane of our Kron J-band selected sample is shown by the points in Fig. 4.8.

Cross et al. (2001) use two different methods to estimate the value of $V_{\rm max}$ associated with each position in this plane. The first method uses the visibility theory of Phillipps et al. (1990). We model the selection characteristics of the 2MASS extended source catalogue by a set of thresholds. The values appropriate in the J-band are a minimum isophotal diameter of 8.5 arcsec at an isophote of 20.5 mag arcsec^{-2}, and an isophotal magnitude limit of J<14.7 at an isophote of 21.0 mag arcsec^{-2} Jarrett et al. (2000). In addition, we impose the limits in the Kron magnitude of 11<J<14.45 that define the sample we analyze. We then calculate for each point on the M_J-μ_0 plane the redshift at which a such a galaxy will drop below one or other of these selection thresholds and hence compute a value of $V_{\rm max}$. The results of this procedure are shown by the contours of constant $V_{\rm max}$ plotted in Fig. 4.8. Note that these estimates of $V_{\rm max}$ are only approximate since we have made the crude assumption that all the galaxies are circular exponential disks. In addition, the diameter and isophotal limits are only approximate and vary with observing conditions.

The second method developed by Cross et al. (2001) consists of making an empirical estimate of $V_{\rm max}$ in bins in the M_J-μ_0 plane. They look at the distribution of observed redshifts in a given bin and adopt the 90$^{\rm th}$ percentile of this distribution to define $z_{\rm max}$ and hence $V_{\rm max}$. It is more robust to use the 90$^{\rm th}$ percentile rather than the 100$^{\rm th}$ percentile and the effect of this choice can easily be compensated for when estimating the luminosity function (Cross et al. 2001). Note that in our application to the 2MASS data we do not apply corrections for incompleteness or the effects of clustering. The result of this procedure is to confirm that for the populated bins, the $V_{\rm max}$ values given by the visibility theory are a good description of the data.

In Fig. 4.8 we see that the distribution of galaxies in the M_J-μ_0 plane is well separated from the low surface brightness limit of approximately 20.5 mag arcsec^{-2} where the $V_{\rm max}$ contours indicate that the survey has very little sensitivity. Thus, there is no evidence that low-surface brightness galaxies are missing from the 2MASS catalogue. Furthermore, in the region occupied by the observed data, the $V_{\rm max}$ contours are close to vertical indicating that there is little dependence of $V_{\rm max}$ on surface brightness. The way in which the $V/V_{\rm max}$

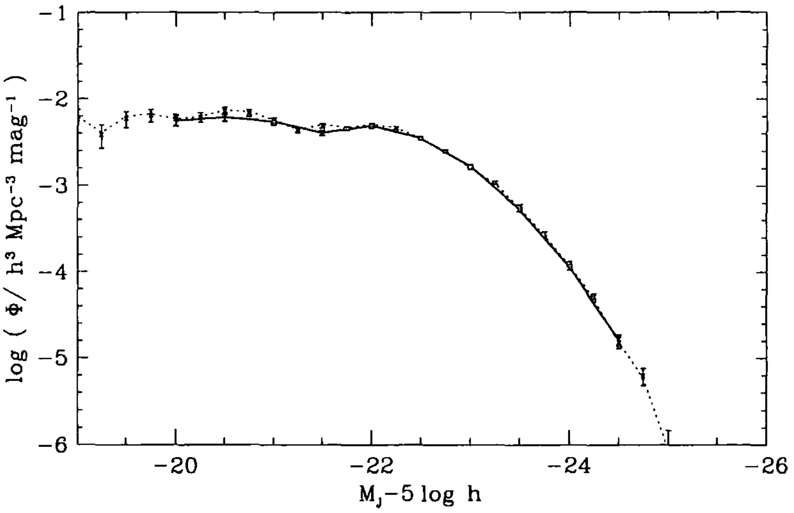


Figure 4.9: Three $1/V_{\rm max}$ estimates of the Kron J-band luminosity function. The data points with errorbars show the estimate based on assuming that $V_{\rm max}$ depends only upon absolute magnitude and ignoring any possible surface brightness dependence. The dotted line and heavy solid line show the estimates in which the surface brightness dependence of $V_{\rm max}$ is derived from visibility theory and from the empirical method of Cross et al. (2000) respectively.

distribution is modified by including this estimate of the surface brightness dependence is shown by the dotted histogram in the top-left panel of Fig. 4.7. Its effect is to increase the mean $V/V_{\rm max}$ slightly, suggesting that this estimate perhaps overcorrects for the effect of surface brightness selection. Even so, the change in the estimated luminosity function is negligible as confirmed by the three estimates of the Kron J-band luminosity function shown in Fig. 4.9. These are all simple $1/V_{\rm max}$ estimates, but with $V_{\rm max}$ computed either ignoring surface brightness effects or using one of the two methods described above. These luminosity functions differ negligibly, indicating that no bias is introduced by ignoring surface brightness selection effects.

4.3 Modelling the Stellar Populations

The primary aim of this chapter is to determine the present-day J and K_S-band luminosity functions and also the stellar mass function of galaxies. Since the 2MASS survey spans a range of redshift (see Fig. 4.10), we must correct for both the redshifting of the filter bandpass (k-correction) and for the effects of galaxy evolution (e-correction). In practice,

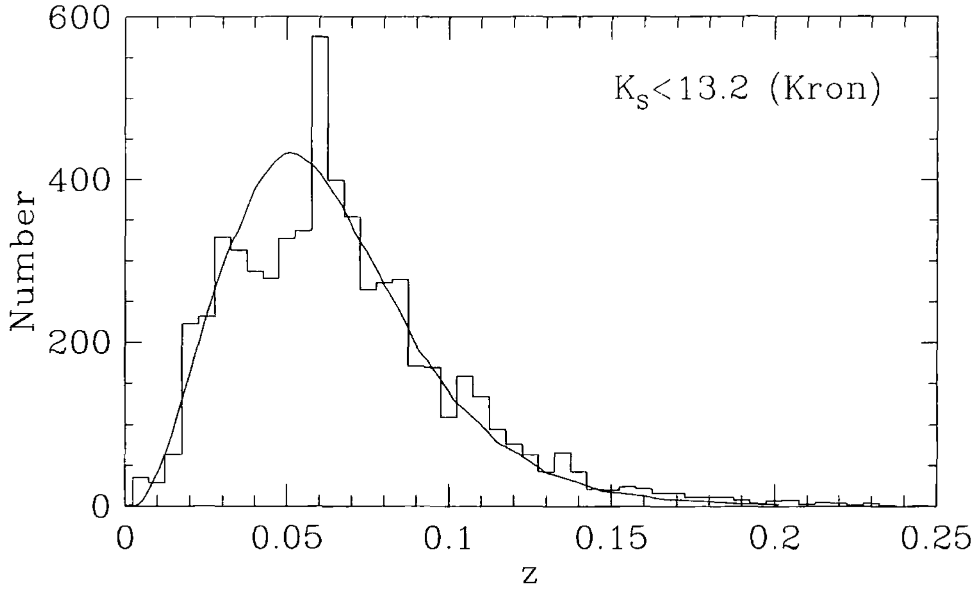


Figure 4.10: The redshift distribution of the K_S<13.2 (Kron) sample selected from the matched 2MASS–2dFGRS catalogue. The smooth curve is the model prediction based on the SWML estimate of the K_S-band luminosity function (c.f. section 4.4). The model prediction is very insensitive to the assumed k+e correction and cosmology.

the k and e-corrections at these wavelengths are both small and uncertainties in them have little effect on the estimated luminosity functions. This is because these infrared bands are not dominated by young stars and also because the 2MASS survey does not probe a large range of redshift. We have chosen to derive individual k and e-corrections for each galaxy using the stellar population synthesis models of Bruzual & Charlot (1993) and Bruzual & Charlot (2001 in preparation). We have taken this approach not because such detailed modelling is necessary to derive robust luminosity functions, but because it enables us to explore the uncertainties in the derived galaxy stellar mass functions, which are, in fact, dominated by uncertainties in the properties of the stellar populations.

The latest models of Bruzual & Charlot (2001 in preparation) provide, for a variety of different stellar initial mass functions (IMFs), the spectral energy distribution (SED), $l_\lambda(t, Z)$, of a single population of stars formed at the same time with a single metallicity, as a function of both age, t, and metallicity, Z. We convolve these with an assumed star formation history, $\psi(t')$, to compute the time-evolving SED of the model galaxy,

$$L_\lambda(t) = \int_0^t l_\lambda(t - t', Z)\, \psi(t')\, dt'. \qquad (4.2)$$

We take account of the effect of dust extinction on the SEDs using the Ferrara et al.

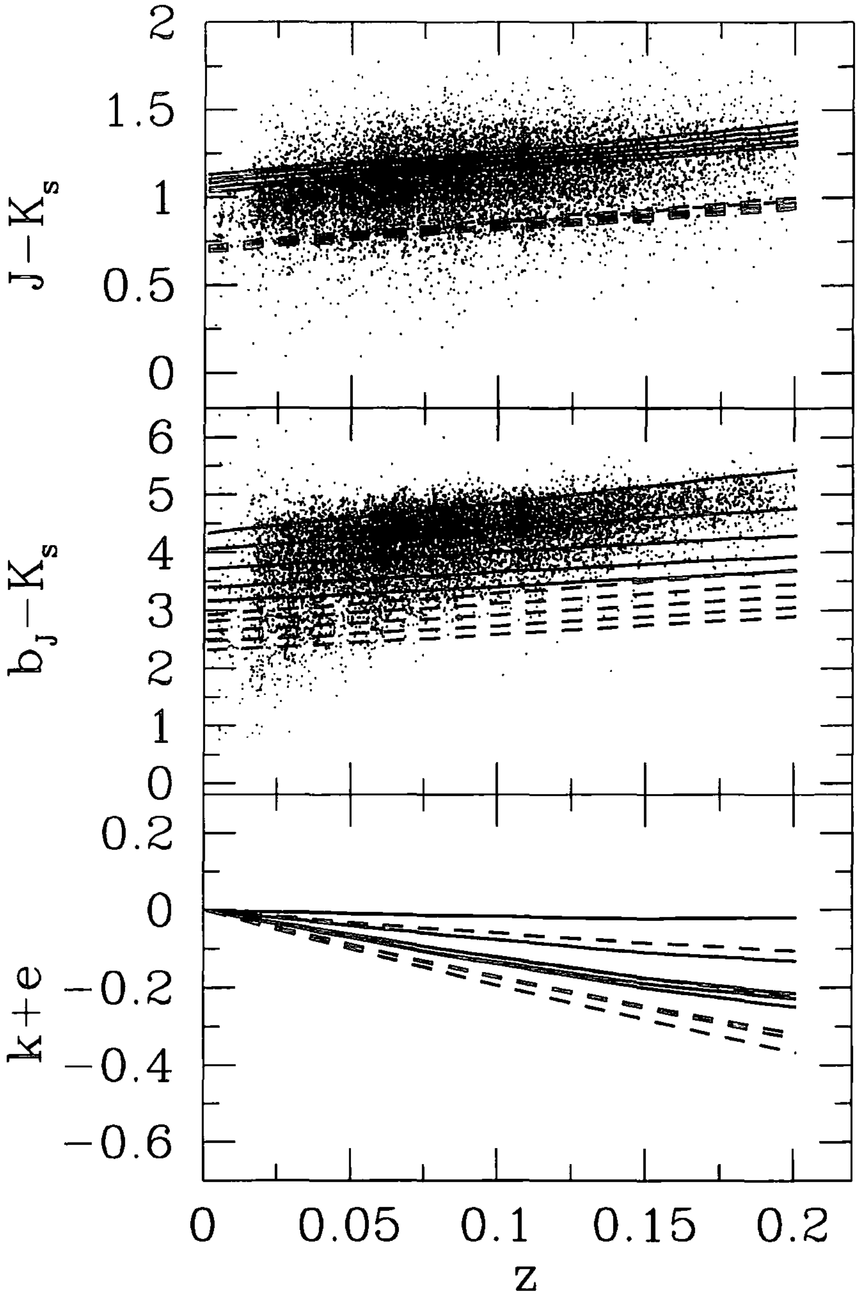


Figure 4.11: The points in the upper two panels show the observed distributions of J-K$_S$ and b$_J$-K$_S$colours as a function of redshift for our matched 2MASS-2dFGRS catalogue with $z < 0.2$ and J< 14.45 (Kron). Overlaid on these points are some examples of model tracks. The solid curves are for solar metallicity, $Z = 0.02$, and the dashed curves for $Z = 0.004$. Within each set, the tracks show different choices of the star formation time scale, τ. The grid of values we use has τ =1, 3, 5, 10 and 50 Gyr. Shorter values of τ lead to older stellar populations and redder colours. The bottom panel shows the k+e corrections in the J-band for these same sets of tracks.

(1999) extinction model normalized so that the V-band central face-on optical depth of the Milk-Way is 10. This value corresponds to the mean optical depth of $L_\star$ galaxies in the model of Cole et al. (2000) which employs the same model of dust extinction. We assume a typical inclination angle of 60 degrees which yields a net attenuation factor of 0.53 in the V-band and 0.78 in the J-band. By varying the assumed metallicity, Z, and star formation history, we build up a two-dimensional grid of models. Then, for each of these models, we extract tracks of $\mathrm{b_J}-\mathrm{K_S}$ and $\mathrm{J}-\mathrm{K_S}$ colours and stellar mass-to-light ratio as a function of redshift.

Our standard set of tracks assumes a cosmological model with $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$, Hubble constant $H_0 = 70$ km s^{-1} Mpc^{-1}, and star formation histories with an exponential form, $\psi(t) \propto \exp(-[t(z) - t(z_\mathrm{f})]/\tau)$. Here, $t(z)$ is the age of the Universe at redshift z and the galaxy is assumed to start forming stars at $z_\mathrm{f} = 20$. For these tracks, we adopt the Kennicutt IMF (Kennicutt 1983) and include the dust extinction model. The individual tracks are labelled by a metallicity, Z, which varies from $Z = 0.0001$ to $Z = 0.05$ and a star formation timescale, τ, which varies from $\tau = 1$ Gyr to $\tau = 50$ Gyr. Examples of these tracks are shown in Fig. 4.11, along with the observed redshifts and colours of the 2MASS galaxies. We can see that the infrared $\mathrm{J}-\mathrm{K_S}$ colour depends mainly on metallicity while the $\mathrm{b_J}-\mathrm{K_S}$ colour depends both on metallicity and star formation timescale. Thus, the use of both colours allows a unique track to be selected. Note from the bottom panel that, for all the tracks, the k+e correction is always small for the range of redshift spanned by our data.

We can gauge how robust our results are by varying the assumptions of our model. In particular, we vary the IMF, the dust extinction and cosmological models, and include or exclude the evolutionary contribution to the k+e correction. Also, we consider power-law star formation histories, $\psi(t) \propto [t(z)/t(z_\mathrm{f})]^{-\gamma}$, as an alternative to the exponential model. The results are discussed at beginning of section 4.5.

The procedure for computing the individual galaxy k+e corrections is straightforward. At the measured redshift of a galaxy, we find the model whose $\mathrm{b_J}-\mathrm{K_S}$ and $\mathrm{J}-\mathrm{K_S}$ colours most closely match that of the observed galaxy. Having selected the model we then follow it to $z = 0$ to predict the galaxy's present-day J and $\mathrm{K_S}$-band luminosities and also its total stellar mass. We also use the model track to follow its k+e correction to higher redshift in order to compute z_max, the maximum redshift at which this galaxy would have passed the selection criteria for inclusion into the analysis sample.

4.4 Luminosity Function Estimation

We use both the simple $1/V_{\rm max}$ method and standard maximum likelihood methods to estimate luminosity functions. We present Schechter function fits computed using the STY method (Sandage, Tammann & Yahil 1979) and also non-parametric estimates using the stepwise maximum likelihood method (SWML) of Efstathiou, Ellis & Peterson (1988). Our implementation of each of these methods is briefly described in chapter 3. The advantage of the maximum likelihood methods is that they are not affected by galaxy clustering (provided that the galaxy luminosity function is independent of galaxy density). By contrast the $1/V_{\rm max}$ method, which makes no assumption about the dependence of the luminosity function with density, is subject to biases produced by density fluctuations.

The two maximum likelihood methods determine the shape of the luminosity function, but not its overall normalization. We have chosen to normalize the luminosity functions by matching the galaxy number counts of Jarrett et al. (in preparation). These were obtained from a 184 deg^2 area selected to have low stellar density and in which all the galaxy classifications have all been confirmed by eye. The counts are reproduced in Fig. 4.6. By using the same 7 arcsec aperture magnitudes as Jarrett et al. (in preparation) and scaling the galaxy counts in our redshift survey, we deduce that the effective area of our redshift catalogue is 619 $\pm$ 25 deg^2. Note that normalizing in this way by-passes the problem of whether or not some fraction of the missed 2MASS objects are stars. Fig. 4.6 also shows the Kron and extrapolated magnitude J and $\rm K_S$ counts of the 2MASS-2dFGRS redshift survey. In the lower panel, these counts are seen to be in agreement with the published K-band counts of Gardner et al. (1996) and Glazebrook et al. (1994).

We also checked the normalization using the following independent estimate of the effective solid angle of the redshift survey. For galaxies in the 2dFGRS parent catalogue brighter than $\rm b_J < B_{\rm limit}$, we computed the fraction that have both measured redshifts and match a 2MASS galaxy. For a faint $B_{\rm limit}$ this fraction is small as the 2dFGRS catalogue is much deeper than the 2MASS catalogue, but as $B_{\rm limit}$ is made brighter, the fraction asymptotes to the fraction of the area of the 2dFGRS parent catalogue covered by the joint 2MASS-2dFGRS redshift survey. By this method we estimate that the effective area of our redshift catalogue is 642 $\pm$ 22 deg^2, which is in good agreement with the estimate from the counts of Jarrett et al. .

It should be noted that for neither of these estimates of the effective survey area do the quoted uncertainties take account of variations in the number counts due to large

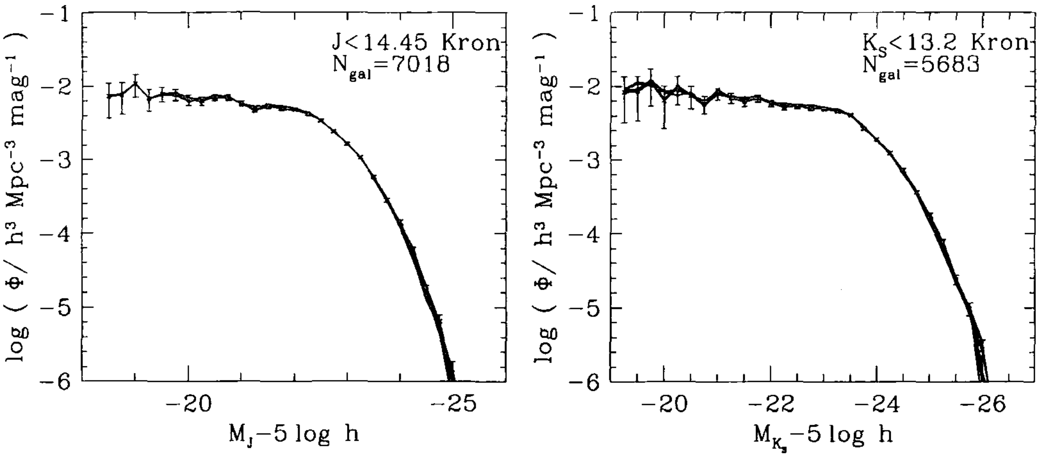


Figure 4.12: SWML estimates of the Kron magnitude J (left) and K_S-band (right) luminosity functions (points with error bars). Our default model of k+e corrections (Kennicutt IMF and standard dust extinction) is adopted. The set of curves on each plot shows the effects of neglecting dust extinction and/or switching to a Salpeter IMF and/or changing the Hubble constant to $H_0 = 50$ km s^{-1}Mpc^{-1} and/or adopting power-law star formation histories and/or making a k-correction but no evolution correction.

scale structure. To estimate the expected variation in the galaxy number counts within the combined 2MASS-2dFGRS survey due to large scale structure we constructed an ensemble of mock catalogues from the ΛCDM Hubble volume simulation of the VIRGO consortium (Evrard 1999). Mock 2dFGRS catalogues constructed from the VIRGO Hubble Volume simulations (Baugh et al. in preparation) can be found at http://star-www.dur.ac.uk/~cole/mocks/hubble.html . We simply took these catalogues and sampled them to the depth of 2MASS over a solid angle of 619 deg^2. To this magnitude limit we found an rms variation in the number of galaxies of 15%. We took this to be a realistic estimate of the uncertainty in the 2MASS number counts and propagated this error through when computing the error on the normalization of the luminosity function.

4.5 Results

4.5.1 Near infra-red galaxy luminosity functions

Fig. 4.12 shows SWML estimates of the Kron J and K_S luminosity functions. The points with errorbars show results for our default choice of k+e corrections, namely those obtained for an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1} cosmology with a Kennicutt

Table 4.1: The dependence of the J and K_S-band Schechter function parameters on cosmological parameters and evolutionary corrections. The parameters refer to STY estimates of the luminosity function for the 2MASS Kron magnitudes and are derived using k or k+e corrections based on model tracks that include dust extinction and assume the Kennicutt IMF. To convert to total magnitudes we estimate that the $M^\star$ values should be brightened by between $\Delta_{\rm Kron} + \Delta_{\rm cal}^{\rm Kron} - \Delta_{\rm conv} = 0.08$ and 0.15 magnitudes. Note that the statistical errors we quote for the $\Phi_\star$ values include the significant contribution that we estimate is induced by large scale structure.

Ω_0	Λ_0	Model Tracks	$M^\star_{\rm J} - 5\log h$	$\alpha_{\rm J}$	$\Phi^\star_{\rm J}/{\rm h}^3{\rm Mpc}^{-3}$	$M^\star_{\rm K_S} - 5\log h$	$\alpha_{\rm K_S}$	$\Phi^\star_{\rm K_S}/{\rm h}^3{\rm Mpc}^{-3}$
0.3	0.7	k+e	-22.36±0.02	-0.93±0.04	$1.04 \pm 0.16 \times 10^{-2}$	-23.44±0.03	-0.96±0.05	$1.08 \pm 0.16 \times 10^{-2}$
0.3	0.7	k only	-22.47±0.02	-0.99±0.04	$0.90 \pm 0.14 \times 10^{-2}$	-23.51±0.03	-1.00±0.04	$0.98 \pm 0.15 \times 10^{-2}$
0.3	0.0	k+e	-22.29±0.03	-0.89±0.04	$1.16 \pm 0.18 \times 10^{-2}$	-23.36±0.03	-0.93±0.05	$1.21 \pm 0.18 \times 10^{-2}$
0.3	0.0	k only	-22.38±0.03	-0.95±0.04	$1.02 \pm 0.15 \times 10^{-2}$	-23.43±0.03	-0.96±0.05	$1.10 \pm 0.16 \times 10^{-2}$
1.0	0.0	k+e	-22.22±0.02	-0.87±0.03	$1.26 \pm 0.19 \times 10^{-2}$	-23.28±0.03	-0.89±0.05	$1.34 \pm 0.20 \times 10^{-2}$
1.0	0.0	k only	-22.34±0.02	-0.93±0.04	$1.08 \pm 0.16 \times 10^{-2}$	-23.38±0.03	-0.93±0.05	$1.18 \pm 0.17 \times 10^{-2}$

Table 4.2: The SWML J and K_S-band luminosity functions for Kron magnitudes as plotted in Fig. 4.13. The units of both ϕ and its uncertainty $\Delta\phi$ are number per h^{-3} Mpc^3 per magnitude.

$M - 5\log h$	$\phi_J \pm \Delta\phi_J$	$\phi_{K_S} \pm \Delta\phi_{K_S}$
-18.75	$(7.94\pm3.59)\times10^{-3}$	$(4.65\pm4.10)\times10^{-3}$
-19.00	$(1.11\pm3.82)\times10^{-2}$	$(5.76\pm4.32)\times10^{-3}$
-19.25	$(6.98\pm2.26)\times10^{-3}$	$(9.16\pm5.67)\times10^{-3}$
-19.50	$(8.14\pm1.80)\times10^{-3}$	$(1.12\pm0.64)\times10^{-2}$
-19.75	$(8.17\pm1.45)\times10^{-3}$	$(1.05\pm0.57)\times10^{-2}$
-20.00	$(7.16\pm1.12)\times10^{-3}$	$(8.58\pm4.63)\times10^{-3}$
-20.25	$(6.62\pm0.88)\times10^{-3}$	$(8.82\pm3.86)\times10^{-3}$
-20.50	$(7.30\pm0.76)\times10^{-3}$	$(6.94\pm2.44)\times10^{-3}$
-20.75	$(7.07\pm0.64)\times10^{-3}$	$(6.09\pm1.63)\times10^{-3}$
-21.00	$(5.84\pm0.48)\times10^{-3}$	$(9.26\pm1.69)\times10^{-3}$
-21.25	$(4.97\pm0.39)\times10^{-3}$	$(6.96\pm1.18)\times10^{-3}$
-21.50	$(5.69\pm0.35)\times10^{-3}$	$(7.29\pm0.98)\times10^{-3}$
-21.75	$(5.15\pm0.28)\times10^{-3}$	$(6.99\pm0.79)\times10^{-3}$
-22.00	$(4.89\pm0.21)\times10^{-3}$	$(5.98\pm0.61)\times10^{-3}$
-22.25	$(4.49\pm0.17)\times10^{-3}$	$(5.93\pm0.52)\times10^{-3}$
-22.50	$(3.41\pm0.12)\times10^{-3}$	$(5.39\pm0.42)\times10^{-3}$
-22.75	$(2.37\pm0.09)\times10^{-3}$	$(5.85\pm0.37)\times10^{-3}$
-23.00	$(1.59\pm0.06)\times10^{-3}$	$(5.24\pm0.28)\times10^{-3}$
-23.25	$(1.06\pm0.04)\times10^{-3}$	$(4.96\pm0.22)\times10^{-3}$
-23.50	$(5.41\pm0.27)\times10^{-4}$	$(4.18\pm0.17)\times10^{-3}$
-23.75	$(2.66\pm0.17)\times10^{-4}$	$(2.72\pm0.11)\times10^{-3}$
-24.00	$(1.19\pm0.10)\times10^{-4}$	$(1.88\pm0.08)\times10^{-3}$
-24.25	$(4.69\pm0.54)\times10^{-5}$	$(1.21\pm0.06)\times10^{-3}$
-24.50	$(1.20\pm0.22)\times10^{-5}$	$(6.54\pm0.37)\times10^{-4}$
-24.75	$(5.40\pm1.34)\times10^{-6}$	$(3.46\pm0.23)\times10^{-4}$
-25.00	$(5.42\pm3.88)\times10^{-7}$	$(1.48\pm0.13)\times10^{-4}$
-25.25		$(5.55\pm0.65)\times10^{-5}$
-25.50		$(2.13\pm0.33)\times10^{-5}$

IMF and including dust extinction. The figure also illustrates that the luminosity functions are very robust to varying this set of assumptions. The various curves in each plot are estimates made neglecting dust extinction and/or switching to a Salpeter IMF and/or changing the Hubble constant to $H_0 = 50$ km s^{-1}Mpc^{-1} and/or adopting power-law star formation histories and/or making a k-correction but no evolution correction. The systematic shifts caused by varying these assumptions are all comparable with or smaller than the statistical errors. The biggest shift results from applying or neglecting the evolutionary correction. In terms of the characteristic luminosity in the STY Schechter function fit, the estimates which include evolutionary corrections are 0.05 to 0.1 magnitudes fainter than those that only include k-corrections (see Table 4.1).

In Fig. 4.13 we compare $1/V_{\rm max}$ and SWML Kron luminosity function estimates (for our default choice of k+e corrections) with STY Schechter function estimates. In general, the luminosity functions are well fit by Schechter functions, but there is marginal evidence for an excess of very luminous galaxies over that expected from the fitted Schechter functions. We tabulate the SWML estimates in Table 4.2. Integrating over the luminosity function gives luminosity densities in the J and K$_{\rm S}$-bands of $\rho_J = (2.75 \pm 0.41) \times 10^8 h{\rm L}_\odot\, h^{-1}\,{\rm Mpc}^{-3}$ and $\rho_{K_S} = (5.74 \pm 0.86) \times 10^8 h{\rm L}_\odot\, h^{-1}\,{\rm Mpc}^{-3}$ respectively, where we have adopted $M_J^\odot = 3.73$ and $M_{K_S}^\odot = 3.39$ (Allen 1973, Johnson 1966). In this analysis, we have not taken account of the systematic and random measurement errors in the galaxy magnitudes. In the case of the STY estimate, the random measurement errors can be accounted for by fitting a Schechter function which has been convolved with the distribution of magnitude errors. However, for the Kron magnitudes, the rms measurement error is only 0.1 magnitudes, as indicated by the comparison in the top right hand panel of Fig 4.5, and such a convolution has only a small effect on the resulting Schechter function parameters. We find that the only parameter that is affected is $M^\star$ which becomes fainter by just $\Delta_{\rm conv} = 0.02$ magnitudes. The comparison to the Loveday (2000) data also indicates a systematic error in the 2MASS Kron magnitudes of $\Delta_{\rm cal}^{\rm Kron} = 0.061 \pm 0.031$. Combining these two systematic errors results in a net brightening of $M^\star$ by $\Delta_{\rm cal}^{\rm Kron} - \Delta_{\rm conv} = 0.041 \pm 0.031$ magnitudes. As this net systematic error is both small and uncertain we have chosen not to apply a correction to our quoted Kron magnitude luminosity function parameters. We recall also that to convert from Kron to total magnitudes requires brightening $M^\star$ by between $\Delta_{\rm Kron} = 0.044$ and 0.11 depending on whether the luminosity profile of a typical galaxy is fit well by an exponential or $r^{1/4}$-law.

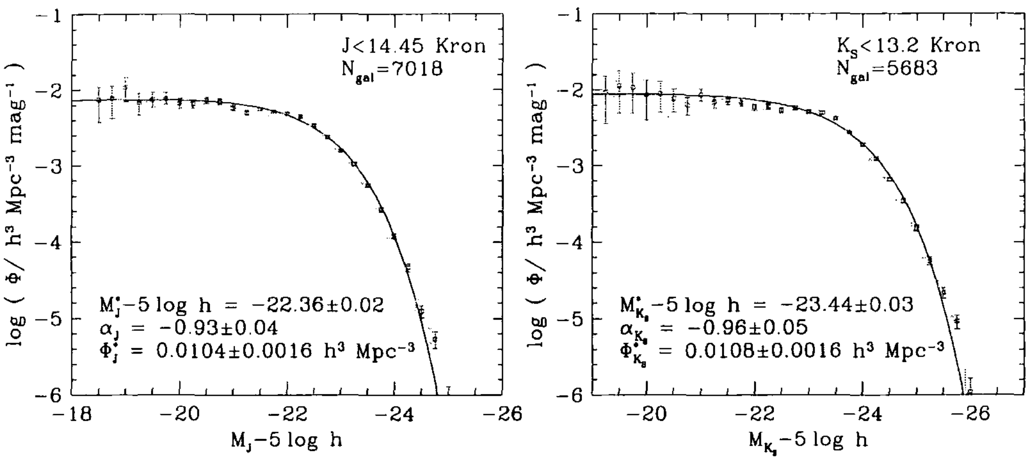


Figure 4.13: SWML estimates of the Kron magnitude J (left) and K_S-band (right) luminosity functions (data points with errorbars) and STY Schechter function estimates (lines). The parameter values and error estimates of the Schechter functions are given in the legends. The errorbars without data points show $1/V_{\rm max}$ estimates of the luminosity functions. For clarity these have been displaced to the left by 0.1 magnitudes.

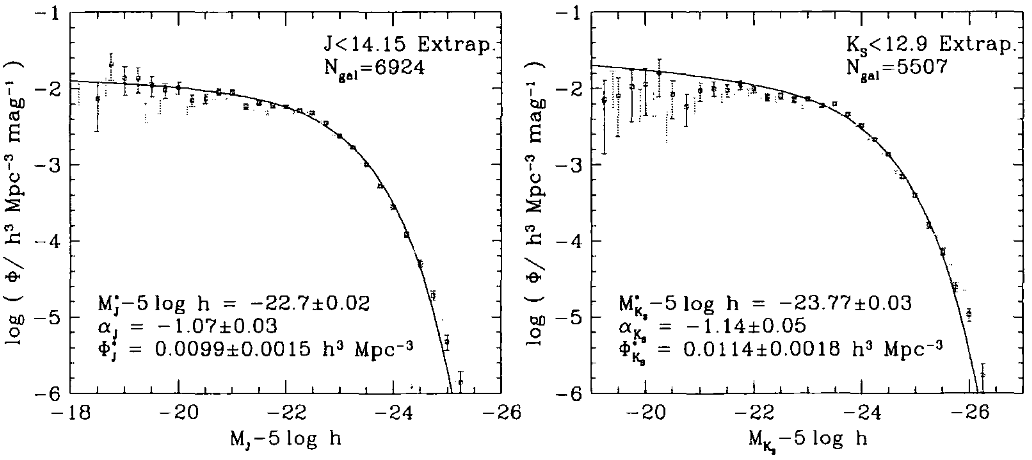


Figure 4.14: SWML estimates of the extrapolated magnitude J (left) and K_S-band (right) luminosity functions (data points with errorbars) and STY Schechter function estimates (lines). The parameter values and error estimates of the Schechter functions are given in the legends. The errorbars without data points show $1/V_{\rm max}$ estimates of the luminosity functions. For clarity these have been displaced to the left by 0.1 magnitudes.

Fig. 4.14 shows the SWML and STY luminosity function estimates for samples defined by the 2MASS extrapolated, rather than Kron, magnitudes. With this definition of magnitude, the luminosity functions differ significantly from those estimated using Kron magnitudes. In particular, the characteristic luminosities are 0.34 and 0.28 magnitudes brighter in J and K_S respectively. Most of this difference is directly related to the systematic offset in the J-band Kron and extrapolated magnitudes, which can be seen in either the middle panel of Fig. 4.4 or the right hand panels of Fig. 4.5 to be approximately 0.23 magnitudes. Note that even in the K_S-band, it is this J-band offset that is relevant as the K_S-band magnitudes we use are derived from the J-band values using the measured aperture colours. We have argued in section 4.2.3 that this offset is caused by the J-band Kron 2MASS magnitudes being fainter than true total magnitudes by between $\Delta_{\rm cal}^{\rm Kron} + \Delta_{\rm Kron} = 0.1$ and 0.17 and the extrapolated magnitudes being systematically too bright by $-\Delta_{\rm cal}^{\rm extrap.} - \Delta_{\rm Kron} = 0.05$ to 0.11 magnitudes. Subtracting this 0.23 magnitude offset results in Kron and extrapolated luminosity functions that differ in $M^{\star}$ by only 0.11 magnitudes. In the K_S-band, the faint end slope of the best-fit Schechter function is significantly steeper in the extrapolated magnitude case, but note that this function is not a good description of the faint end of the luminosity function since the SWML and $1/V_{\rm max}$ estimates lie systematically below it. The Schechter function fit is constrained mainly around $M_{\star}$ and in this case the χ^2 value indicates it is not a good fit overall.

The residual differences between the Kron and extrapolated magnitude luminosity functions arise from the scatter in the relation between extrapolated and Kron magnitudes. If this scatter is dominated by measurement error, then these differences represent small biases, which are largest for the less robust, extrapolated magnitudes. However, it is possible that the scatter is due to genuine variations in galaxy morphology and light profiles. To assess which of these alternatives is correct requires independent deep photometry of a sample of 2MASS galaxies to quantify the accuracy of the extrapolated magnitudes. However, we note that the $V/V_{\rm max}$ distributions for the extrapolated magnitudes shown in Fig. 4.7 have mean $\langle V/V_{\rm max}\rangle$ values significantly greater than 0.5, which is probably an indication that the extrapolated magnitudes are not robust. Thus, overall we favour adopting Kron magnitudes, noting the small offset of $\Delta_{\rm cal}^{\rm Kron} + \Delta_{\rm Kron} - \Delta_{\rm conv} = 0.08$ to 0.15 required to convert to total magnitudes and correct for the convolving effect of measurement errors.

The parameters of the STY Schechter function fits shown in Fig. 4.13 are listed in the first row of Table 4.1. The subsequent rows illustrate how the best-fit parameters

Table 4.3: Schechter function fits to K-band luminosity functions. Where necessary, the values quoted have been converted from the cosmological model assumed in the original work to the $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ assumed here. In addition, we have shifted the $M^\star_{\rm K}$ of Kochanek et al. (2000) brightward by 0.05 magnitudes corresponding to the mean difference between 2MASS Kron and isophotal magnitudes in the Kochanek et al. sample. We have also shifted $M^\star_{\rm K}$ of Glazebrook et al. (1995) brightward by 0.3 magnitudes and that of Mobasher et al. (1993) faintward by 0.22 magnitudes as advocated by Glazebrook et al. (1995) to make aperture corrections and consistent k-corrections respectively.

Sample	$M^\star_{\rm K}$	$\alpha_{\rm K}$	$\Phi_{\rm K}/{\rm h}^3{\rm Mpc}^{-3}$
Mobasher et al. 1993	-23.37 ± 0.30	1.0 ± 0.3	$1.12 \pm 0.16 \times 10^{-2}$
Glazebrook et al. 1995	-23.14 ± 0.23	1.04 ± 0.3	$2.22 \pm 0.53 \times 10^{-2}$
Gardner et al. 1997	-23.30 ± 0.17	1.0 ± 0.24	$1.44 \pm 0.20 \times 10^{-2}$
Szokoly et al. 1998	-23.80 ± 0.30	1.3 ± 0.2	$0.86 \pm 0.29 \times 10^{-2}$
Loveday et al. 2000	-23.58 ± 0.42	1.16 ± 0.19	$1.20 \pm 0.08 \times 10^{-2}$
Kochanek et al. 2000	-23.43 ± 0.05	1.09 ± 0.06	$1.16 \pm 0.1 \times 10^{-2}$
This work	-23.36 ± 0.02	0.93 ± 0.04	$1.16 \pm 0.17 \times 10^{-2}$

change when the cosmological model is varied and the evolutionary correction is included or excluded. The $M^\star$ values are approximately 0.14 magnitudes fainter for the $\Omega_0 = 1$ case than for our standard $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology. This shift is largely due to the difference in distance moduli between the two cosmologies at the median redshift of the survey. This, and the difference in the volume-redshift relation, cause $\phi_\star$ to change in order to preserve the same galaxy number counts.

Fig. 4.15 compares our estimates of the K_S-band luminosity function for our standard $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology, with the estimates of Mobasher et al. (1993), Glazebrook et al. (1995), Gardner et al. (1997), Mobasher et al. (1993), Szokoly et al. (1998), Loveday (2000) and Kochanek et al. (2001). In general, these authors assumed different cosmological models when analysing their data. We have therefore modified the estimates from each survey. First, we apply a shift in magnitude reflecting the difference in distance moduli, at the median redshift, between the assumed cosmological model and our standard $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ model. We then apply a shift in number density so as to keep fixed the surface density of galaxies per square degree at the survey magnitude limit. In

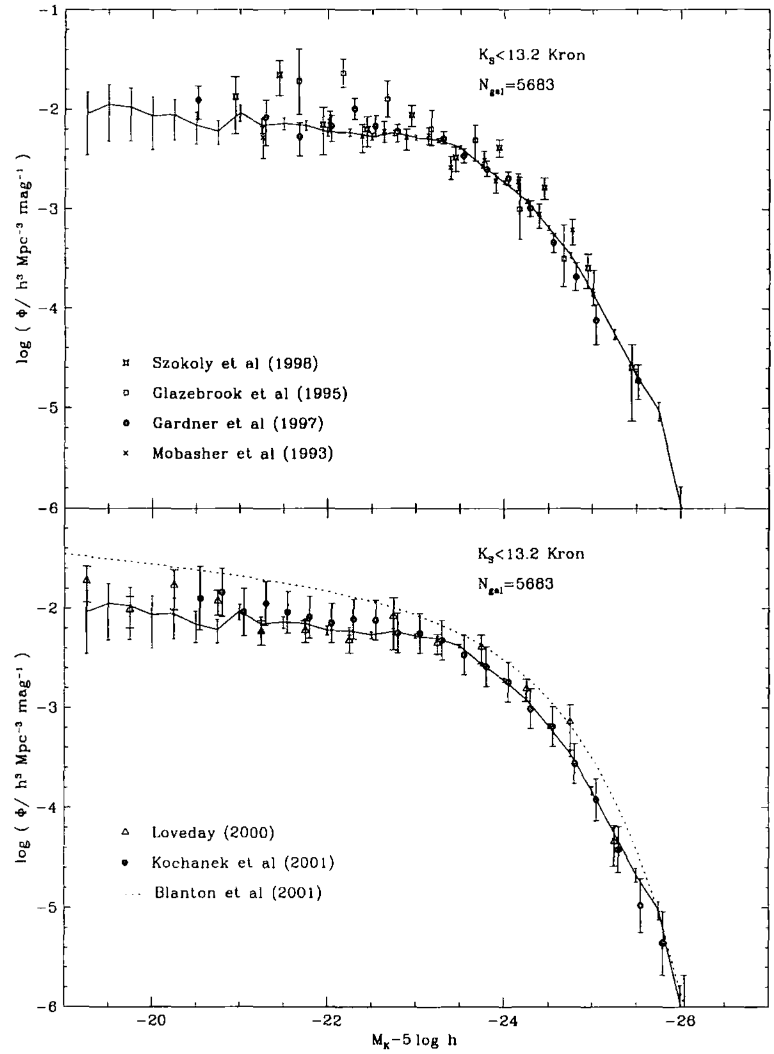


Figure 4.15: Comparison of various estimates of the K-band luminosity function. In both panels the solid line shows our SWML estimate of the K_S-band luminosity function for Kron magnitudes. The symbols and errorbars in the top panel show the estimates of Mobasher et al. (1993), Glazebrook et al. (1995), Gardner et al. (1997) and Szokoly et al. (1998) as indicated in the legend. We have shifted the data of Glazebrook et al. (1995) brightward by 0.3 magnitudes and the data of Mobasher et al. (1993) faintward by 0.22 magnitudes as advocated by Glazebrook et al. (1995) to make aperture corrections and consistent k-corrections respectively. In the lower panel the symbols and errorbars show the recent estimates of Loveday (2000) and Kochanek et al. (2000). The estimate of Kochanek et al. has been shifted brightward by 0.05 magnitudes to account for the difference between isophotal and Kron magnitudes. The dotted line shows a Schechter function estimate of the K-band luminosity function inferred from SDSS z*-band luminosity function of Blanton et al. (2000) (see text for details).

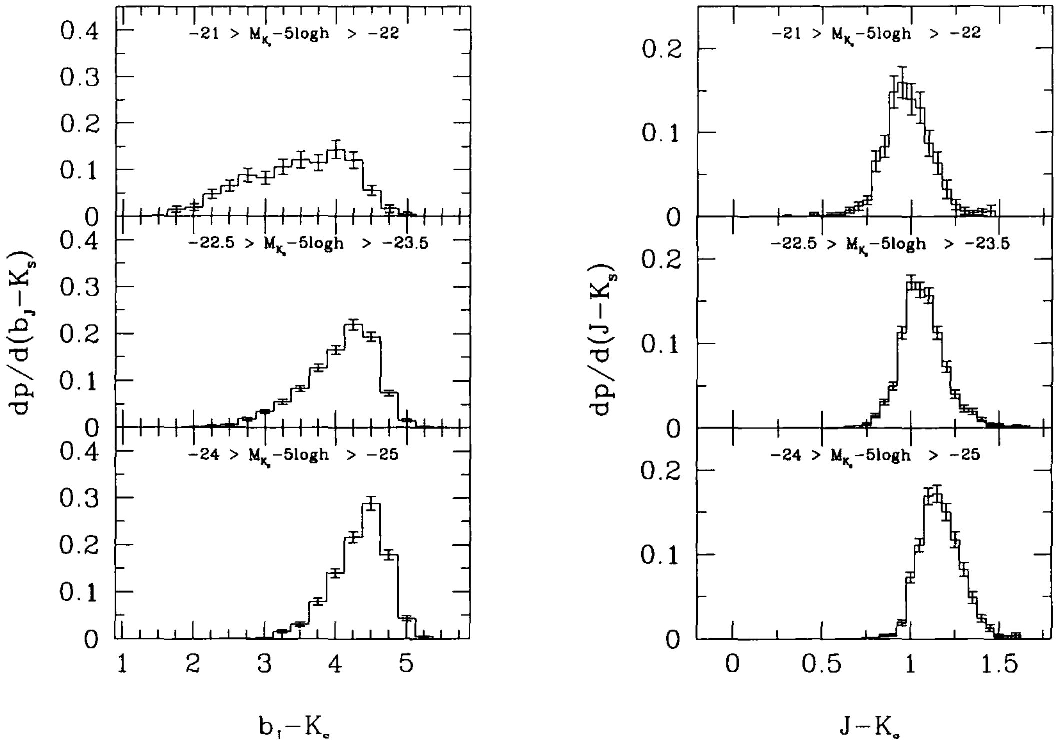


Figure 4.16: The distribution of rest frame $b_J - K_S$ (left) and $J - K_S$ (right) colours in three bins of K_S absolute magnitude, computed using our default set of k+e corrections.

the case of the Kochanek et al. (2001) luminosity function we have shifted the data points brightwards by 0.05 magnitudes to account for the mean difference between the 2MASS isophotal magnitudes used by Kochanek et al. (2001) and the Kron magnitudes we have adopted. Schechter function parameters scaled and adjusted in this manner are given for each survey in Table 4.3. Note that due to the correlations between the Schechter function parameters it is better to judge the agreement between the different estimates by reference to Fig. 4.15 rather than by the parameter values in Table 4.3. Our new estimate of the K_S-band luminosity function is in excellent agreement with the independent estimates and has the smallest statistical errors at all magnitudes brighter than $M_{K_S} - 5\log h = -22$. For very faint magnitudes, from -20 to -16, the sparsely sampled survey of Loveday (2000) has smaller statistical errors. Note that many previous analyses of the K-band luminosity function ignored the contribution of large scale structure to the error in Φ_*, and so the errors in Table 4.3 are likely to be underestimated.

Also shown on the lower panel of Fig. 4.15 is an estimate of the K-band luminosity function inferred from the Sloan Digital Sky Survey (SDSS) near infrared, z*-band luminosity function of Blanton et al. (2001). To convert from z* (AB system) to standard K we have simply subtracted 2.12 magnitudes from the SDSS z* magnitudes. This offset consists of a contribution of 0.51 magnitudes to convert from AB magnitudes to the standard Vega system, and a mean z^*-K_S colour of 1.61, which we find is typical of the model spectra discussed in section 4.3 that match our observed b_J-K_S colours. As has been noted by Wright (2001) the luminosity function inferred from the SDSS data is offset compared to our estimate. One suggestion put forward by Wright (2001) is that the 2MASS magnitudes could be systematically too faint. The systematic error would have to amount to 0.5 magnitudes to reconcile the luminosity density inferred from the SDSS data with that which we infer from the 2MASS-2dFGRS catalogue. Such an error is comprehensively excluded by the very small offset that was found in section 4.2.3 between the 2MASS Kron magnitudes and the data of Loveday (2000). Also, a direct galaxy-by-galaxy comparison of the z^*-J and z^*-K_S colours computed using the SDSS Petrosian and 2MASS Kron magnitudes produced galaxy colours in good accord with expectations based on model spectra (Ivezic, Blanton and Loveday private communication). Finally, we note that a good match to our estimate of the K_S-band luminosity function cannot be achieved by simply moving the SDDS curve in Fig. 4.15 horizontally. If slid by 0.5 magnitudes to match the luminosity density then it falls well below our estimate at bright magnitudes. However if the SDSS curve is moved vertically, by a factor of 1.6, then the two estimates come into reasonable agreement at all magnitudes. Thus, the most likely explanation of the difference between the SDSS and 2MASS-2dFGRS luminosity functions is the uncertainty in the overall normalization induced by large-scale density fluctuations. It is to be hoped that as the sky coverage of the SDSS and 2MASS-2dFGRS surveys increases this discrepancy will be reduced.

This last issue should be contrasted with the similar problem we encountered in chapter 3, when comparing our estimate of the b_J-band galaxy luminosity function with the one measured by Blanton et al. (2001). The largest discrepancy between our two estimates was the overall normalisation, which had been obtained from two different methods. As in chapter 3, we choose to normalize to the number counts, whereas Blanton et al. (2001) used the method of Davis & Huchra (1982b), which weights galaxies as a function of redshift in order to get a minimum variance estimate of the galaxy density. This method gives more weight to galaxies at high redshift than normalizing to the counts. The choice of the method by which one normalises the luminosity function can probably explain the difference seen in various luminosity function estimates, and particularly if one compares two surveys of different depths.

4.5.2 Colour distributions

Since our combined 2MASS–2dFGRS catalogue includes b_J-band and infrared magnitudes, it is also possible to estimate the b_J-band optical luminosity function and the optical/infrared bivariate luminosity function. We do not present the b_J-band optical luminosity function here as estimates from the 2dFGRS are discussed in detail in chapter 3 and decomposed into luminosity functions of different spectral types in Folkes et al. (1999) and Madgwick et al. (2001). Instead, we present the bivariate b_J/K_S and J/K_S luminosity functions in Fig. 4.16, in the form of the rest-frame b_J-K_S and $J-K_S$ colour distributions, split by K_S-band absolute magnitude. Results are shown for just our default set of k+e corrections, but the colour distributions are extremely insensitive to this choice and to whether evolutionary corrections are ignored or included. The shape of b_J-K_S colour distribution varies systematically with K_S-band luminosity. At fainter magnitudes there is an increasingly large population of bluer, star-forming galaxies. The star formation rate has less effect on the infrared $J-K_S$ colours. Here, the shape of the $J-K_S$ colour distribution varies little with luminosity, but the position of the peak moves gradually redder with increasing luminosity. Colour distributions such as these are sensitive to both the distribution of stellar age and the metallicity, and therefore provide important constraints on models of galaxy formation (Cole et al. (2000) and references therein).

4.5.3 Spectral type distribution

Another interesting issue that we can address with our data is the distribution of spectral types in the 2MASS catalogue. For this, we make use of the spectral information in the 2dF galaxies extracted by a principal component analysis (Folkes et al. 1999). Specifically, we use the new continuous variable introduced in Madgwick, Lahav & Taylor (2000) which is defined by a linear combination of the first 2 principal component projections, $\eta \equiv 0.44pc_1 - pc_2$. This variable was chosen to be robust to instrumental uncertainties whilst, at the same time, preserving physical information about the galaxy. The dominant influence on the η parameter is the relative strength of absorption and emission lines ($\eta<0$ implies less than average emission-line strength while $\eta>0$ implies stronger than average emission-line strength). A more detailed description is presented in Madgwick et al. (2001).

We can now gain insight into the population mix of our 2MASS sample by simply

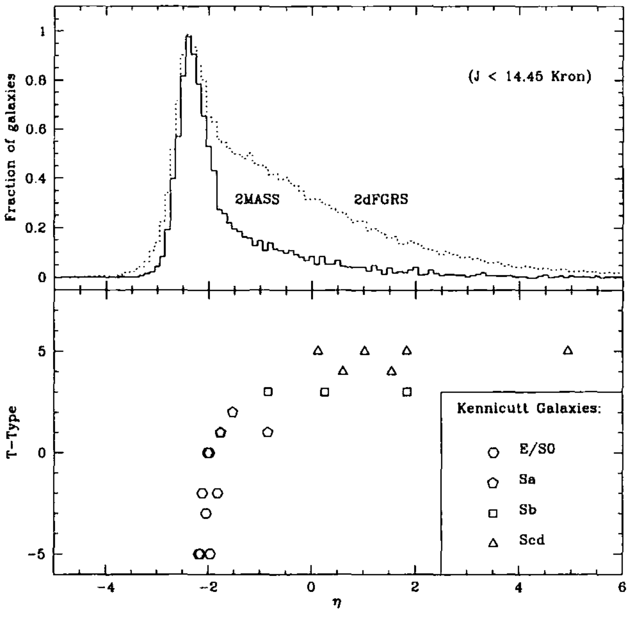


Figure 4.17: The distribution of the spectral type parameter, η, in the full 2dFGRS and our matched 2MASS catalogue (upper panel). The lower panel uses a sample of galaxies from Kennicutt (1992) to show how η is correlated with morphological type.

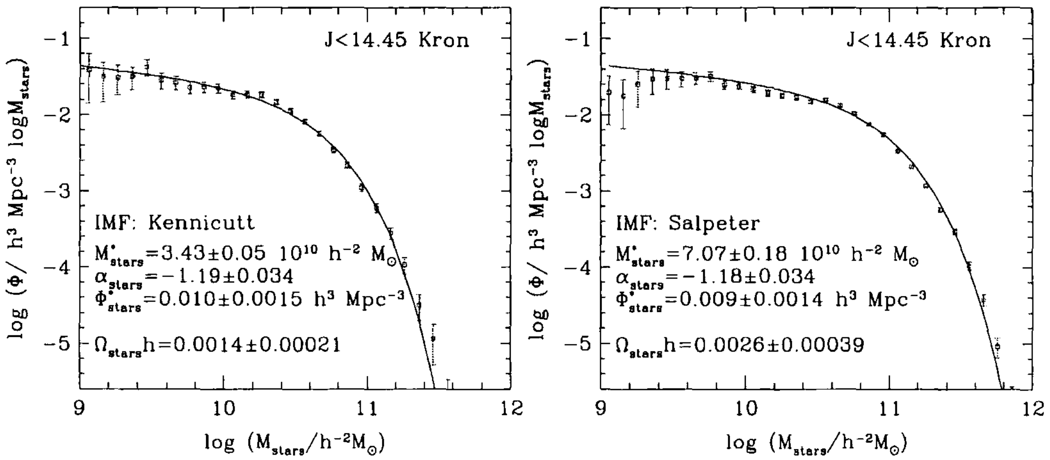


Figure 4.18: SWML estimates of the stellar mass function (open symbols with error bars) and STY Schechter function estimates (lines). The parameter and error estimates of the Schechter function fits are given in the legends. The left-hand panel is for a Kennicutt IMF with recycled fraction $R = 0.42$ and the right-hand panel for a Salpeter IMF with $R = 0.28$.

creating a histogram of the η values for the corresponding 2MASS–2dFGRS matched galaxies with J<14.45 (Kron). We plot this in Fig. 4.17 where we also show data for the entire 2dFGRS sample as comparison. Also shown in Fig. 4.17 (bottom panel) is the morphology-η relation derived from a sample of galaxy spectra from the Kennicutt Atlas (Kennicutt 1992b).

It can be clearly seen from Fig. 4.17 that the predominant population in the 2MASS sample is has $\eta<-2$. By contrasting this with values of η obtained from the spectra of galaxies with known morphological type (Kennicutt 1992), we can see that this corresponds to galaxies of E/S0 morphologies. More precisely, the fraction of galaxies in our matched sample with spectral types corresponding to E/S0 morphologies is 62% (compared with $\sim$ 35% in the full 2dFGRS). Sa-Sb galaxies make up a further 22% and the remaining 16% are galaxies of later morphological types.

4.5.4 Galaxy stellar mass function

In contrast to optical light, near-infrared luminosities are relatively insensitive to the presence of young stars and can be more accurately related to the underlying stellar mass. Thus, with relatively few model assumptions, we can derive the distribution of galaxy stellar masses. The integral of this distribution is the total mass density in stars, which can be expressed in units of the critical density as $\Omega_{\rm stars}$. Attempts to estimate this quantity date back many decades, but even recent estimates such as those by Persic & Salucci (1992), Gnedin & Ostriker (1992), Fukugita, Hogan & Peebles (1998) and Salucci & Persic (1999) have very large uncertainties because they are based on B-band light and require uncertain B-band mass-to-light ratios. The much more accurate estimate that we provide here should prove very useful for a variety of purposes.

To estimate the galaxy stellar mass function, we use the modelling of the stellar populations described in section 4.3 to obtain estimates of the present luminosity and stellar mass-to-light ratio for each galaxy in the survey. This is done on a galaxy-by-galaxy basis as described in section 4.3. The sample we analyze is defined by the 11<J<14.45 (Kron) apparent magnitude limits. The stellar mass that we estimate for each galaxy is the mass locked up in stars and stellar remnants. This differs from the time integral of the star formation rate because some of the material that goes into forming massive stars is returned to the interstellar medium via winds and supernovae. For a given IMF, this recycled fraction, R, can be estimated reasonably accurately from stellar evolution theory. Here, we adopt the values $R = 0.42$ and 0.28 for the Kennicutt (1983) and Salpeter (1955)

IMFs respectively, as described in section 5.2 of Cole et al. (2000) who made use of the models of Renzini & Voli (1981) and Woosley & Weaver (1995). Hence, the stellar masses we choose to estimate are $(1-R)$ times the time integral of the star formation rate to the present day. Note that the IMFs we consider assume that only stars with mass greater than $0.1\mathrm{M}_\odot$ ever form and so we are not accounting for any mass that may be locked up in the form of brown dwarfs.

Our results are presented in Fig. 4.18 which shows both SWML and Schechter function estimates of the present-day galaxy stellar mass function for two choices of IMF. The SWML estimates are tabulated in Table 4.4. Just as for the luminosity functions, the stellar mass function is quite well described by the Schechter functional form. Integrating over these Schechter functions to determine the total stellar mass gives $\Omega_{\mathrm{stars}}h = (1.4 \pm 0.21) \times 10^{-3}$ for the Kennicutt IMF and $\Omega_{\mathrm{stars}}h = (2.6 \pm 0.39) \times 10^{-3}$ for the Salpeter IMF. Note that the integral converges rapidly at both limits and, in particular, the contribution to Ω_{stars} from objects with $M_{\mathrm{stars}} < 10^9\ \mathrm{h}^{-2}\ \mathrm{M}_\odot$ is negligible. We find that these values vary by less than the quoted errors when we alter the assumed (k+e)-corrections by either ignoring evolution, ignoring dust or changing Ω_0. Taken together with our estimates of the $\mathrm{K_S}$-band luminosity density these, estimates imply mean stellar mass-to-light ratios of $0.73\,\mathrm{M}_\odot/\mathrm{L}_\odot$ in the case of the Kennicutt IMF and $1.32\,\mathrm{M}_\odot/\mathrm{L}_\odot$ for the Salpeter IMF. If we apply the correction we estimated in section 4.2.3 to transform 2MASS Kron into total magnitudes, then these estimates and their uncertainties increase to $\Omega_{\mathrm{stars}}h = (1.6 \pm 0.24) \times 10^{-3}$ for the Kennicutt IMF and $\Omega_{\mathrm{stars}}h = (2.9 \pm 0.43) \times 10^{-3}$ for the Salpeter IMF. Both of these estimates are consistent with the value, $\Omega_{\mathrm{stars}} = (3.0 \pm 1.0) \times 10^{-3}$, derived by Salucci & Persic (1999) but have fractional statistical errors which are several times smaller. With our method, the uncertainty in Ω_{stars} is clearly dominated by the uncertainty in the IMF. For some purposes, it is not possible to improve upon this without a more precise knowledge of the true IMF – assuming there is a universal IMF. However, for other applications, such as modelling the star formation history of the universe, it is necessary to assume a specific IMF to convert the observational tracers of star formation to star formation rates. Hence, in this case, it is the much smaller statistical errors that are relevant.

It is interesting to compare our values with what is inferred by integrating the observational estimates of the mean star formation history of the universe. Fig. 4.19 shows observational estimates for one particular choice of cosmology and IMF and illustrates how the rates are sensitive to the assumed dust extinction. By fitting a smooth curve

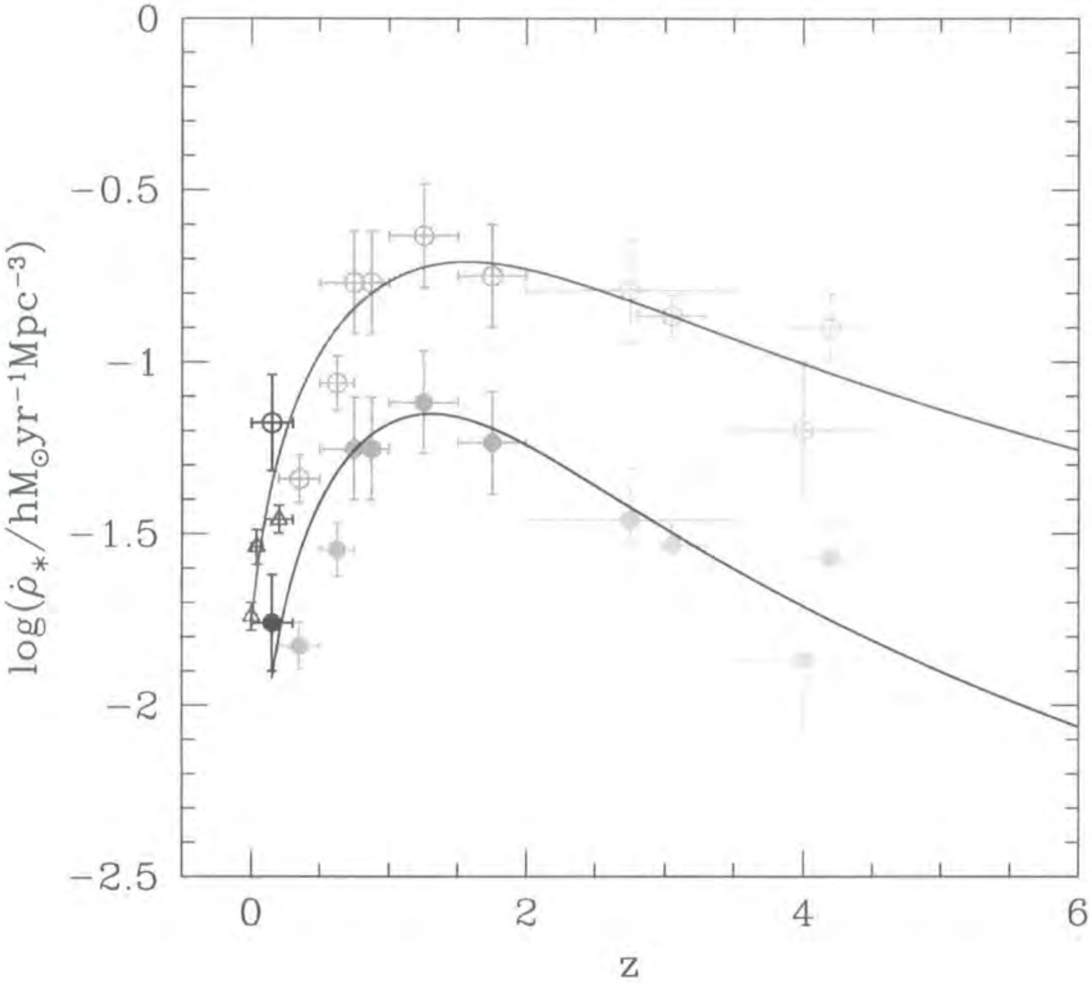


Figure 4.19: Observational estimates of the star formation history of the universe. The points with error bars show estimates of the mean star formation rate per unit volume at various redshifts (see Steidel et al. 1999 and references therein). The solid symbols are the star formation rates implied if there is no absorption by dust. The open symbols show estimates corrected for dust absorption using a Calzetti (1999) extinction law with a mean E(B − V) = 0.15 (Steidel et al. 1999). In both cases an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology has been used to calculate the volume as a function of redshift and a Salpeter IMF to convert luminosity to star formation rate. The smooth curves are the fits we use when integrating over time to estimate the total mass density of stars formed by the present.

Table 4.4: The SWML stellar mass functions as plotted in Fig. 4.18. The units of both ϕ and its uncertainty $\Delta\phi$ are number per h^{-3} Mpc^3 per decade of mass.

	Kennicutt	Salpeter
$\log_{10} M$	$\phi \pm \Delta\phi$	$\phi \pm \Delta\phi$
9.06	$(4.24\pm2.62)\times10^{-2}$	$(1.37\pm1.05)\times10^{-2}$
9.16	$(3.42\pm1.80)\times10^{-2}$	$(2.41\pm1.35)\times10^{-2}$
9.26	$(3.01\pm1.31)\times10^{-2}$	$(2.06\pm1.13)\times10^{-2}$
9.36	$(3.33\pm1.11)\times10^{-2}$	$(3.01\pm1.13)\times10^{-2}$
9.46	$(4.21\pm1.04)\times10^{-2}$	$(3.25\pm0.92)\times10^{-2}$
9.56	$(2.75\pm0.67)\times10^{-2}$	$(2.87\pm0.67)\times10^{-2}$
9.66	$(2.70\pm0.55)\times10^{-2}$	$(3.10\pm0.56)\times10^{-2}$
9.76	$(2.31\pm0.42)\times10^{-2}$	$(3.30\pm0.47)\times10^{-2}$
9.86	$(2.20\pm0.35)\times10^{-2}$	$(2.67\pm0.34)\times10^{-2}$
9.96	$(2.21\pm0.31)\times10^{-2}$	$(2.51\pm0.27)\times10^{-2}$
10.06	$(1.77\pm0.23)\times10^{-2}$	$(2.03\pm0.20)\times10^{-2}$
10.16	$(1.91\pm0.20)\times10^{-2}$	$(1.93\pm0.17)\times10^{-2}$
10.26	$(1.77\pm0.16)\times10^{-2}$	$(1.86\pm0.15)\times10^{-2}$
10.36	$(1.46\pm0.12)\times10^{-2}$	$(1.62\pm0.11)\times10^{-2}$
10.46	$(1.11\pm0.08)\times10^{-2}$	$(1.49\pm0.09)\times10^{-2}$
10.56	$(8.15\pm0.61)\times10^{-3}$	$(1.61\pm0.08)\times10^{-2}$
10.66	$(5.62\pm0.43)\times10^{-3}$	$(1.30\pm0.06)\times10^{-2}$
10.76	$(3.39\pm0.29)\times10^{-3}$	$(1.06\pm0.04)\times10^{-2}$
10.86	$(2.08\pm0.20)\times10^{-3}$	$(7.40\pm0.30)\times10^{-3}$
10.96	$(1.07\pm0.12)\times10^{-3}$	$(5.50\pm0.22)\times10^{-3}$
11.06	$(5.95\pm0.82)\times10^{-4}$	$(3.29\pm0.15)\times10^{-3}$
11.16	$(2.75\pm0.49)\times10^{-4}$	$(2.02\pm0.10)\times10^{-3}$
11.26	$(1.05\pm0.26)\times10^{-4}$	$(1.13\pm0.07)\times10^{-3}$
11.36	$(2.77\pm1.11)\times10^{-5}$	$(5.56\pm0.40)\times10^{-4}$
11.46	$(9.51\pm5.65)\times10^{-6}$	$(2.90\pm0.26)\times10^{-4}$
11.56	$(2.05\pm2.38)\times10^{-6}$	$(9.87\pm1.26)\times10^{-5}$
11.66	$(6.87\pm13.6)\times10^{-7}$	$(3.73\pm0.66)\times10^{-5}$
		$(8.46\pm2.58)\times10^{-6}$
		$(2.22\pm1.20)\times10^{-6}$

Table 4.5: Estimates of the present-day mass in stars and stellar remnants obtained by integrating over observational estimates of the star formation history of the universe. We express this stellar mass density in terms of the critical density and give values of $\Omega_{\rm stars}h^2$ estimated for different assumed IMF's and dust corrections. All values are for an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology and assume stellar populations of solar metallicity.

Dust Extinction	Kennicutt IMF	Salpeter IMF
E(B-V)=0.05	0.80×10^{-3}	1.30×10^{-3}
E(B-V)=0.10	1.17×10^{-3}	1.86×10^{-3}
E(B-V)=0.15	1.63×10^{-3}	2.66×10^{-3}

through these estimates, we can calculate the mass of stars formed by the present day and how this depends on the IMF and assumed dust extinction. The upper smooth curve shown in Fig. 4.19 is of the form $\dot{\rho}_\star = (a+bz)/(1+(z/c)^d)hM_\odot{\rm yr}^{-1}{\rm Mpc}^{-3}$, where $(a, b, c, d) = (0.0166, 0.1848, 1.9474, 2.6316)$. The data points uncorrected for dust extinction are fit with $(a, b, c, d) = (0.0, 0.0798, 1.658, 3.105)$. As for our estimates above, we assume that no mass goes into forming brown dwarfs and multiply the star formation rate by $1-R$, where R is the recycled fraction, so as to form an estimate of the mass locked up in stars. Values of $\Omega_{\rm stars}h^2$ estimated in this way are listed in Table 4.5. The values in this table are for an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology, but they are insensitive to this choice. They depend slightly on the assumed metallicity of the stellar population and would be 10% lower if half solar, rather than solar metallicity were assumed. Note that the $\Omega_{\rm stars}$ values inferred from the star formation history of the universe scale differently with the assumed Hubble constant than those inferred above from the IR luminosity functions. For $h = 0.7$ our estimates from 2MASS become $\Omega_{\rm stars}h^2 = (1.12 \pm 0.16) \times 10^{-3}$ for the Kennicutt IMF and $\Omega_{\rm stars}h^2 = (2.03 \pm 0.30) \times 10^{-3}$ for the Salpeter IMF. Comparison with Table 4.5 shows that these values are consistent with those inferred from the cosmic star formation history only if the dust correction assumed in the latter is modest, E(B-V)$\approx$ 0.1. This value is 50% smaller than the value preferred by Steidel et al. (1999).

4.6 Conclusions

The new generation of very large surveys currently underway make it possible to characterize the galaxy population with unprecedented accuracy. In this chapter, we have combined two such large surveys, the infrared imaging 2MASS and the 2dF Galaxy Redshift Survey to obtain a complete dataset which is more than an order of magnitude larger than previous datasets used for statistical studies of the near-infrared properties of the local galaxy population. We have used this combined catalogue to derive the most precise estimates to date of the galaxy J and K_S-band luminosity functions and of the galaxy stellar mass function.

Characterizing the near-infrared properties of galaxies offers several advantages. Firstly, the near-infrared light is dominated by established, old stellar populations rather than by the recent star formation activity that dominates the blue light. Thus, the J and K-band luminosity functions reflect the integrated star formation history of a galaxy and, as a result, provide particularly important diagnostics of the processes of galaxy formation. For the same reason, the distribution of stellar mass in galaxies -the galaxy stellar mass function- can be derived from the near-infrared luminosities in a relatively straightforward way, with only a weak model dependence. Finally, corrections for dust extinction as well as k-corrections are much smaller in the near-infrared than in the optical.

Due to the size of our sample, our determination of the J- and K_S-band galaxy luminosity functions have, for the most part, smaller statistical errors than previous estimates. Furthermore, since our sample is infrared-selected, our estimates are free from any potential biases that might affect infrared luminosity functions derived from optically-selected samples. We find that the J- and K_S-band galaxy luminosity functions are fairly well described by Schechter functions, although there is some evidence for an excess of bright galaxies relative to the best-fit Schechter functional form. In general, the SWML estimates are a truer representation of the luminosity functions. Our K-band estimates are in overall agreement with most previous determinations, but have smaller statistical errors.

The exception is the K-band luminosity function inferred from the near infrared SDDS photometry (Blanton et al. 2001). The difference between the K-band luminosity function we infer from their data and our own estimate is too large to be explained by photometric differences. The difference between the two estimates is better described by a difference in overall number density of a factor of 1.6. A similar discrepancy is seen in the b_J-band between the SDSS and 2dFGRS luminosity function estimates (see chapter 3 for further details). The suspicion is that the uncertainty in the overall normalization of the luminosity functions induced by large-scale structure within the large, but finite, survey volumes could be to blame. However, the errors that we quote for the 2dFGRS-2MASS luminosity functions already include an estimate of this sampling uncertainty as derived from realistic mock catalogues. A similar exercise for a catalogue with the same area and depth as that of Blanton et al. (2001) indicates that the required overdensity of a factor of 1.6 is unlikely. So probably there is more than one contributory factor at work and the hope is that these will be identified as the surveys progress.

Chapter 5

Quantifying Redshift Space Distortions

The large-scale structure in the distribution of galaxies is thought to arise from the gravitational instability of small fluctuations in the initial density field of the universe. A key test of this hypothesis is that superclusters of galaxies in the process of formation should generate systematic infall of other galaxies. This would be evident in the pattern of recessional velocities, causing an anisotropy in the inferred 3-dimensional clustering of galaxies. Here we report a precise measurement of this clustering, using the redshifts of more than 141,000 galaxies* from the 2dFGRS. We determine the parameter $\beta \equiv \Omega^{0.6}/b = 0.43 \pm 0.07$, where Ω is the total mass-density parameter and b is a measure of the 'bias' of the luminous galaxies in the survey. Combined with the anisotropy of the cosmic microwave background, our results favour a low-density universe with $\Omega \simeq 0.3$.

5.1 Introduction

Hubble (1934) showed that the pattern of galaxies on the sky is non-random, and successive years have seen ever more ambitious attempts to map the distribution of visible matter on cosmological scales. In order to obtain a three-dimensional picture, redshift surveys use Hubble's law, $v = H_0 r$, to infer approximate radial distances to a set of galaxies. The first major surveys of this sort took place in the early 1980s (Kirshner et al. 1981; Davis & Peebles 1983; Bean et al. 1983; de Lapparent, Geller & Huchra 1986), and were limited to a few thousand redshifts, owing to the limited speed of single-object spectroscopy. In the 1990s, redshift surveys were extended to much larger volumes by a 'sparse sampling' strategy (Kaiser 1986). These studies (Saunders et al. 1991; Shectman et al. 1996) established that the universe was close to uniform on scales above about $100\,h^{-1}$ Mpc ($h \equiv H_0/100\,\mathrm{km\,s^{-1}Mpc^{-1}}$), but with a complex nonlinear supercluster network of walls, filaments and voids on smaller scales.

---

*The analysis done in this chapter used all 2dFGRS data available in December 2000

The origin of this large-scale structure is one of the key issues in cosmology. A plausible assumption is that structure grows via gravitational collapse of density fluctuations that are small at early times – but it is vital to test this idea. One important signature of gravitational instability is that collapsing structures should generate 'peculiar' velocities, $\delta\mathbf{v}$, which distort the uniform Hubble expansion. We measure a redshift, z, which combines Hubble's law with the radial component of these peculiar velocities: $cz \simeq H_0 r + \delta\mathbf{v}\cdot\hat{\mathbf{r}}$. The apparent density field seen in a redshift survey is thus not a true three-dimensional picture, but this can be turned to our advantage. The redshift-space distortions have a characteristic form, whose detection can both verify the general idea that structure forms by gravitational instability, and also measure the density of the universe. The present chapter presents measurements of this effect, based on the 2dFGRS. The full survey is briefly outlined in chapter 1, whereas chapter 2 gives a more detailed description of the survey and the observing strategy. In the analysis presented hereafter we make full use of the masks developed in section 2.3, as well as the weighting schemes described in section 2.5.

This chapter is organized as follows: in section 5.2, we present clustering results in redshift space for a magnitude limited sample, which is followed in section 5.3 by a quantitative analysis of the redshift space distortions, with a measure of the flattening parameter β . We end this chapter by comparing our estimate of β within an independent estimate inferred from the real space galaxy power spectrum and anisotropies of the microwave background.

5.2 Galaxy correlations in redshift space

The simplest statistical indicator of peculiar velocities in cosmological structure is the two-point correlation function, $\xi(\sigma,\pi)$. This measures the excess probability over random of finding a pair of galaxies with a transverse separation σ and a line-of-sight separation π. In an isotropic universe, this function should be independent of direction, but this is not true in redshift space. Transverse separations are true measures of distance, but apparent radial separations are distorted by peculiar velocities. This redshift-space anisotropy should cause two characteristic effects, operating respectively on small and large scales. On small scales, random orbital velocities within galaxy groups cause an apparent radial smearing, known as 'fingers of God'. Of greater interest is the large-scale effect; if cosmological structure forms via gravitational collapse, there should exist coherent infall

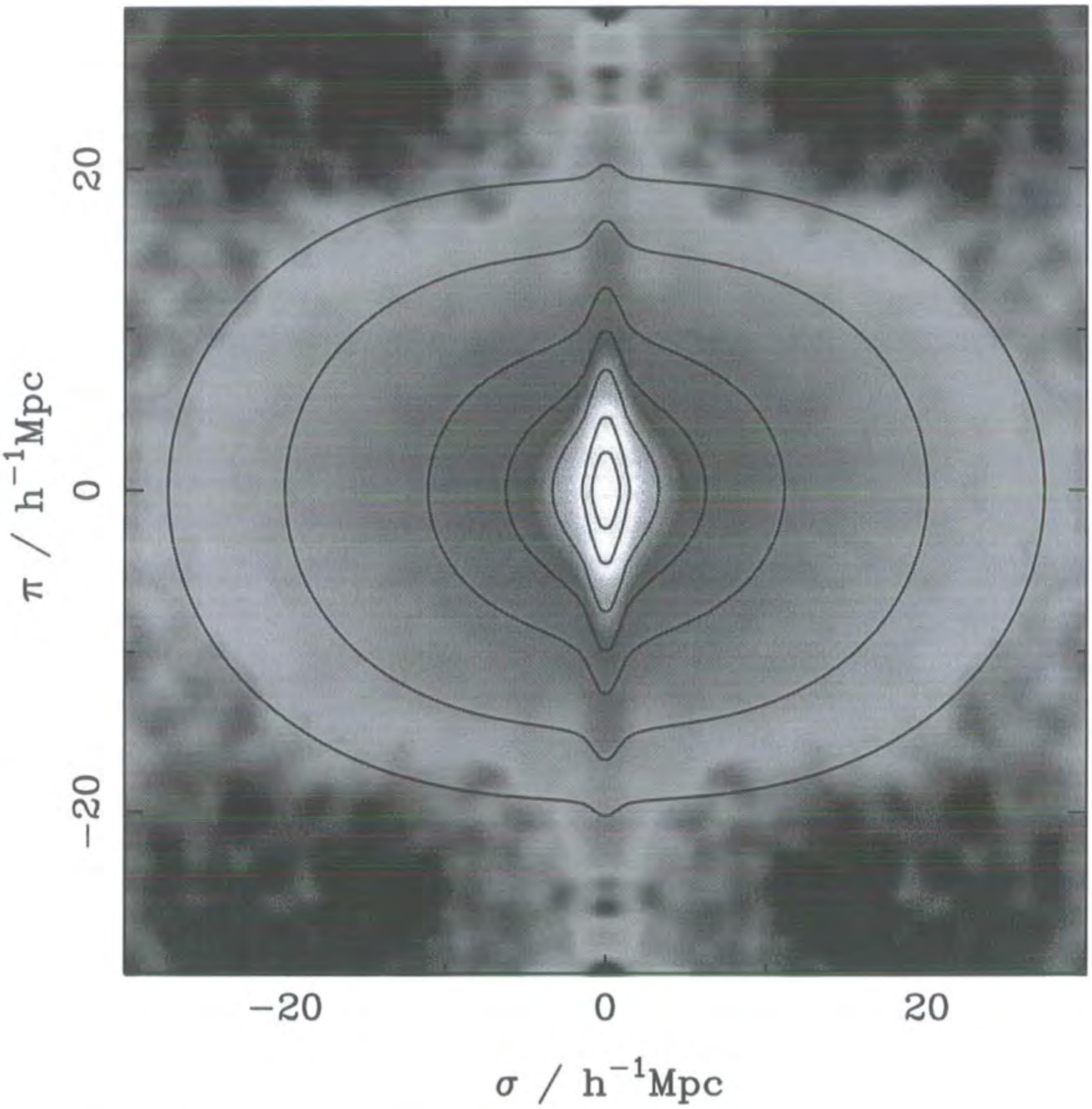


Figure 5.1: The redshift-space correlation function for the 2dFGRS, $\xi(\sigma, \pi)$, plotted as a function of transverse (σ) and radial (π) pair separation. The function was estimated by counting pairs in boxes of side $0.2\,h^{-1}$ Mpc (assuming a flat geometry), and then smoothing with a Gaussian of rms width $0.5\,h^{-1}$ Mpc. To illustrate deviations from circular symmetry, the data from the first quadrant are repeated with reflection in both axes. This plot clearly displays redshift distortions, with 'fingers of God' elongations at small scales and the coherent Kaiser flattening at large radii. The overplotted contours show model predictions (see text) with flattening parameter $\beta \equiv \Omega^{0.6}/b = 0.4$ and a pairwise dispersion of $\sigma_p = 400\,\mathrm{km\,s^{-1}}$. Contours are plotted at $\xi = 10, 5, 2, 1, 0.5, 0.2, 0.1$.

velocities, and the effect of these is to cause an apparent *flattening* of structures along the line of sight. The general existence of redshift-space distortions was recognized in the first redshift surveys (Kirshner et al. 1981; Davis & Peebles 1983; Bean et al. 1983), but the first comprehensive analysis of the phenomenon was performed by Kaiser (1987), who showed that they could be used to measure the quantity

$$\beta \equiv \Omega^{0.6}/b,$$

where Ω is the cosmological mass density parameter and b is the bias parameter that relates the relative density fluctuations of the galaxies and of the total mass:

$$\left.\frac{\delta\rho}{\rho}\right|_{\text{galaxies}} = b\left.\frac{\delta\rho}{\rho}\right|_{\text{mass}} .$$

The presence of bias is an inevitable consequence of the nonlinear nature of galaxy formation, and the relation between mass and galaxy tracers is complex (Hamilton, Tegmark & Padmanabhan 2000; Pen 1998; Dekel & Lahav 1998). However, there are good theoretical reasons to expect that b can indeed be treated as a constant on large scales, where the density fluctuations are linear (Benson et al. 2000b; Kauffmann, Nusser & Steinmetz 1997). Redshift-space distortions have thus been seen as an important method for weighing the universe (Strauss & Willick 1995; Hamilton 1998). To date, a number of papers have made significant detections of the Kaiser effect (Hamilton, Tegmark & Padmanabhan 2000; Outram, Hoyle & Shanks 2001), but the 2dFGRS is the first survey that is large enough for the effect to be studied in detail.

In order to estimate $\xi(\sigma, \pi)$, we follow standard methods (Hamilton 1993; Landy & Szalay 1993) that compare the observed count of galaxy pairs with the count estimated using a random distribution that obeys the same selection effects in redshift and sky position. These selection effects are well defined, but complex as explained in detail in chapter 2. We present here a brief summary of the relevant parts of that chapter for this analysis done in this chapter. The survey is tessellated into a pattern of 'sectors' defined by the overlap of the 2° diameter survey tiles, whose positions are chosen adaptively with the aim of being able to place a fibre on $> 95\%$ of the galaxies in the input catalogue. At the present intermediate stage of the survey, many tiles remain to be observed, and some regions of the survey presently contain redshifts for $< 50\%$ of the galaxies. Furthermore, the spectroscopic success rate (redshifts per allocated fibre) is $> 95\%$ in good conditions, but can fall to $\simeq 80\%$ in marginal weather (as shown by Fig. 2.11). We have implemented a number of independent algorithms for estimating the resulting survey selection effects (explained both in chapter 2 and in chapter 3), and are confident that we can measure the galaxy correlations robustly out to a separation of $25\,h^{-1}$ Mpc. For example, the redshift distribution in sectors with low spectroscopic completeness is biased to low redshifts, but it makes no significant difference whether or not we correct for this, or indeed whether the low-completeness regions are simply excised. In addition to allowing for survey completeness, it is necessary to give higher weight to regions with a low sampling density, to achieve the optimum balance between cosmic variance and shot noise (Kaiser 1986). In practice, we have chosen to truncate the analysis at a maximum redshift of $z = 0.25$. Within this volume, the exact optimum weight per galaxy varies very nearly as the reciprocal of the number density, so that all volume elements receive approximately equal weight. The redshift-space correlation function for the 2dFGRS computed in this way is shown in Fig. 5.1. The correlation-function results display very clearly the two signatures of redshift-space distortions discussed above. The 'fingers of God' from small-scale random velocities are very clear, as indeed has been the case from the first redshift surveys (Davis & Peebles 1983). However, this is the first time that the detailed signature of large-scale flattening from coherent infall has been seen with high signal-to-noise.

5.3 Quantifying redshift-space distortions

The large-scale flattening of the correlation function may be quantified by measuring the quadrupole moment of $\xi(\sigma, \pi)$ as a function of radius. More generally the multipole moments of the correlation function are defined as

$$\xi_\ell(r) \equiv (2\ell+1)/2 \int_{-1}^{1} \xi(\sigma = r\sin\theta, \pi = r\cos\theta)\; P_\ell(\cos\theta)\; d\cos\theta, \tag{5.1}$$

where $P_\ell(\cos\theta)$ is the Legendre polynomial of order ℓ. The quadrupole-to-monopole ratio should be a clear indicator of coherent infall. In linear theory, it is given by

$$\xi_2/\xi_0 = f(n)\,\frac{4\beta/3 + 4\beta^2/7}{1 + 2\beta/3 + \beta^2/5}. \tag{5.2}$$

(Hamilton 1992). Here $f(n) = (3+n)/n$, where n is the power-spectrum index of the density fluctuations: $\xi \propto r^{-(3+n)}$.

A negative quadrupole moment implies flattening, whereas the finger-of-God distortion tends to yield a positive quadrupole moment. Fig. 5.2 shows that the quadrupole-to-monopole ratio is positive on small scales, but that it falls with separation, becoming progressively more negative up to the largest separations at which it can be reliably measured. This arises partly because the underlying power spectrum is not a simple power

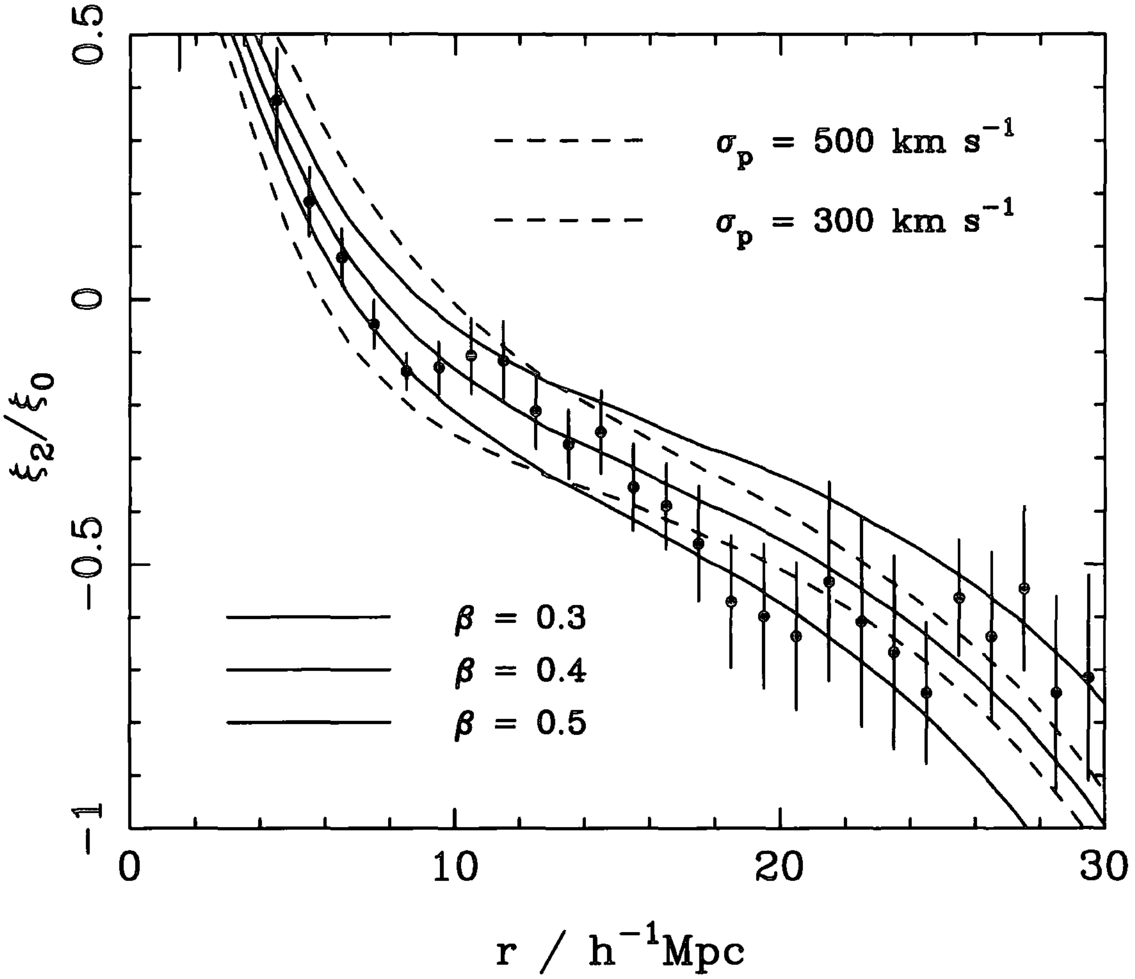


Figure 5.2: The flattening of the redshift-space correlation function is quantified by the quadrupole-to-monopole ratio, ξ_2/ξ_0. This quantity is positive where fingers-of-God distortion dominates, and is negative where coherent infall dominates. The solid lines show model predictions for $\beta = 0.3$, 0.4, and 0.5, with a pairwise velocity dispersion of $\sigma_p = 400\,\mathrm{km\,s^{-1}}$ (solid lines), plus $\beta = 0.4$ with $\sigma_p = 300$ and $500\,\mathrm{km\,s^{-1}}$ (dashed lines). The ξ_2/ξ_0 ratio becomes more negative as β increases and as σ_p decreases. At large radii, the effects of fingers-of-God become relatively small, and values of $\beta \simeq 0.4$ are clearly appropriate.

law function of scale, so that the peculiar velocities have a different effect at different radii. By integrating over the correlation function, it is possible to construct quantities in which this effect is eliminated. We shall not do this here, firstly because it seems desirable to keep the initial analysis as direct as possible. More importantly, finger-of-God smearing is a significant correction that will also cause the flattening to depend on radius. We therefore have to fit the data with a two-parameter model. The model predictions assume that the redshift-space power spectrum (P_s) may be expressed as a product of the linear Kaiser distortion and a radial convolution (Ballinger, Peacock & Heavens 1996):

$$P_s(\mathbf{k}) = P_r(k)\,(1+\beta\mu^2)^2\,(1+k^2\sigma_p^2\mu^2/2H_0^2)^{-1}, \tag{5.3}$$

where $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{r}}$, and σ_p is the rms pairwise dispersion of the random component of the galaxy velocity field. This model gives a very accurate fit to exact nonlinear simulations (Hatton & Cole 1998). For the real-space power spectrum, $P_r(k)$, we take the estimate obtained by deprojecting the angular clustering in the APM survey (Maddox, Efstathiou & Sutherland 1996; Baugh & Efstathiou 1994). This agrees very well with estimates that can be made directly from the 2dFGRS (Percival et al. 2001). We use this model only to estimate the scale dependence of the quadrupole-to-monopole ratio (although Fig. 5.1 shows that it does match the full $\xi(\sigma,\pi)$ data very well).

The parameters are β and a measure of the size of the random dispersion in the relative velocities of galaxies, σ_p. In practice, σ_p plays the role of an empirical fitting parameter to describe the scale on which the distortions approach the linear-theory predictions. It therefore also incorporates other possible effects, such as a scale dependence of bias.

The results for the quadrupole-to-monopole ratio are shown in Fig. 5.2, which shows the average of the estimates for the NGP and SGP slices. The difference between the NGP and SGP allows an estimate of the errors to be made: these slices are independent samples for the present analysis of clustering on relatively small scales. For model fitting, it is necessary to know the correlation between the values at different r. A simple way of addressing this is to determine the effective number of degrees of freedom from the value of χ^2 for the best-fitting model. A more sophisticated approach is to generate realizations of $\xi(\sigma,\pi)$, and construct the required covariance matrix directly. One way of achieving this is to analyze large numbers of mock surveys drawn from numerical simulations (Cole et al. 1998), which we do in chapters 6 and 7 when measuring the correlation length as function of luminosity. A more convenient method (computationally affordable) is to generate direct realizations of the redshift-space power spectrum, using Gaussian fluctuations on

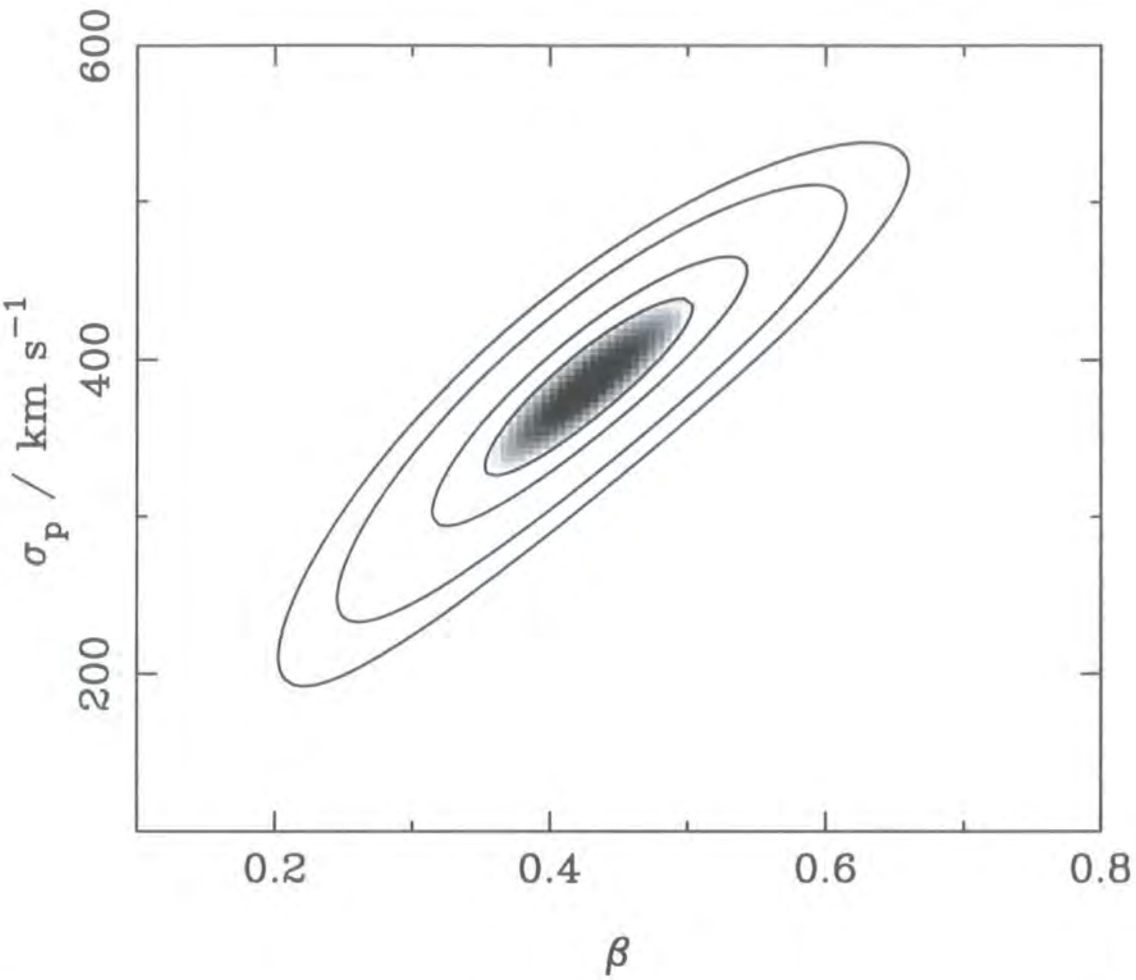


Figure 5.3: Likelihood contours for β and the fingers-of-God smearing parameter σ_p, based on the data in Fig. 5.2 (considering $8\,h^{-1}\,\mathrm{Mpc} < r < 25\,h^{-1}\,\mathrm{Mpc}$). These are plotted at the usual positions for one-parameter confidence of 68% (shaded region), and two-parameter confidence of 68%, 95% and 99% (i.e. $\Delta\chi^2 = 1, 2.3, 6.0, 9.2$). The maximum-likelihood solution is $\beta = 0.43$ and $\sigma_p = 385\,\mathrm{km\,s^{-1}}$. The value for the large-scale pairwise dispersion is in reasonable agreement with previously suggested values (Jing, Mo & Börner 1998). If we marginalize over σ_p (i.e. integrate over σ_p, treating the likelihood as a probability distribution), the final estimate of β and its rms uncertainty is $\beta = 0.43 \pm 0.07$.

large scales, but allowing for enhanced variance in power on non-linear scales (Feldman, Kaiser & Peacock 1994; Meiksin & White 1999; Scoccomarro, Zaldarriaga & Hui 1999). In practice, the likelihood contours resulting from this approach agree well with those from the simple approach, and we are confident that the resulting errors on β are realistic. These contours are shown in Fig. 5.3, and show that there is a degree of correlation between the preferred values of β and σ_p, as expected. For present purposes, σ_p is an uninteresting parameter, so we marginalize over it to obtain the following estimate of β and its rms uncertainty:

$$\beta = 0.43 \pm 0.07.$$

This result is the first precise determination of β from redshift-space distortions. The best previous studies (Hamilton, Tegmark & Padmanabhan 2000; Outram, Hoyle & Shanks 2001) have achieved no more than a detection of the effect at the 3σ level. We believe that this result is robust, in the sense that systematic errors in the modelling are smaller than the random errors. We have tried assuming that the power spectrum for $k < 0.1\,h\,\mathrm{Mpc}^{-1}$ has the shape of a $\Omega = 0.3$ ΛCDM model, rather than the APM measurement; this has a very small effect. A more serious issue is whether the pairwise velocity dispersion of galaxies may depend strongly on separation, as is found for mass particles in numerical simulations (Jenkins et al. 1998). Assuming that the pairwise velocity dispersion σ_p rises to twice its large-scale value below $1\,h^{-1}$ Mpc reduces the best-fit β by 0.04. This correction is small because our analysis excludes the nonlinear data at $r < 8\,h^{-1}$ Mpc.

Before discussing the implications of our result, we should therefore consider some possible small systematic corrections that have been unimportant in earlier work. First, the Kaiser analysis applies only in the small-angle approximation, and in principle corrections might be needed for wide-angle surveys such as ours (Szalay, Matsubara & Landy 1998). However, with our weighting scheme, the mean angular separation of pairs with spatial separations $< 30\,h^{-1}$ Mpc is only $2.5°$, so this is not a concern. There is potentially a significant correction for luminosity effects. The optimal weighting means that our mean luminosity is high: it is approximately $M_{b_J} = -20.3$, or 1.9 times the characteristic luminosity, $L^\star$, of the overall galaxy population (chapter 3; see also Folkes et al. 1999). A number of studies (Loveday et al. 1995; Benoist et al. 1996; Norberg et al. 2001) have suggested that the strength of galaxy clustering increases with luminosity. This effect has been controversial, as explained in more detail in chapter 6, but the 2dFGRS dataset undeniably favours a luminosity dependence of the galaxy clustering. This effective bias is well fitted by $b/b^\star = 0.85 + 0.15\,L/L^\star$ (see section 6.5). We therefore expect that β

for $L^\star$ galaxies will exceed our directly measured figure. Applying a correction using the given formula for $b(L)$, we deduce

$$\beta(L = L^\star) = 0.49 \pm 0.08.$$

Finally, the 2dFGRS has a median redshift of 0.11. With weighting, the mean redshift in the present analysis is $\bar{z} = 0.17$, and our measurement should be interpreted as β at this epoch. The extrapolation to $z = 0$ is model-dependent, but probably does not introduce a significant change (Carlberg et al. 2000).

5.4 Consistency with microwave-background anisotropies

These results are significant in a number of ways. First, we have verified in some detail that the pattern of redshift-space distortions associated with the gravitationally-driven growth of clustering exists as predicted. Although gravitational instability is well established as the standard model for the formation of large-scale structure, it is an important landmark to have verified such a characteristic feature of the theory. Extracting the full cosmological implications of our measurement of $\Omega^{0.6}/b$ requires us to know the bias parameter in order to determine Ω. For example, our measurement implies $\Omega = 0.30 \pm 0.09$ if $L^\star$ galaxies are unbiased, but it is difficult to justify such an assumption. In principle, the details of the clustering pattern in the nonlinear regime allow the $\Omega - b$ degeneracy to be broken, yielding a direct determination of the degree of bias (Verde et al. 1998). For the present, however, it is interesting to use an independent approach. Observations of anisotropies in the cosmic microwave background (CMB) can in principle measure almost all the cosmological parameters, and current small-scale anisotropy results are starting to tighten the constraints. In a recent analysis (Jaffe et al. 2001), best-fitting values for the densities in collisionless matter (c), baryons (b), and vacuum (Λ) have been obtained: $\Omega_c + \Omega_b + \Omega_\Lambda = 1.11 \pm 0.07$, $\Omega_c h^2 = 0.14 \pm 0.06$, $\Omega_b h^2 = 0.032 \pm 0.005$, together with a power-spectrum index $n = 1.01 \pm 0.09$. Our result for β gives an independent test of this picture, as follows.

The only parameter left undetermined by the CMB data is the Hubble constant, h. Recent work (Mould et al. 2000; Freedman et al. 2001) indicates that this is now determined to an rms accuracy of 10%, and we adopt a central value of $h = 0.70$. This completes the cosmological model, requiring a total matter density parameter $\Omega \equiv \Omega_c + \Omega_b = 0.35 \pm 0.14$. It is then possible to use the parameter limits from the CMB to predict

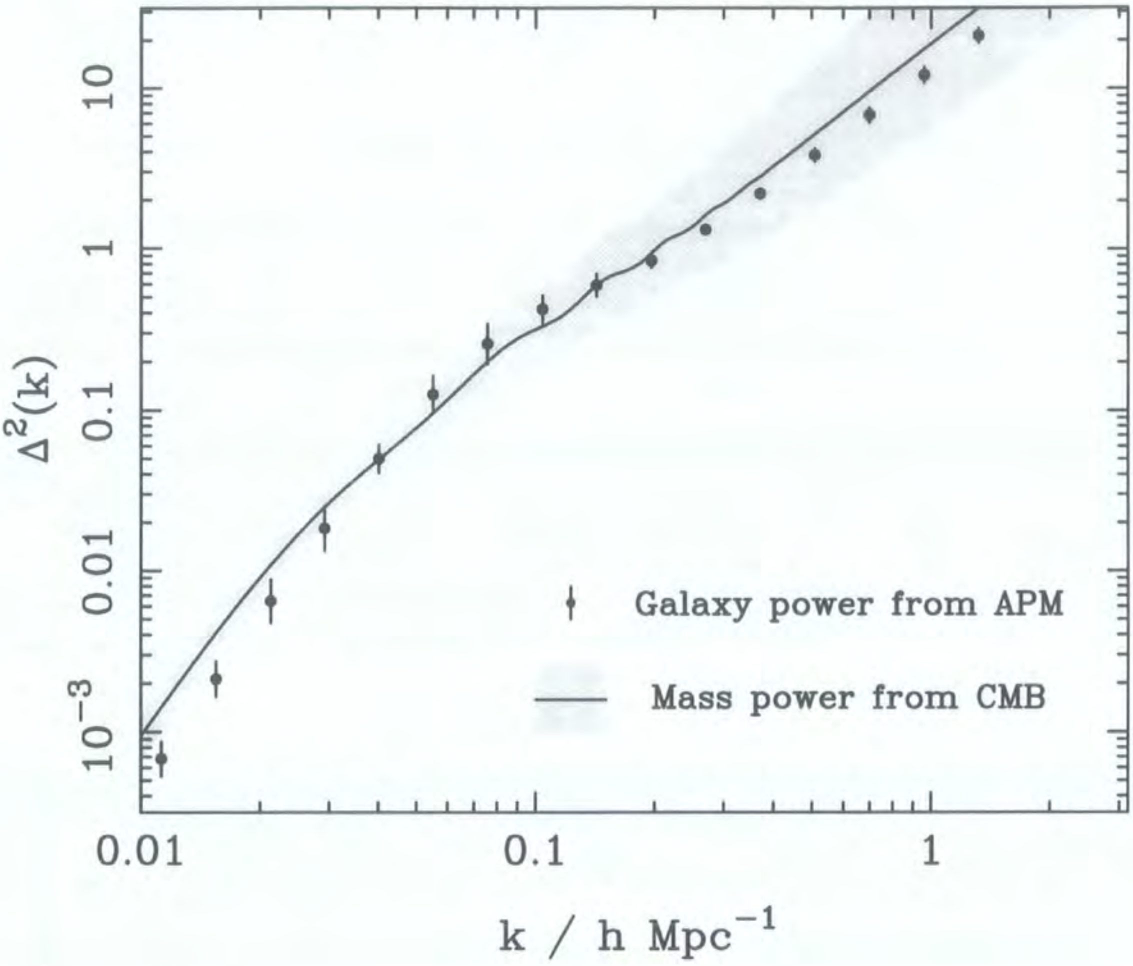


Figure 5.4: The dimensionless matter power spectrum at zero redshift, $\Delta^2(k)$, as predicted from the allowed range of models that fit the microwave-background anisotropy data, plus the assumption that $H_0 = 70\,\mathrm{km\,s^{-1}Mpc^{-1}} \pm 10\%$. The solid line shows the best-fit model (Jaffe et al. 2001) (power-spectrum index $n = 1.01$, and density parameters in baryons, CDM, and vacuum of respectively 0.065, 0.285, 0.760). The effects of nonlinear evolution have been included (Peacock & Dodds 1996). The shaded band shows the 1σ variation around this model allowed by the CMB data. The solid points are the real-space power spectrum measured for APM galaxies. The clear conclusion is that APM galaxies are consistent with being essentially unbiased tracers of the mass on large scales. Since the CMB data also constrain the range of Ω, this allows β to be predicted.

a conservative range for the mass power spectrum at $z = 0$, which is shown in Fig. 5. A remarkable feature of this plot is that the mass power spectrum appears to be in good agreement with the clustering observed in the APM survey. For each model allowed by the CMB, we can predict both b (from the ratio of galaxy and mass spectra) and also β (since a given CMB model specifies Ω). In practice, we determine b by determining the mean ratio of power spectra over the range $0.02 < k < 0.1\,h\,\mathrm{Mpc}^{-1}$, where the APM measurement is robust and where scale-dependent bias and nonlinearities should be unimportant. Considering the allowed range of models, we then obtain the prediction $\beta_{\mathrm{CMB+APM}} = 0.57 \pm 0.17$. A flux-limited survey such as the APM will have a mean luminosity close to $L^\star$, so the appropriate comparison is with the 2dFGRS corrected figure of $\beta = 0.49 \pm 0.08$ for $L^\star$ galaxies. These numbers are, at a 1σ level, in good agreement.

This analysis of galaxy clustering in the 2dFGRS thus gives strong support to the simplest picture of cosmological structure formation, in which the primary mechanism is gravitational instability in a sea of collisionless dark matter. We have shown that the fluctuations seen in the CMB (which measure structure at a redshift $z \simeq 1100$) can be extrapolated to the present to predict the peculiar velocities that distort redshift-space clustering. The agreement between this extrapolation and direct observations from the 2dFGRS is a remarkable and highly non-trivial test of the basic model. The precision of data in both areas should improve rapidly, and the use of β as a meeting ground between studies of the CMB and large-scale structure will undoubtedly lead to more demanding tests of the theory in years to come. For the present, we can say that there is complete consistency between clustering in the 2dFGRS and the emerging 'standard model' of cosmology: a spatially flat, vacuum-dominated universe with density parameter $\Omega \simeq 0.3$.

Chapter 6

Dependence of Galaxy Clustering on Luminosity

In this chapter, we investigate the dependence of the strength of galaxy clustering on intrinsic luminosity using the two degree field galaxy redshift survey (2dFGRS). We measure the projected two-point correlation function of galaxies in a series of volume-limited samples. The projected correlation function is free from any distortion of the clustering pattern induced by peculiar motions and we find that it is well described by a power-law in pair separation over the range $0.1 < (r/\,h^{-1}\,\mathrm{Mpc}) < 10$. The clustering of $L^{\star}$ ($M_{b_J} - 5\log_{10} h = -19.7$) galaxies in real space is well fit by a correlation length $r_0 = 4.9 \pm 0.3\,h^{-1}$ Mpc and power-law slope $\gamma = 1.71 \pm 0.06$. The clustering amplitude increases slowly with absolute magnitude for galaxies fainter than $M^{\star}$, but rises more strongly at higher luminosities. At low luminosities, our results agree with measurements from the SSRS2 by Benoist et al. However, we find a weaker dependence of clustering strength on luminosity at the highest luminosities. The correlation function amplitude increases by a factor of 4.0 between $M_{b_J} - 5\log_{10} h = -18$ and -22.5, and the most luminous galaxies are 3.0 times more strongly clustered than $L^{\star}$ galaxies. The power-law slope of the correlation function shows remarkably little variation for samples spanning a factor of 20 in luminosity. Our measurements are in very good agreement with the predictions of the hierarchical galaxy formation models of Benson et al. (2001).

6.1 Introduction

A major issue to be addressed by any successful theory of the formation of large scale structure is the problem of how galaxies trace the distribution of matter in the Universe. Measurements of differential galaxy clustering as a function of colour (Willmer et al. 1998), morphological type (Davis & Geller 1976; Iovino et al. 1993) and selection passband

(Peacock 1997; Hoyle et al. 1999) imply the existence of biases between the distributions of galaxies and mass.

A generic prediction of hierarchical structure formation models is that rarer objects should be more strongly clustered than average (Davis et al. 1985; White et al. 1987). Correspondingly, if more luminous galaxies are associated with more massive haloes, then these galaxies are expected to exhibit stronger clustering than the galaxy population as a whole. For the special case of bright galaxies at high redshift, see for example Baugh et al. 1998 and Governato et al. 1998. However, the form of the dependence of the amplitude of galaxy clustering on luminosity remains controversial even after more than twenty years of constructing and analysing redshift surveys of the local Universe. In the literature, claims of a dependence of galaxy clustering on luminosity (e.g. Davis et al. 1988; Hamilton 1988; Maurogordato & Lachieze-Rey 1991; Park et al. 1994; Benoist et al. 1996; Willmer et al. 1998; Guzzo et al. 2000) have been made with similar regularity to claims of non-detections (e.g. Phillipps & Shanks 1987; Hasegawa & Umemura 1993; Loveday et al. 1995; Szapudi et al. 2000; Hawkins et al. 2001). Part of the reason for this disagreement is a mismatch in the range of luminosities and clustering length scales considered in some of those studies. However, the main problem with earlier work is the small size of the redshift surveys analysed, both in terms of volume and number of galaxies. With previous surveys, the dynamic range in luminosity for which clustering can be measured reliably is limited, particularly when volume-limited samples are used. Due to the small volumes probed, it has generally not been possible to compare the clustering of galaxies of different luminosity measured within the same volume. Thus, these results have generally been affected by sampling fluctuations that are difficult to quantify. This problem is compounded by underestimation of the errors on the measured correlation functions and on the power-law fits traditionally employed in this subject.

In this chapter, we use the largest extant local survey, the Anglo-Australian two degree field galaxy redshift survey (hereafter 2dFGRS), to address the issue of how clustering depends upon galaxy luminosity. We describe the galaxy sample and the construction of volume-limited samples in section 6.2, and our estimation of the correlation function is described in section 6.3. Our results for the real space correlation function are given in section 6.4. We compare our results with those from previous studies and with the predictions of simulations of hierarchical galaxy formation in section 6.5.

6.2 The Data

6.2.1 The 2dFGRS sample

In this chapter, we use 2dFGRS galaxy redshifts obtained prior to January 2001, over 160 000 in total. As we are mainly interested in measuring clustering out to separations of order $20\,h^{-1}$ Mpc, we do not include galaxies that lie in the random fields in our analysis. A detailed description of the 2dFGRS is given in chapter 2.

In order to select an optimal sample for the measurement of the two point correlation function, we apply a weighting scheme to objects in the 2dFGRS. A weight is assigned to each measured redshift based upon the redshift completeness mask, the construction of which is explained in detail in section 2.3. We require a relatively high completeness in a given direction on the sky, so that, in practice, our results are fairly insensitive to the precise details of the weighting scheme. Excluding areas below our completeness threshold (which arise mainly as a result of the tiling strategy adopted to make optimal use of telescope time, coupled with the fact that the survey is not yet finished), we estimate the effective solid angle used in the SGP region is $\sim 420\ \square^\circ$, and in the NGP $\sim 190\ \square^\circ$.

6.2.2 Constructing a volume-limited sample

We analyse here a series of volume-limited subsamples drawn from the 2dFGRS. The advantage of this approach is that the radial selection function is uniform, and the only variations in the space density of galaxies within each volume are due to clustering. By contrast, in a flux-limited survey, the galaxy number density is a strong function of radial distance and this needs to be corrected for when measuring the clustering. The disadvantage of using a volume-limited sample is that a large number of galaxies in the flux-limited survey do *not* satisfy the selection cuts (which are explained below). This was a serious problem for previous surveys, but not for a survey the size of the 2dFGRS. As we demonstrate in section 6.4, the volume-limited samples we analyse give robust clustering measurements and contain over an order of magnitude more galaxies than similar samples constructed from previous surveys (see Table 6.1).

The construction of a volume-limited sample drawn from a flux-limited redshift survey requires a range of absolute magnitudes to be specified. Since a flux-limited survey has both bright and faint apparent magnitude limits, the selected range of absolute magnitudes requires that both a minimum ($z_{\rm min}$) and a maximum ($z_{\rm max}$) redshift cut be applied to the volume-limited sample. Thus, in principle, a galaxy included in the volume-limited

Table 6.1: Properties of the combined NGP & SGP volume-limited sub-samples analysed. The second column gives the median magnitude of each sample. Columns 6 and 7 list the best fitting correlation length, r_0, and power-law slope γ of the correlation function in real space, fitted over the range $0.5 \leq \sigma/(h^{-1}\,\mathrm{Mpc}) \leq 10$. Column 8 gives the value of $A(\gamma)$, defined by Eq. 6.4, evaluated for the best fitting value of γ.

Mag. range		Med. mag.	N_{gal}	$z_{\min}$	$z_{\max}$	r_0	γ	$A(\gamma)$
$M_{b_J} - 5\log_{10} h$						(h^{-1} Mpc)		
−18.0	− 18.5	−18.11	7061	0.010	0.086	4.14 ± 0.64	1.78 ± 0.10	3.75
−18.5	− 19.0	−18.61	9382	0.013	0.104	4.43 ± 0.45	1.75 ± 0.08	3.80
−19.0	− 19.5	−19.11	13690	0.016	0.126	4.75 ± 0.44	1.68 ± 0.08	4.14
−19.5	− 20.0	−19.60	15123	0.020	0.152	4.92 ± 0.27	1.71 ± 0.06	4.01
−20.0	− 20.5	−20.09	13029	0.025	0.182	5.46 ± 0.28	1.68 ± 0.06	4.14
−20.5	− 21.0	−20.58	9114	0.031	0.220	6.49 ± 0.29	1.63 ± 0.06	4.39
−21.0	− 21.5	−21.06	3644	0.039	0.270	7.58 ± 0.48	1.76 ± 0.09	3.82
−18.0	− 19.0	−18.22	12594	0.013	0.086	4.06 ± 0.53	1.79 ± 0.09	3.72
−19.0	− 20.0	−19.19	21874	0.020	0.126	4.75 ± 0.44	1.70 ± 0.08	4.06
−20.0	− 21.0	−20.13	17383	0.031	0.182	5.65 ± 0.30	1.69 ± 0.06	4.10
−21.0	− 22.0	−21.07	4013	0.048	0.270	8.12 ± 0.46	1.78 ± 0.12	3.75
−21.5	− 22.5	−21.55	1002	0.059	0.280	9.38 ± 1.48	1.69 ± 0.15	4.10

sample could be displaced to any redshift between $z_{\rm min}$ and $z_{\rm max}$ and still remain within the bright and faint *apparent* magnitude limits of the flux-limited survey.

In order to estimate the absolute magnitude of 2dFGRS galaxies at redshift zero, it is necessary to apply corrections for band shifting ($k-$correction) and evolution in the stellar populations ($e-$correction). We adopt a global $k + e$ correction of the form $k + e = 0.03z/(0.01 + z^4)$, which is a good fit to the correction calculated for the $\rm b_J$ selected ESO Slice Project survey using population synthesis models (see Fig. 1 of Zucca et al. 1997). This form for the $k + e$ correction gives consistent luminosity functions for the 2dFGRS when the survey is divided into redshift bins, indicating that it adequately accounts for the degree of evolution in galaxy luminosity over the lookback time spanned by the survey. Further details about $k + e$ corrections for the average galaxy sample are given in chapter 3. Our results are unchanged if we use the mean of the k corrections for different spectral types given by Madgwick et al. (2001), or the mean weighted k corrections use in chapter 7. The values of $z_{\rm min}$ and $z_{\rm max}$ that define a volume-limited sample drawn from the 2dFGRS vary slightly with position on the sky. This is due to revisions made to the map of galactic extinction (Schlegel, Finkbeiner & Davis 1998) and to the CCD calibration of APM plate zero points since the definition of the original input catalogue, as explained in more detail in §2.1.2 and §2.3.1. Throughout this chapter, we adopt an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology to convert redshift into comoving distance.

6.3 Estimating the Two-Point Correlation Function

The galaxy correlation function is estimated on a two dimensional grid of pair separations parallel (π) and perpendicular (σ) to the line-of-sight. To estimate the mean density of pairs, a catalogue of unclustered points is generated with the same angular selection and ($z_{\rm min}$, $z_{\rm max}$) values as the data. The correlation function is estimated by

$$\xi = \frac{DD - 2DR + RR}{RR}, \tag{6.1}$$

where DD, DR and RR are the suitably normalised number of weighted data-data, data-random and random-random pairs respectively in each bin (Landy & Szalay 1993).

Contours of constant clustering amplitude in the redshift space correlation function, $\xi(\sigma, \pi)$, are distorted as a result of the peculiar motions of galaxies, as explained in chapter 5. On small scales, random motions inside virialised structures elongate the constant-ξ contours in the π direction, whereas on large scales, coherent flows flatten

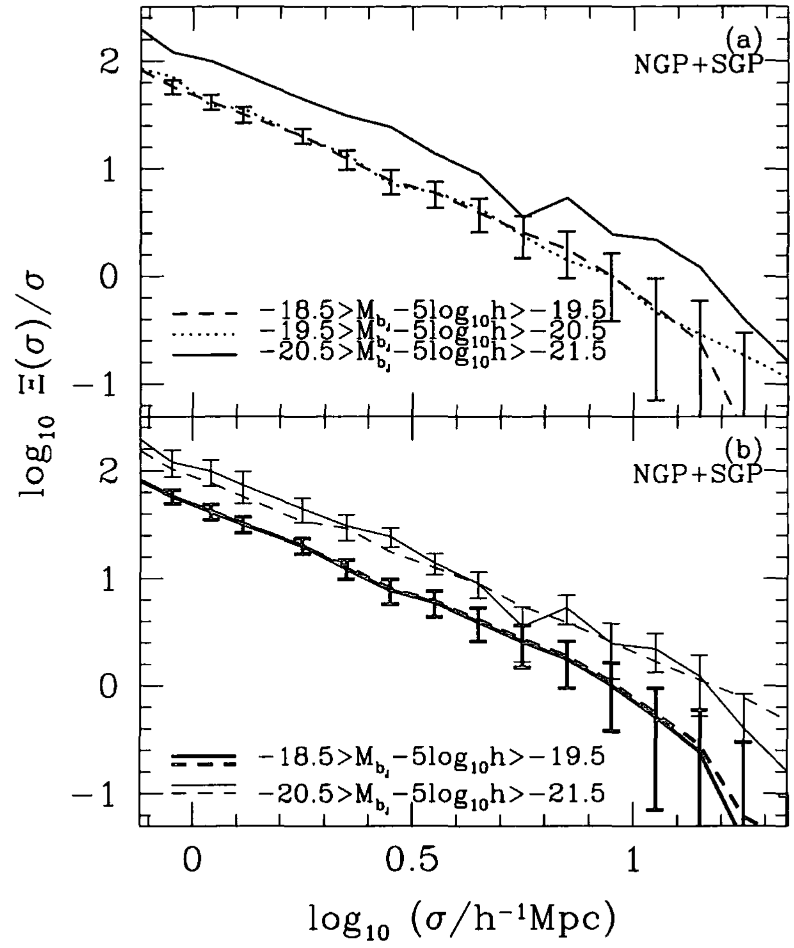


Figure 6.1: (a) The projected correlation function measured for galaxies in three different absolute magnitude bins in the *same* volume. The faintest sample contains 16 134 galaxies, the middle sample contains 6 186 galaxies and the brightest sample contains 985 galaxies. For clarity, error bars are plotted only on the correlation function of galaxies with $-18.5 \geq M_{b_J} - 5\log_{10} h \geq -19.5$. (b) A comparison of the correlation function of galaxies in the same absolute magnitude bins but measured in different (although not completely independent) volumes. The heavy lines show results for galaxies with $-18.5 \geq M_{b_J} - 5\log_{10} h \geq -19.5$ and the light lines show results for a brighter bin with $-20.5 \geq M_{b_J} - 5\log_{10} h \geq -21.5$. In each case, the dashed line shows the estimate from the optimal sample (see text) for the selected magnitude bin, whilst the solid line shows an estimate of the correlation function from the volume analysed in Fig. 6.1(a). For the $-20.5 \geq M_{b_J} - 5\log_{10} h \geq -21.5$ magnitude bin, the optimal estimate is measured using 10 962 galaxies, which should be contrasted with the 985 galaxies used to make the measurement shown by the light solid line.

the contours. The latter effect was measured clearly for the first time for galaxies using the 2dFGRS (Peacock et al. 2001). The dependence of the redshift space correlation function on galaxy luminosity is not going to be analysed in this chapter*. Thus, we consider here only clustering in real space, which we infer by projecting the measured correlation function along the line-of-sight. We compute a dimensionless quantity, $\Xi(\sigma)/\sigma$ by integrating over the measured $\xi(\sigma, \pi)$ grid:

$$\frac{\Xi(\sigma)}{\sigma} = \frac{1}{\sigma} \int_{-\infty}^{\infty} \xi(\sigma, \pi) \mathrm{d}\pi. \tag{6.2}$$

Note that $\Xi(\sigma)$ is sometimes referred to as $w(r_p)$ in the literature. In practice, the integral converges by a pair separation of $\pi = 75\,h^{-1}\,\mathrm{Mpc}$. The projected correlation function can, in turn, be written as an integral over the spherically averaged real space correlation function, $\xi(r)$,

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \int_{\sigma}^{\infty} \xi(r) \frac{r\mathrm{d}r}{(r^2 - \sigma^2)^{1/2}}, \tag{6.3}$$

(Davis & Peebles 1983). If the real space correlation function is a power-law (which is a reasonable approximation for APM galaxies out to separations around $r \sim 10\,h^{-1}\,\mathrm{Mpc}$, see e.g. Baugh 1996), then

$$\frac{\Xi(\sigma)}{\sigma} = \left(\frac{r_0}{\sigma}\right)^{\gamma} \frac{\Gamma(1/2)\Gamma([\gamma-1]/2)}{\Gamma(\gamma/2)} = \left(\frac{r_0}{\sigma}\right)^{\gamma} A(\gamma), \tag{6.4}$$

where $\xi(r) = (r_0/r)^{\gamma}$ and r_0 is the correlation length.

Previous studies have estimated the error on the measured correlation function from the Poisson statistics of the pair counts in each bin (Peebles 1980) or by bootstrap resampling of the data (e.g. Benoist et al. 1996). Since we study a range of samples corresponding to different luminosity bins and also compare samples from different volumes, it is important to include an estimate of the sampling fluctuations in the error budget for the correlation function. This we derive from analysis of 22 mock 2dFGRS catalogues constructed from the ΛCDM Hubble Volume dark matter simulation, in the manner explained by Baugh et al. (2001, in preparation; see also Cole et al. 1998). In order to mimic the clustering of the 2dFGRS, a biasing scheme is employed to select particles in the simulations with a probability which is a function of the final dark matter density field, smoothed with a Gaussian filter (model 2 of Cole et al. 1998). The mock

---

*Within the 2dFGRS, we agreed to split clustering analyses via the two-point correlation function into two parts. The Nottingham group have concentrated their efforts on redshift space clustering (Hawkins et al. 2001, in preparation), while the Durham group have considered clustering in real space.

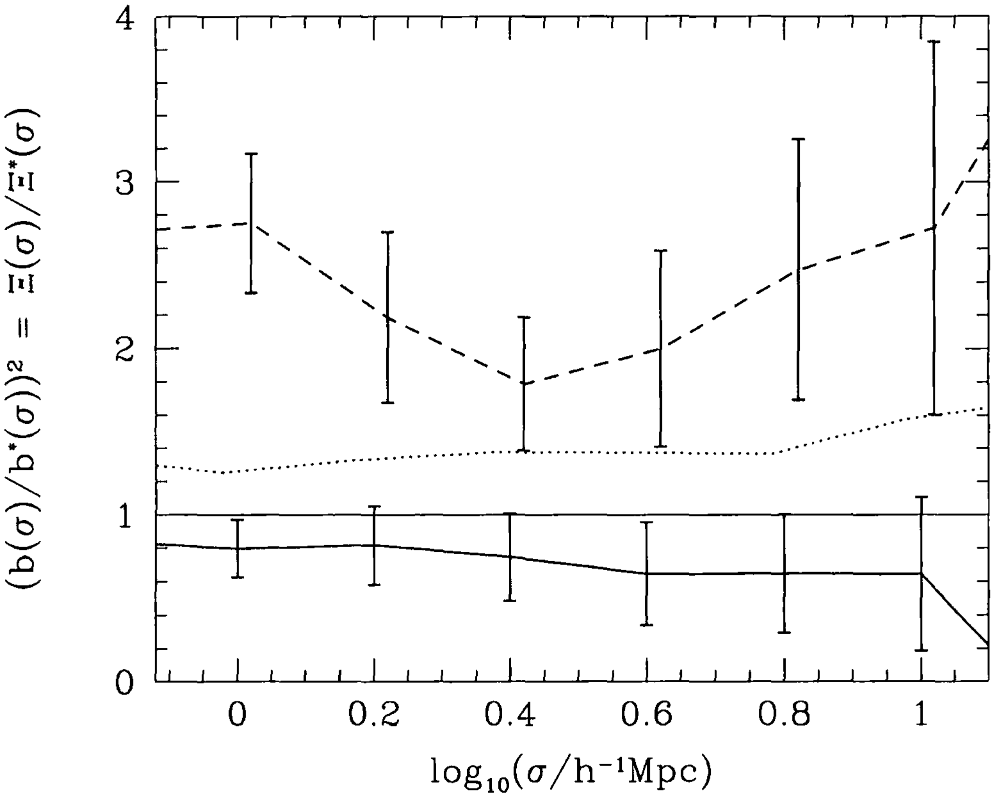


Figure 6.2: The ratio of the projected correlation function of galaxies in different magnitude slices to the projected correlation function of galaxies with $-19 \geq M_{b_J} - 5\log_{10} h \geq -20$. Note that the ratio is plotted on a linear scale, whilst the pair separation is on a log scale. The solid line shows the ratio for galaxies with absolute magnitudes in the range $-18 \geq M_{b_J} - 5\log_{10} h \geq -19$, the dotted line for $-20 \geq M_{b_J} - 5\log_{10} h \geq -21$ and the dashed line for $-21 \geq M_{b_J} - 5\log_{10} h \geq -22$. For clarity, error bars have been omitted from the dotted line but these are comparable in size with those plotted on the solid line.

catalogues have the same clustering amplitude as galaxies in the flux-limited 2dFGRS and the same selection criteria that are applied to the data are used in the construction of the mock surveys. The clustering amplitude in the mocks is independent of luminosity. The error bars that we plot on correlation functions measured from the 2dFGRS are the *rms* found by averaging over the 22 mock catalogues.

6.4 Results

We first demonstrate the robustness of the approach of measuring the correlation function in volume-limited samples. Unless stated otherwise, we have added the pair counts in the

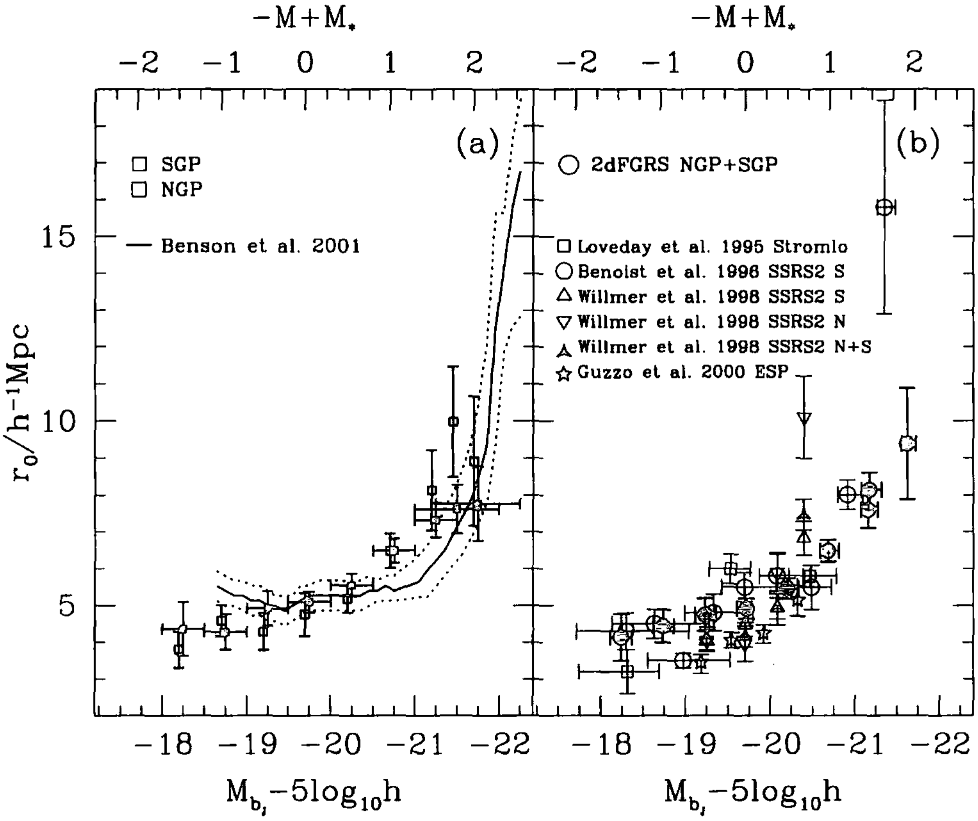


Figure 6.3: (a) The correlation length in real space as a function of absolute magnitude. Results are shown for the SGP and NGP regions separately. The NGP points are plotted with an offset of 0.05 mag for clarity. Horizontal error bars on the SGP points indicate the absolute magnitude range of each bin, and each point is plotted at the bin centre. In both cases, the brightest data points are for galaxies in one magnitude wide bins, whereas the other points are for galaxies in 0.5 magnitude wide bins. The solid line shows the predictions of the semi-analytic model of Benson et al. (2001), computed in a series of overlapping bins, each 0.5 magnitudes wide. The dotted curves show an estimate of the errors on this prediction, including the sample variance expected for a volume equal to that of the N-body simulation used. (b) The real space correlation length estimated combining pairs counts in the NGP and SGP (filled circles). The open symbols show a selection of recent data from other studies. The data for surveys selected in the B-band have been corrected to the b_J band using the approximate relation $M_{b_J} = M_B - 0.2$. In order to compare samples defined by cumulative and differential magnitude bins, the data points are plotted at the median magnitude of each sample.

NGP and SGP regions to compute correlation functions. In Fig. 6.1(a), we show the correlation function of galaxies in three disjoint absolute magnitude bins measured in the same volume. The sampling fluctuations are therefore virtually the same for each subsample, although the number of galaxies varies between them. There is a clear difference in the clustering amplitude of galaxies in the brightest absolute magnitude bin. Next, we demonstrate that sampling fluctuations are not important in a survey the size of the 2dFGRS. For this purpose, we show, in Fig. 6.1(b), the correlation function in two fixed absolute magnitude bins measured in different volume-limited sub-samples. Specifically, the dashed lines show the correlation function for the optimal volume-limited sample, appropriate to the selected absolute magnitude bin (see Table 6.1). Such a sample contains the maximum number of galaxies in that magnitude bin. The different estimates of the correlation function agree within the errors, demonstrating that our results are not sensitive to the choice of volume.

We now focus attention on the series of volume-limited subsamples covering the range $-18 \geq M_{b_J} - 5\log_{10} h \geq -22.5$, whose characteristics are listed in Table 6.1. The shape and amplitude of the projected correlation function in a selection of these samples is compared in Fig. 6.2 with the correlation function of galaxies in the magnitude range $-19 \geq M_{b_J} - 5\log_{10} h \geq -20$. The shape of the correlation function varies relatively little with the absolute magnitude that defines the sample in contrast to the amplitude of the correlation function, which changes significantly for the brightest magnitude slice. Another view of this trend is given in Fig. 6.3(a) where we plot the real space correlation length as a function of absolute magnitude. The best fitting values of the correlation length, r_0, and power-law slope γ, are determined by applying Eq. 6.4 to the measured correlation function over the pair separation range $0.5 \leq \sigma/(h^{-1}\,\mathrm{Mpc}) \leq 10$ and carrying out a χ^2 minimisation. This simple χ^2 approach will not, however, give reliable estimates of the errors on the fitted parameters due to the correlation between the estimates at differing pair separations. We use the mock 2dFGRS catalogues to estimate the errors on the fitted parameters. In brief, the best fitting values of r_0 and γ are found for each mock individually, using the simple χ^2 analysis. The estimated error bar is the *rms* scatter in the fitted parameters over the ensemble of mock catalogues.

In Fig. 6.3(a), we plot the correlation lengths for the NGP and SGP regions separately. These independent estimates are in excellent agreement with one another. The slope of the best fitting power-law correlation function, given in Table 6.1, is similar for all the volume-limited samples considered. The clustering amplitude increases slowly with luminosity for galaxies fainter than $M^\star$(where $M^\star = M_{b_J} - 5\log_{10} h = -19.7$, as found by our analysis in chapter 3 and by Folkes et al. 1999), but rises strongly at higher luminosities. The correlation function amplitude increases by a factor of 4 between $M_{b_J} - 5\log_{10} h = -18$ and -22.5, and the most luminous galaxies are 3 times more strongly clustered than $M^\star$galaxies.

6.5 Discussion

The volume-limited samples analysed in this chapter contain over an order of magnitude more galaxies than previous studies of the dependence of clustering on galaxy luminosity, allowing a more accurate measurement of this effect than was possible before. The sheer volume covered by our samples, $10^6 - 2\times 10^7\, h^{-3}\,\mathrm{Mpc}^3$, ensures that sampling fluctuations have little impact upon our results.

We compare the 2dFGRS results with a selection of recent measurements taken from the literature since 1995 in Fig. 6.3(b). To compare samples defined by cumulative and differential magnitude bins, we plot the datapoints at the median magnitude for the sample, as computed using the Schechter function parameters for the 2dFGRS luminosity function, as measured in chapter 3 and in Folkes et al. (1999). The horizontal bars plotted on selected points show the quartile range of the magnitude distribution in the sample. Benoist et al. (1996) analysed quasi volume-limited samples in the SGP region of the Southern Sky Redshift Survey 2 (SSRS2), and found a sharp increase in the correlation length for galaxies brighter than $M_B - 5\log_{10} h = -20.5$. The Benoist et al. correlation lengths are measured in redshift space, although the authors report that a similar trend with luminosity is seen in real space. Willmer et al. (1998) re-analysed the SSRS2 South using different volume limits and also measured clustering in the SSRS2 North, presenting fits for the correlation length in real and redshift space. Intriguingly, Willmer et al. find a larger correlation length in real space for galaxies with $M_B - 5\log_{10} h \sim -20$ than Benoist et al. find in redshift space. Moreover, the clear disagreement between the results for the brightest galaxies analysed in SSRS2 North and South suggests that sampling fluctuations are significant in a survey of this size and that the errors on these points have been underestimated (as demonstrated in Fig. 4 of Benson et al. 2001). Loveday et al. (1995) measured the clustering in real space by cross-correlating galaxies in the sparsely-sampled Stromlo/APM redshift survey with galaxies in the parent catalogue. Galaxies were considered in three absolute magnitude bins. No difference was found between the

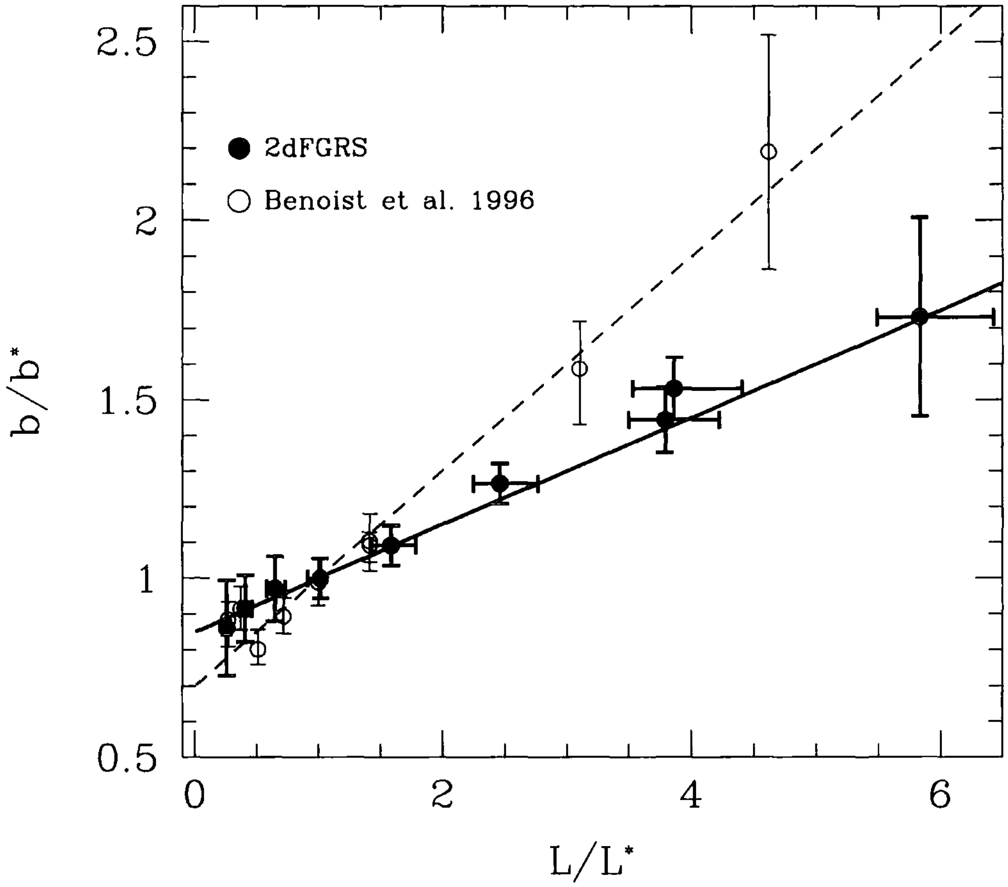


Figure 6.4: The variation of the relative bias as a function of luminosity, using the clustering of $L^{\star}$galaxies as a reference point (see text for definition). The 2dFGRS points are plotted at the median magnitude of each sample and the horizontal bars show the quartile magnitude range. The Benoist et al. (1996) points are taken from their Fig. 5 and are plotted at the median value of $L/L^{\star}$ for each sample. Note that the error bars given by Benoist et al. are obtained by averaging over correlated bins in pair separation. The curves shows parametric fits: the Benoist et al. measurements are well fitted by $b/b^{\star} = 0.7 + 0.3L/L^{\star}$ (dashed line), whereas the 2dFGRS results suggest a more modest dependence on luminosity: $b/b^{\star} = 0.85 + 0.15L/L^{\star}$ (solid line).

clustering amplitude of $L^\star$ and super-$L^\star$ galaxies. However, the median magnitude for the most luminous sample considered by these authors is only 0.5 magnitudes brighter than $M^\star$.

The increase in clustering amplitude with luminosity can be connected with a change in the mix of morphological types with increasing luminosity. The mix of spectral types at the brightest absolute magnitudes is dominated by spectra characteristic of elliptical galaxies, whereas spiral galaxies are more numerous around $L^\star$ (see Fig. 7.1 and Fig. 7.9, but also: Folkes et al. 1999; Cole et al. 2001; Madgwick et al. 2001). With the 2dFGRS, it is possible to analyse the clustering of galaxies as a bivariate function of luminosity and spectral type, which is the aim of chapter 7.

Our clustering results can be characterised in a concise way in terms of a relative bias parameter, $b/b^\star$, that gives the amplitude of the correlation function relative to that of $L^\star$galaxies (where $M^\star = M_{b_J} - 5\log_{10} h = -19.7$). The *relative* bias between the correlation functions of galaxies of different luminosity is assumed to be constant for pair separations spanned by the r_0 values listed in Table 6.1 (see also Fig. 6.2). The relative bias is then defined by $b/b^\star = (r_0/r_0^\star)^{\gamma/2}$, where we take $r_0^\star = 4.9 \pm 0.3\, h^{-1}$ Mpc from Table 6.1 and use $\gamma = 1.7$. The 2dFGRS results are shown by the filled symbols in Fig. 6.4 and are well fitted by the relation $b/b^\star = 0.85 + 0.15L/L^\star$. The 2dFGRS data suggest a significantly weaker dependence of the relative bias on luminosity than the Benoist et al. data, which follow the relation $b/b^\star = 0.7 + 0.3L/L^\star$ (Peacock et al. 2001). (The parametric fit to the Benoist et al. measurements was used by Peacock et al. 2001 to estimate the parameter $\beta = \Omega^{0.6}/b$ for $L^\star$galaxies in the 2dFGRS. Using the above fit to the 2dFGRS measurements changes the inferred value for β by less than 1σ to $\beta = 0.49 \pm 0.08$, as shown in section 5.3.)

Hierarchical models of galaxy formation predict that bright galaxies should be more strongly clustered than faint galaxies (e.g. White et al. 1987; Kauffmann, Nusser & Steinmetz 1997). This generic prediction arises because bright galaxies are expected to occupy more massive dark matter haloes and these haloes are more strongly clustered than the overall distribution of dark matter. The trend of clustering amplitude with luminosity measured for 2dFGRS galaxies is in very good agreement with the predictions of a simulation of hierarchical galaxy formation taken from Fig. 4 of Benson et al. (2001), reproduced as the solid line in Fig. 6.3(a). In the Benson et al. semi-analytic model, the input parameters are set in order to reproduce a subset of local galaxy data, with most emphasis given to the field galaxy luminosity function (see Cole et al. 2000). No reference is made to clustering data in setting the model parameters. In a ΛCDM cosmology, Benson et al. (2000a,b) find excellent agreement with the real space correlation function measured for galaxies in the APM survey by Baugh (1996). It is remarkable that the same model, without any readjustment of parameters, also reproduces the dependence of clustering amplitude on luminosity exhibited by the 2dFGRS in Fig. 6.3(a).

Chapter 7

Dependence of Galaxy Clustering on Luminosity and Spectral Type

We investigate the dependence of galaxy clustering on luminosity and spectral type using data from the 2dFGRS. Spectral types are assigned using the principal component analysis presented by Madgwick et al. (2001). We divide the sample into two broad spectral classes; galaxies with strong emission lines, referred to as late-types, and more quiescent galaxies, denoted as early-types. A series of volume limited samples is constructed and we measure the clustering in real space, free from any distortions to the spatial distribution of galaxies arising from peculiar motions. The projected correlation functions of both spectral types is well described by a power law for transverse pair separations in the range: $2.0 < (\sigma/h^{-1}\mathrm{Mpc}) < 15$. The clustering strength of both early- and late-type galaxies increases with luminosity at approximately the same rate: between $L^{\star}$ (defined by $M_{b_J} - 5\log_{10} h = -19.7$) and $4L_{\star}$ the clustering amplitude of early- and late-types changes by a factor of ~ 2.5. At a given luminosity, early-types have a correlation function amplitude that is typically 50% higher than that measured for late-types. We find some evidence that early-types have a steeper correlation function than late-types.

7.1 Introduction

One of the major scientific drivers behind the 2dFGRS is to make an accurate measurement of the spatial distribution of galaxies. In particular, the unprecedented size of the 2dFGRS makes it possible to subdivide the catalogue and therefore to quantify how the clustering signal changes as intrinsic galaxy properties, such as luminosity or star formation rate, are varied.

The motivation behind such a program is to provide constraints upon theoretical models of structure formation. In the current paradigm, galaxies form inside dark matter haloes that have been built up in a hierarchical way through mergers or by the accretion of smaller objects. The clustering pattern of galaxies is therefore determined by two processes: the spatial distribution of dark matter haloes and the manner in which dark matter haloes are populated by galaxies (Benson et al. 2000; Peacock & Smith 2001; Seljak 2001). The evolution of clumping in the dark matter has been studied extensively using N-body simulations of the growth of density fluctuations via gravitational instability (e.g. Jenkins et al. 1998; 2001). With the advent of powerful theoretical tools that can follow the formation and evolution of galaxies in the hierarchical scenario, the issue of how galaxies are partitioned amongst dark matter haloes can be addressed, and detailed predictions of the clustering of galaxies are now possible (Kauffmann, Nusser & Steinmetz 1997; Kauffmann et al. 1999; Benson et al. 2000a,b; Somerville et al. 2001). The models predict that early-type galaxies are more strongly clustered than late types. This phenomenon arises because of the segregation of morphological types with density, as observed in the Universe (Dressler 1980) and reproduced by the models (Baugh et al. 1996; Benson et al. 2001).

The first attempt to quantify the difference between the clustering signals of early and late type galaxies was made using a shallow angular survey, the Uppsala catalogue, with morphological types assigned to the galaxies following visual examination of the photographic plates (Davis & Geller 1976). Elliptical galaxies were found to have a steeper angular correlation function than spiral galaxies. More recently, the comparison of clustering for different types has been extended to three dimensions, using redshift surveys in which a large fraction of the galaxies have been assigned morphological types. Again, similar conclusions have been reached in these studies, namely that ellipticals have a stronger clustering amplitude than spirals (e.g. Lahav & Saslaw 1992; Santiago & Strauss 1992; Iovino et al. 1993; Hermit et al. 1996; Loveday et al. 1995; Guzzo et al. 1997; Willmer et al. 1998). The subjective process of visual classification can be replaced with an automated approach to quantify the shape of galaxy image; such a scheme, using a "concentration parameter" derived from the radius of different isophotes, has been applied to a sample of 30000 galaxies from the Sloan Digital Sky Survey by Zehavi et al. (2001). Again, based upon cuts in the distribution of concentration parameter, early types are found to be more clustered than late types.

In this chapter, we adopt a different approach to the labelling of galaxies, based upon a principal component analysis (PCA) of galaxy spectra (Madgwick, Lahav & Taylor 2000; Madgwick et al. 2001). This technique has a number of advantages over that of deriving morphological types from photometry. First, the PCA approach is completely objective and reproducible, as the process is well specified. This means that an equivalent analysis can, for example, be applied readily to spectra produced by theoretical models of galaxy formation or to spectra obtained in an independent redshift survey. Second, the PCA can be applied to the full magnitude range of the survey, whenever the spectra are of sufficient signal to noise (see §7.2.2); for the 2dFGRS, the quality of the photometry is only adequate to permit a determination of a morphological type for galaxies brighter than $b_J \sim 17$, which includes roughly only 5% of the spectroscopic sample.

Two previous clustering studies have used information derived from spectra to select galaxy samples. Loveday, Tresse & Maddox (1999) grouped galaxies in the Stromlo-APM redshift survey into three classes based upon the equivalent width of either the Hα or O[II] lines, and found that galaxies with prominent emission lines display weaker clustering than more quiescent galaxies. Tegmark & Bromley (1998) measured the relative bias between different spectral classes in the Las Campanas redshift survey, using the PCA derived classification produced by Bromley et al. (1998) (see also Blanton (2000) for a revision of Tegmark & Bromley's analysis, which takes into account the effect of errors in the survey selection function).

In this chapter we use the 2dFGRS survey to measure the dependence of galaxy clustering on luminosity and spectral type, adding an extra dimension to the analysis carried out in chapter 6*. Previously, a pioneering study of bivariate galaxy clustering, in terms of luminosity and morphological type, was carried out using the Stromlo-APM redshift survey (Loveday et al. 1995). To place the analysis presented here in context, the samples that we consider cover a larger volume and, despite being volume-limited (see §7.2.4), typically contain over an order of magnitude more galaxies than those available to Loveday et al. .

In section 7.2, we give further details about the spectral classification and an explanation of how the samples used in the clustering analysis are constructed. The estimation of the redshift space correlation function and its real space counterpart, the projected correlation function, are outlined in section 7.3. A brief overview of the clustering of

---

*When writing up this analysis, chapter 6 had already appeared as a published paper, referenced as Norberg et al. (2001). Hence any mention to Norberg et al. (2001) in this chapter is also a reference to chapter 6.

2dFGRS galaxies in redshift space, selected by luminosity and spectral type, is given in section 7.4. We present the results of the real space clustering analysis in section 7.5 and we conclude in section 7.6.

7.2 The Data

7.2.1 The 2dFGRS sample

Detailed descriptions of the construction of the 2dFGRS and its properties are given in chapter 2. The sample considered in this chapter consists of 180 000 redshifts measured prior to May 2001. We focus our attention on the two large contiguous volumes of the survey, the SGP and NGP regions.

7.2.2 Spectral classification of 2dFGRS galaxies

The spectral properties of 2dFGRS galaxies are quantified using the principal component analysis (PCA) described by Madgwick, Lahav & Taylor (2000) and Madgwick et al. (2001). The first two projections of the PCA are sensitive to the relative strengths of emission and absorption lines in the rest-frame wavelength range 3700rA to 6650rA (e.g. OII, Hα) and, to a lesser degree, the slope of the continuum. The PCA is dependent upon being able to measure the strength of the Hα line. For galaxies with $z \sim 0.15$, this line is contaminated by sky lines, so we limit our analysis to galaxies with $z < 0.15$, following Madgwick et al. (2001).

The 2dFGRS spectra are classified by a single parameter, η, which is a linear combination of first and second principal components. This parameter is very tightly correlated with the equivalent width of Hα in emission line galaxies (Bland-Hawthorn et al. 2001, in preparation). In this chapter, we divide the 2dFGRS sample into two broad, distinct classes: galaxies with spectra for which the PCA returns $\eta < -1.4$, we refer to, for the sake of brevity, as early-type, and galaxies with $\eta > -1.4$, we call late-type. The distribution of η for 2dFGRS spectra displays a shoulder feature at this value (see Fig. 4a of Madgwick et al. (2001)). Madgwick et al. (2001) show that there is a reasonable correspondence between η and morphological type, using high signal-to-noise spectra and photometry taken from Kennicutt (1992a); $\eta \sim -1.4$ approximately delineates the transition between early and late morphological types. We revisit the comparison between classification using spectral type and morphological type in Fig. 7.1, this time using 2dFGRS spectra and APM photometry. The horizontal axis shows the morphological type

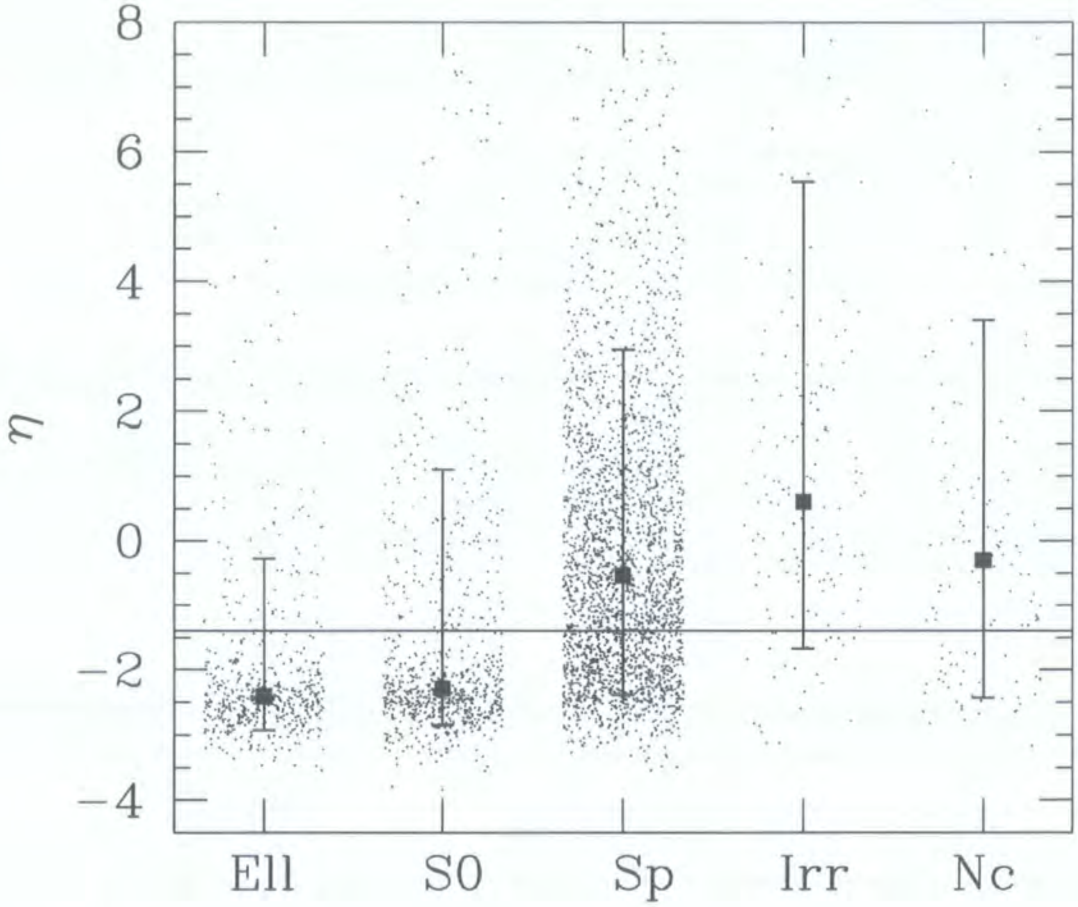


Figure 7.1: A comparison between the morphological classification of bright ($b_J < 17.0$) APM galaxies by Loveday (1996) with the 2dFGRS spectral classification, as quantified by the continuous variable η (see text and Madgwick et al. 2001 for a definition). The morphological classification distinguishes between elliptical (Ell), lenticular (S0), spiral (Sp) and irregular (Irr) galaxies. All galaxies with both a morphological classification and a spectral classification are plotted. The Non-classified (Nc) class incorporates objects for which morphological classification was attempted but for which Loveday was unable to assign a morphological type. The points show the median value of η for each morphological class defined by Loveday, and the error bars show the 10-90 percentile range.

assigned to a subset of bright APM galaxies by Loveday (1996), using images from APM plate scans. The "Nc" class denotes images for which a morphological type could not be assigned. Although there is a substantial amount of scatter in the η values calculated for spectra that lie within a given morphological class, it is reassuring to see that the median η does correlate with morphological class. Moreover, the median η values do match up well with the broad division that we employ to separate early and late types; i.e. galaxies denoted "early type" on the basis of their morphology have a median η that is smaller than our fiducial value of $\eta = -1.4$ and vice-versa for late types. Finally, the total sample of galaxies analysed in this chapter is a subset of the sample used in chapter 6, because of the upper redshift cut applied to the data before the PCA analysis is undertaken.

7.2.3 Sample selection

In order to construct an optimal sample for the measurement of the two point correlation function, we select regions with a high completeness in terms of measured redshifts, using the redshift completeness mask described in section 2.4.2. Indeed, we cannot use the traditional redshift incompleteness mask, as outlined in section 2.3.3, as we need to take into account, in the present analysis, of the additional constraint of the success rate with which spectral types have been assigned to galaxies, which is dependent upon the signal-to-noise ratio of the galaxy spectrum.

In Fig. 7.2, we show histograms of the spectral classification success rate for two different ranges of field completeness, whose definition is given in §2.3.3. The spectral classification success rate has two contributions. The first of these is the redshift incompleteness, shown by the dotted curve. This incompleteness arises because we do not always succeed to measure a redshift for targeted galaxies. The redshift incompleteness is necessarily small for the high completeness fields contributing to the histograms. The second contribution is a spectral classification incompleteness, which occurs when a galaxy redshift is measured successfully, but the spectrum has a low signal to noise ratio (< 10) and is not considered in the PCA (shown by the dashed line). Note that the galaxies used to construct the dashed curve have $z < 0.15$. The spectral classification success rate is given by the product of these two contributions. Our model for this effect, plotted as the solid curves in each panel of Fig. 7.2, is in good agreement with the success rate realised in the 2dFGRS, shown by the histograms.

Rather than weighting the data to compensate for a spectral classification success rate below 100%, we instead modulate the number of unclustered or random points laid down in each field in the clustering analysis to take into account the varying success rate. We have conducted a number of tests in which we varied the completeness thresholds used, adopted different weighting schemes using samples of higher completeness, and we have also compared our results with those obtained in chapter 6, which samples are not subject to spectral classification incompleteness. The results of these tests confirm that our clustering measurements are robust to changes to the details of the weighting scheme adopted; this is largely due to our practice of restricting the analysis to high completeness fields. Excluding areas below our relatively high sector completeness threshold, we estimate that the effective solid angle used in the SGP region is $\sim 380\ \square^\circ$, and in the NGP $250\ \square^\circ$.

7.2.4 Constructing a volume-limited sample

In this chapter, we analyse, like in chapter 6, a series of volume-limited subsamples drawn from the 2dFGRS. The main advantage of this approach is simplicity; the radial selection function is uniform apart from modulations in the space density of galaxies due to clustering. Therefore the complication of modelling the radial selection function of galaxies in a flux limited survey is avoided. This is particularly appealing for the current analysis, as separate selection functions would be required for each class of spectral types studied, since Madgwick et al. (2001) have demonstrated that different galaxies with different spectral types have different luminosity functions.

The disadvantage of using volume-limited samples is that a large fraction of galaxies in the flux limited catalogue do not satisfy the selection criteria of the sample. As we pointed it out in §6.2.2, a volume limited sample specified by a range in absolute magnitude has both a lower ($\mathrm{z_{min}}$) and an upper redshift cut ($\mathrm{z_{max}}$), because the flux limited catalogue has, in practice, bright and faint apparent magnitude limits. This profligate use of galaxies was a serious problem for previous generations of redshifts surveys. This is not the case, however, for the 2dFGRS, which contains sufficient galaxies to permit the construction of volume limited samples defined both by luminosity and spectral type. As we demonstrate in section 7.5, the volume-limited samples we analyse are large enough, both in terms of volume and number of galaxies, to give extremely robust clustering measurements.

To construct a volume limited sample, it is necessary to estimate the absolute magnitude of a galaxy at $z = 0$. Therefore assumptions are required regarding how the luminosity of the galaxy changes with wavelength and with redshift, or equivalently with cosmic time. We make use of the class dependent k-corrections derived by Madgwick et

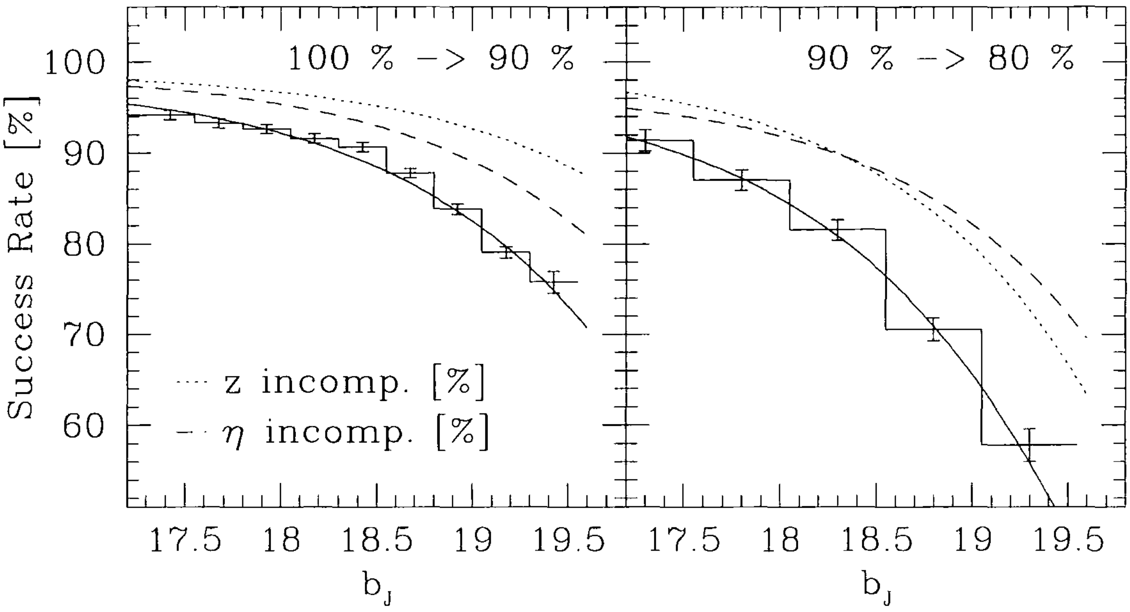
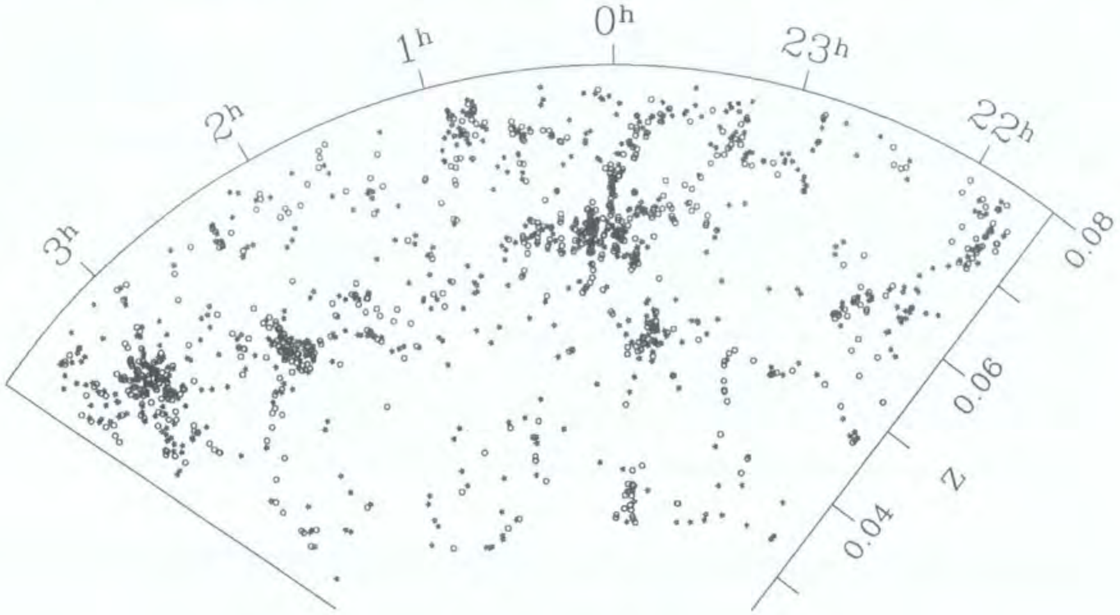
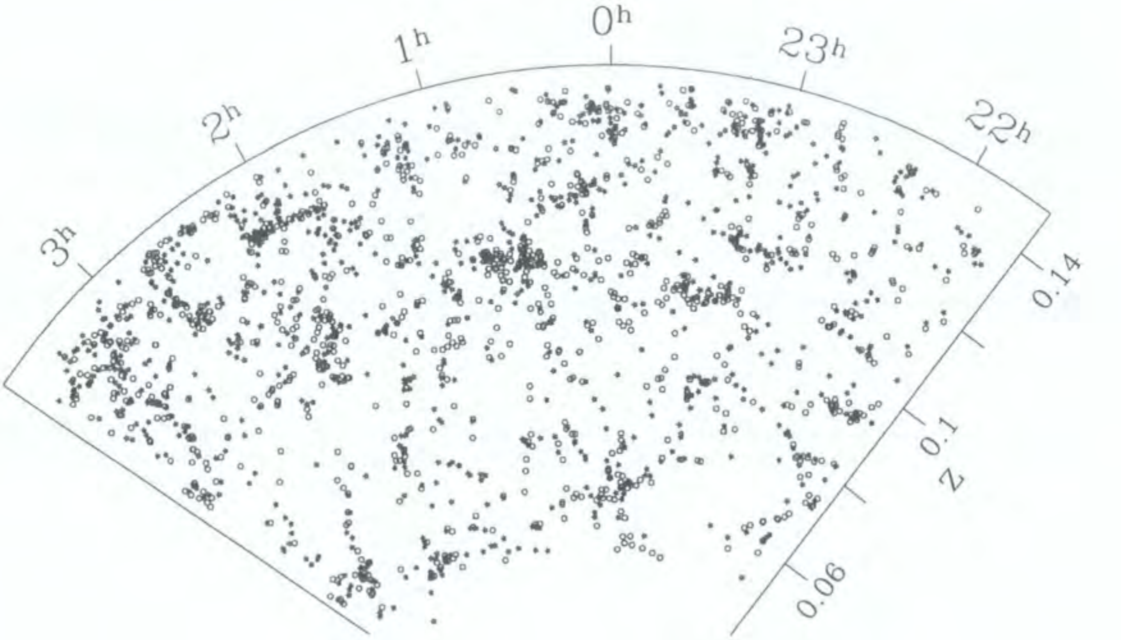


Figure 7.2: The histograms, plotted with Poisson error bars, show the success rate of assigning a spectral type to a galaxy as a function of apparent magnitude. Two field completeness ranges are shown, as indicated by the values listed in the top right of each panel: the left hand panel shows high field completeness regions and the right hand panel shows regions of more moderate field completeness, which definition is given in §2.3.3. The redshift incompleteness, *i.e.* the fraction of targeted galaxies for which a redshift is measured, is shown by the dotted lines. The spectral classification incompleteness is shown by the dashed lines; the dashed lines show the fraction of galaxies with measured redshifts below $z = 0.15$ that have spectra of sufficient signal-to-noise to be used in the PCA. The model for the spectral classification success rate, shown by the solid lines, is the product of the dotted and dashed lines in each panel, and is a good fit to the histogram in each case.



(a) $-18.0 \geq M_{b_J} - 5\log_{10} h \geq -19.0$



(b) $-20.0 \geq M_{b_J} - 5\log_{10} h \geq -21.0$

Figure 7.3: The spatial distribution of 2dFGRS galaxies. The cone plots show the redshift and right ascension of galaxies in a three degree thick strip in declination. Both cones are taken from the SGP. Different absolute magnitude slices are plotted in the two panels; note that the redshift scales are different. Stars mark the locations of late-type galaxies and circles show the position of early-type galaxies.

al. (2001). The mean weighted k-corrections are given by the following expressions:

$$\begin{aligned} k(z) &= 2.6z + 4.3z^2 \text{ (early types)} &(7.1)\\ k(z) &= 1.5z + 2.1z^2 \text{ (late types)} &(7.2)\\ k(z) &= 1.9z + 2.7z^2 \text{ (full sample).} &(7.3) \end{aligned}$$

These k-corrections have the appeal that they are extracted directly in a self-consistent way from 2dFGRS spectra. However, no account is taken of any evolution in the galaxy spectrum. The explicit inclusion of evolution in the spectrum could lead to the ambiguous situation whereby a galaxy's spectral type changes with redshift. We have checked that our results are in fact insensitive to the precise choice of k-correction, comparing clustering results obtained with the spectral type dependent k-corrections given above with those obtained when a global $k + e$-correction (i.e. making an explicit attempt to account for galaxy evolution, albeit in an average sense) is applied, as used for example in chapter 6.

As the $k-$corrections are class dependent, this means that the $z_{\rm min}$ and $z_{\rm max}$ values corresponding to a given absolute magnitude range are also slightly class dependent. Hence, the volumes defined for two different classes with the same range in absolute magnitude will not coincide exactly. In addition to this subtle class dependent definition of the volumes, the values of $z_{\rm min}$ and $z_{\rm max}$ vary slightly with position on the sky. This is due to revisions made to the map of galactic extinction (Schlegel, Finkbeiner & Davis 1998) and to the CCD calibration of APM plate zero-points since the definition of the original input catalogue.

Finally, throughout this chapter, we adopt an $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology to convert redshift into comoving distance.

7.3 Estimating the Two-Point Correlation Function

The galaxy correlation function is estimated on a two dimensional grid of pair separation parallel (π) and perpendicular (σ) to the line of sight. To estimate the mean density of galaxy pairs, a catalogue of randomly positioned points is generated with the same angular distribution and the same values of $z_{\rm min}$ and $z_{\rm max}$ as the data. The correlation function is estimated using

$$\xi_{\rm H} = \frac{DD\,RR}{DR^2}, \tag{7.4}$$

where DD, DR and RR are the number of weighted data-data, data-random and random-random pairs respectively in each bin (Hamilton 1993). This estimator does not require an explicit estimate of the mean galaxy density. We have also cross-checked our results by using the estimator proposed by Landy & Szalay (1993):

$$\xi_{\mathrm{LS}} = \frac{DD - 2DR + RR}{RR}, \tag{7.5}$$

where DD, DR and RR needs to be suitably normalized. We find that the two estimators give equivalent results over the range of pair separations we are interested in.

The clustering pattern of galaxies is distorted when radial positions are inferred from redshifts, as expected in the gravitational instability scenario for structure formation (Kaiser 1987; Cole, Fisher & Weinberg 1994). Clear evidence for this effect is seen in the shape of the two point correlation function when plotted as $\xi(\sigma, \pi)$, as demonstrated for galaxies in chapter 5 (see also Peacock et al. (2001)) and for groups of galaxies by Padilla et al. (2001). After giving a brief flavour of the clustering of 2dFGRS galaxies in redshift space in section 7.5, we focus our attention on clustering in real space in the remainder of the chapter. The clustering signal in real space is inferred by integrating $\xi(\sigma, \pi)$ in the π direction (i.e. along the line of sight):

$$\frac{\Xi(\sigma)}{\sigma} = \frac{1}{\sigma}\int_{-\infty}^{\infty} \xi(\sigma, \pi)\mathrm{d}\pi. \tag{7.6}$$

For the samples that we consider, the integral converges by a pair separation of $\pi \geq 50\,h^{-1}$Mpc. The projected correlation function can then be written as an integral over the spherically averaged real space correlation function, $\xi(r)$,

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma}\int_{\sigma}^{\infty} \xi(r)\frac{r\mathrm{d}r}{(r^2 - \sigma^2)^{1/2}}, \tag{7.7}$$

(Davis & Peebles 1983). As we demonstrate in section 7.4.2, the projected correlation function is well described by a power law. As a working definition, we choose to set the characteristic clustering length of the power law, σ_1, as the scale for which $\Xi(\sigma)/\sigma = 10$:

$$\frac{\Xi(\sigma)}{\sigma} = 10\left(\frac{\sigma_1}{\sigma}\right)^{\gamma}. \tag{7.8}$$

If we assume that the real space correlation function is also a power law (which is a fair approximation for APM galaxies out to separations around $r \sim 10\,h^{-1}$ Mpc, see e.g. Baugh 1996), then Eq. 7.7 can be written as

$$\frac{\Xi(\sigma)}{\sigma} = \left(\frac{r_0}{\sigma}\right)^{\gamma}\frac{\Gamma(1/2)\Gamma([\gamma-1]/2)}{\Gamma(\gamma/2)} = \left(\frac{r_0}{\sigma}\right)^{\gamma}A(\gamma), \tag{7.9}$$

where $\xi(r) = (r_0/r)^{\gamma}$, r_0 is the real space correlation length and γ is equal to the slope of the projected correlation function $\Xi(\sigma)/\sigma$.

We study a range of samples containing different numbers of galaxies and covering different volumes of the Universe. It is imperative to include the effects of sampling fluctuations when estimating the errors on the measured correlation function, to allow a meaningful comparison of the results obtained from different samples. This contribution to the errors has often been neglected in previous work. Following the method used in chapter 6, we employ a sample of 22 mock 2dFGRS catalogues drawn from the ΛCDM Hubble Volume simulation to estimate the error bars on the measured correlation functions. The construction of these mock catalogues is explained in Baugh et al. (2001, in preparation; see also Cole et al. 1998). The catalogues have the same clustering amplitude as measured for galaxies in the flux limited 2dFGRS. The selection criteria applied to the data are used in the construction of the mock surveys. We have experimented with ensembles of catalogues constructed to have different clustering strengths to ascertain the most appropriate way to assign error bars in the case where the clustering amplitude in a data sample is different from that in the mocks. The error bars are reproduced most closely by using the fractional *rms* error obtained from the ensemble of 22 mocks.

7.4 Clustering in Redshift Space

In this section we give a brief overview of the redshift space clustering of 2dFGRS galaxies split by luminosity and spectral type. First, in §7.4.1, we give a qualitative impression of the clustering differences by plotting the spatial distribution of galaxies in volume limited samples. Then we quantify these differences by measuring the spherically averaged correlation function, $\xi(s)$. As mentioned already in chapter 6, we will not make any comprehensive analysis of the clustering of 2dFGRS galaxies in redshift space in this Thesis. Clustering results in redshift space will be presented by Hawkins et al. (2001, in preparation).

7.4.1 Spatial distribution of 2dFGRS galaxies

It is instructive to gain a visual impression of the spatial distribution of 2dFGRS galaxies before interpreting the measured correlation functions. In Fig. 7.3, we show the spatial distribution of galaxies in two ranges of absolute magnitude: in panel (a) we show a sample of faint galaxies ($-18.0 \geq M_{b_J} - 5\log_{10} h \geq -19.0$) and in panel (b) we show bright galaxies ($-20.0 \geq M_{b_J} - 5\log_{10} h \geq -21.0$). Within each panel, early and late type galaxies, as distinguished by their spectral types, are plotted with different symbols;

the positions of early types are indicated by circles and the late types are marked by stars. For clarity, we show only a three degree declination slice cut from the SGP region and we have sparse sampled galaxies, so that the space densities of the two spectral type classes are comparable. We have also restricted the range of redshift plotted, taking a subset of the full volume limited sample in each case. (Note that the redshift ranges differ between the two panels.)

A hierarchy of structures is readily apparent in these plots, ranging from isolated objects, to groups of a handful of galaxies and on through to rich clusters containing over a hundred members. It is revealing to see how structures are traced by galaxies in the different luminosity bins by comparing common structures between the two panels. For example, the prominent structure (possibly a supercluster of galaxies) seen at $\alpha \simeq 0^h$ and $z \simeq 0.061$ is clearly visible in both panels. The same is true for the overdensity seen at $\alpha \simeq 03^h15'$ at $z \simeq 0.068$.

This is the first time that a large enough survey has been available, both in terms of the volume spanned and the number of measured redshifts, to allow a comparison of the clustering of galaxies of different spectral types in representative volume limited samples, without the complication of the strong radial gradient in number density seen in flux limited samples.

It is apparent from a comparison of the distribution of the different spectral types in Fig. 7.3(a), that the faint early type galaxies tend to be grouped into structures on small scales whereas the faint late types are more spread out. One would therefore anticipate that the early types should have a stronger clustering amplitude than the late types, an expectation that is borne out in section 7.4.2.

In the lower panel of Fig. 7.3, the distinction between the distribution of the spectral types is less apparent. This is partly due to projection effects being more important in the declination direction, as the cone extends to a greater redshift than in Fig. 7.3(a). However, close examination of the largest structures suggests that early types are more abundant than late types, again implying a stronger clustering amplitude.

7.4.2 $\xi(s)$ as function of luminosity and spectral type

In Fig. 7.4, we show the spherically averaged redshift space correlation function $\xi(s)$, as a function of luminosity and spectral type. Results are shown for samples selected in one magnitude wide bins, as indicated by the legend in each panel. The top panel shows the correlation functions for all galaxies that have been assigned a spectral type,

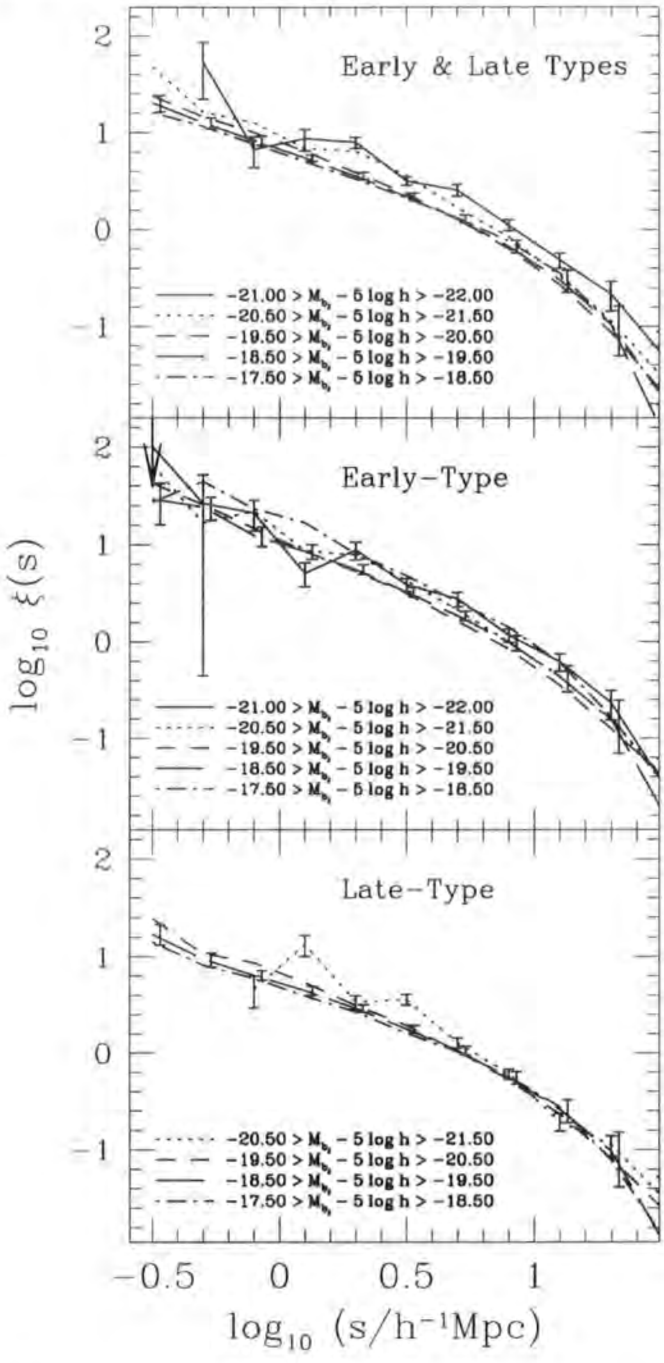


Figure 7.4: The spherically averaged redshift space correlation function of galaxies in disjoint absolute magnitude bins, as indicated by the key in each panel. The panels show the results for different samples: the top panel shows the correlation functions for all galaxies that have been assigned a spectral type, the middle panel show the clustering of galaxies with $\eta < -1.4$ and the lower panel shows $\xi(s)$ for galaxies with $\eta > -1.4$. The error bars are obtained using 2dFGRS mock catalogues, as described in the text. For clarity, error bars are only plotted on the $-18.5 \geq M_{b_J} - 5\log_{10} h \geq -19.5$ sample curve and for the brightest sample in each panel.

the middle panel shows the results for galaxies denoted as early-types ($\eta < -1.4$) and the bottom panel shows the results for late-types ($\eta > -1.4$). Note that, at present, there are insufficient numbers of late-type galaxies to permit a reliable measurement of the correlation function for the magnitude bin $-21.0 \geq M_{b_J} - 5\log_{10} h \geq -22.0$.

Several deductions can be drawn immediately from Fig. 7.4. In all cases, the redshift space correlation function can be described by a power law only over a fairly limited range of scales. The correlation functions of early type galaxies are steeper than those measured for late-types. The early-type galaxies also have a stronger clustering amplitude than late type galaxies. The correlation length, read off as the pair separation for which $\xi(s_0) = 1$, varies for early-types from $s_0 = 7.1h^{-1}$Mpc for galaxies with absolute magnitudes around $M_{b_J} - 5\log_{10} \sim -19.5$ to $s_0 = 8.9h^{-1}$Mpc for the brightest sample with $-21.0 \geq M_{b_J} - 5\log_{10} h \geq -22.0$. The faintest early-types, with magnitudes $-17.5 \geq M_{b_J} - 5\log_{10} h \geq -18.5$, also appear to display a clustering amplitude that is similar to that measured for the brightest early-types. However, the measurement of the correlation function for this faint sample is relatively noisy, as the volume in which galaxies are selected is small compared with the volumes used for brighter samples. The late-type galaxies, show, in contrast, little change in clustering amplitude with increasing luminosity, with a redshift space correlation length of $s_0 = 5.6h^{-1}$Mpc. Only a slight steepening of the redshift space correlation function is apparent with increasing luminosity, until the brightest sample, which displays a modest increase in the redshift space correlation length.

7.5 Clustering in Real Space

7.5.1 Robustness of clustering results

The approach adopted to study the dependence of galaxy clustering on luminosity relies upon being able to compare correlation functions measured in different volumes. It is important to ensure that there are no systematic effects, such as significant sampling fluctuations, that could undermine such an analysis. We demonstrated the robustness of this approach in two ways in chapter 6. First, we constructed a volume limited sample defined using a broad magnitude range, so that the sample could be divided into co-spatial subsamples of galaxies in different luminosity bins, i.e. subsamples within the same volume and therefore subject to the same large-scale structure fluctuations. A clear increase in clustering amplitude was found for the brightest galaxies in the volume, establishing the dependence of clustering on galaxy luminosity (see Fig. 6.1(a)). Second, we demonstrated

Table 7.1: Properties of the combined NGP & SGP volume-limited samples for the full sample of galaxies that have been assigned a spectral type. The first column gives the absolute magnitude range used to define the volume limited sample, the second column gives the median absolute magnitude, $M_{\rm med}$, estimated directly from the data and the third column gives the number of galaxies in the sample. Columns 4 and 5 give the redshift limits of the sample, for a faint limit of b_J=19.45 and a bright limit of b_J=15.0. Columns 6 and 7 list the best fitting correlation length, r_0, and power-law slope, γ, of the real space correlation function, fitted to the projected correlation function, $\Xi(\sigma)/\sigma$, over the range $2.0 \leq \sigma/(h^{-1}\,{\rm Mpc}) \leq 15$. Column 8 gives the value of $A(\gamma)$, defined by Eq. 7.9, evaluated for the best fitting value of γ. The last column lists the value of σ_1, as defined by Eq. 7.8. This table can be compared directly with Table 6.1.

M range $M_{b_J} - 5\log_{10} h$	$M_{\rm med}$	N_{gal}	$z_{\rm min}$	$z_{\rm max}$	r_0 (h^{-1} Mpc)	γ	$A(\gamma)$	σ_1 (h^{-1} Mpc)
-17.5 -18.5	-17.98	8510	0.0164	0.0724	5.19 ± 0.79	1.68 ± 0.12	4.11	3.04 ± 0.46
-18.0 -19.0	-18.46	13795	0.0204	0.0886	4.36 ± 0.76	1.83 ± 0.10	3.57	2.47 ± 0.43
-18.5 -19.5	-18.93	19207	0.0255	0.1077	4.65 ± 0.51	1.80 ± 0.08	3.67	2.67 ± 0.30
-19.0 -20.0	-19.40	24675	0.0317	0.1302	4.93 ± 0.53	1.79 ± 0.10	3.69	2.81 ± 0.31
-19.5 -20.5	-19.85	22555	0.0394	0.1500	4.89 ± 0.33	1.79 ± 0.05	3.69	2.79 ± 0.19
-20.0 -21.0	-20.30	10399	0.0487	0.1500	5.37 ± 0.51	1.78 ± 0.11	3.71	3.04 ± 0.29
-20.5 -21.5	-20.74	3423	0.0602	0.1500	6.57 ± 0.89	1.83 ± 0.23	3.56	3.73 ± 0.51
-21.0 -22.0	-21.19	751	0.0739	0.1500	8.47 ± 2.16	1.80 ± 0.29	3.69	4.87 ± 1.24

Table 7.2: Properties of the combined NGP & SGP volume-limited samples of early-type galaxy samples. The columns have the same definitions as used in Table 7.1.

Mag. range $M_{b_J} - 5\log_{10} h$	$M_{\rm med}$	N_{gal}	$z_{\rm min}$	$z_{\rm max}$	r_0 (h^{-1} Mpc)	γ	$A(\gamma)$	σ_1 (h^{-1} Mpc)
-17.5 -18.5	-18.05	1909	0.0163	0.0707	8.33 ± 1.57	1.87 ± 0.23	3.46	4.72 ± 0.89
-18.0 -19.0	-18.53	3717	0.0203	0.0887	6.28 ± 1.15	1.98 ± 0.11	3.18	3.52 ± 0.64
-18.5 -19.5	-18.98	6405	0.0253	0.1041	5.92 ± 0.84	1.83 ± 0.10	3.57	3.36 ± 0.48
-19.0 -20.0	-19.44	10135	0.0314	0.1249	5.71 ± 0.57	1.87 ± 0.09	3.46	3.23 ± 0.32
-19.5 -20.5	-19.89	11346	0.0388	0.1486	5.66 ± 0.46	1.87 ± 0.09	3.46	3.21 ± 0.26
-20.0 -21.0	-20.33	6434	0.0480	0.1500	6.10 ± 0.57	1.80 ± 0.12	3.69	3.50 ± 0.33
-20.5 -21.5	-20.77	2587	0.0590	0.1500	7.60 ± 1.21	1.87 ± 0.26	3.46	4.31 ± 0.69
-21.0 -22.0	-21.21	686	0.0722	0.1500	9.74 ± 2.93	1.95 ± 0.37	3.27	5.50 ± 1.65

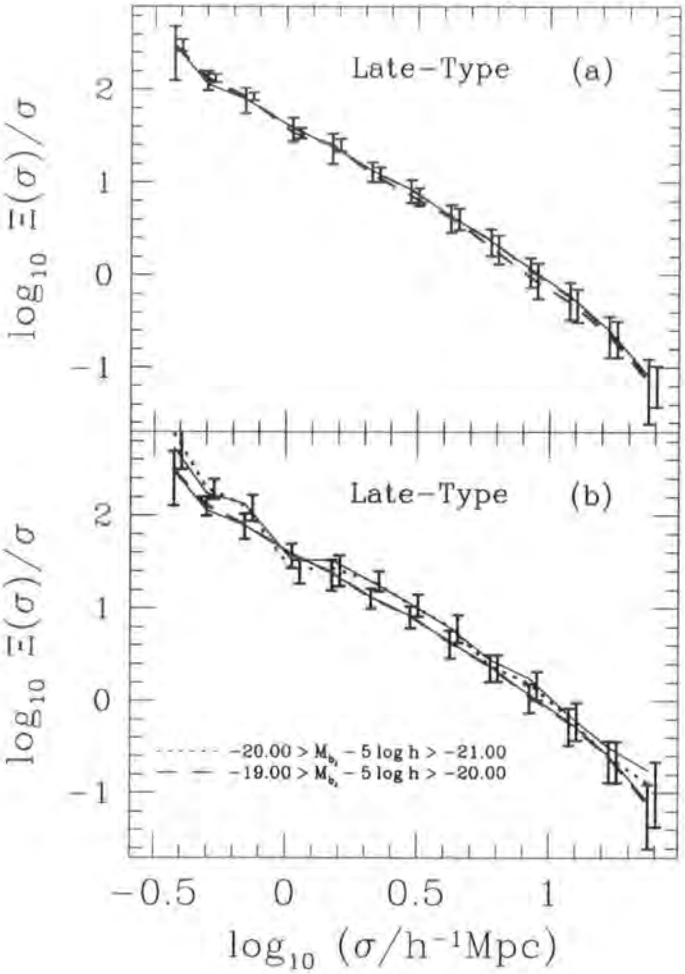


Figure 7.5: (a) The projected correlation function of late-type galaxies in a fixed absolute magnitude bin taken from different, almost independent volumes. We show the correlation function of galaxies with $-19.0 \geq M_{b_J} - 5\log_{10} h \geq -20.0$ taken from volumes defined by $-18.0 \geq M_{b_J} - 5\log_{10} h \geq -20.0$ and $-19.0 \geq M_{b_J} - 5\log_{10} h \geq -21.0$ (both shown by heavy dashed lines). The thin solid line shows the estimate from the optimal sample for the $-19.0 \geq M_{b_J} - 5\log_{10} h \geq -20.0$ magnitude bin. The different measurements are in almost perfect agreement. (b) The projected correlation function measured for late-type galaxies in two *different* absolute magnitude bins but taken from the *same* volume. The volume is defined by the magnitude range $-19.0 \geq M_{b_J} - 5\log_{10} h \geq -21.0$. Within a fixed volume, there is clear evidence for an increase (albeit small) in the clustering amplitude with luminosity. The two thin solid lines show estimates obtained from the corresponding optimal samples for the stated magnitude bins. In both panels the error bars come from the analysis of mock 2dFGRS catalogues.

Table 7.3: Properties of the combined NGP & SGP volume-limited samples of late-type galaxy samples. The columns have the same definitions as used in Table 7.1. Note that the brightest sample listed contains too few galaxies to permit a reliable measurement of the projected correlation function.

Mag. range $M_{b_J} - 5\log_{10}h$	$\mathrm{M}_{\mathrm{med}}$	N_{gal}	$z_{\min}$	$z_{\max}$	r_0 (h^{-1} Mpc)	γ	$A(\gamma)$	σ_1 (h^{-1} Mpc)
-17.5 -18.5	-17.96	6674	0.0164	0.0734	4.27 ± 0.64	1.65 ± 0.12	4.28	2.55 ± 0.38
-18.0 -19.0	-18.44	9992	0.0205	0.0901	3.71 ± 0.69	1.76 ± 0.11	3.81	2.14 ± 0.40
-18.5 -19.5	-18.90	12619	0.0256	0.1099	4.17 ± 0.55	1.79 ± 0.10	3.69	2.38 ± 0.31
-19.0 -20.0	-19.37	14420	0.0319	0.1333	4.45 ± 0.50	1.76 ± 0.09	3.81	2.57 ± 0.29
-19.5 -20.5	-19.82	11122	0.0397	0.1500	4.59 ± 0.28	1.76 ± 0.07	3.81	2.65 ± 0.16
-20.0 -21.0	-20.26	4300	0.0492	0.1500	5.52 ± 0.92	1.87 ± 0.13	3.46	3.13 ± 0.52
-20.5 -21.5	-20.71	1118	0.0608	0.1500	6.33 ± 1.01	2.01 ± 0.29	3.10	3.53 ± 0.56
-21.0 -22.0	-21.17	198	0.0749	0.1500	–	–	–	–

that measuring the correlation function of galaxies in a fixed luminosity bin, but using samples taken from different volumes, gave consistent results (see Fig. 6.1(b)).

In this section, we repeat these tests. The motivation for this exercise is that the samples considered in the present analysis contain fewer galaxies than the full survey, as spectra are only subject to the PCA for galaxies with $z < 0.15$, and because the samples are in this case more dilute as they have been selected on the basis of spectral type as well as luminosity. In Fig. 7.5(a) we plot the projected correlation function of late-type galaxies in a fixed absolute magnitude bin ($-19.0 \geq M_{b_J} - 5\log_{10}h \geq -20.0$), but measured for samples taken from volumes defined by different $\mathrm{z}_{\min}$ and $\mathrm{z}_{\max}$ values. The clustering results are in excellent agreement with one another. In Fig. 7.5(b), we compare the projected correlation function of late-type galaxies in different absolute magnitude ranges taken from the same volume. A clear difference in the clustering amplitude is seen. We have also performed these tests for early-type galaxies and arrive at similar conclusions.

As an additional test, we also show in Fig. 7.5 the correlation function measured in what we refer to as the *optimal sample* for a given magnitude bin. The optimal sample contains the maximum number of galaxies for the stated magnitude bin, because the magnitude bin also defines the redshift range of the volume limited sample. The correlation functions of galaxies in optimal samples are shown by thin solid lines in both panels and are in excellent agreement with the other measurements shown.

7.5.2 Projected correlation function

Fig. 7.6 shows how the real space clustering of galaxies of different spectral type depends on luminosity. We use the optimal sample for each magnitude bin, i.e. the volume-limited sample with the maximum possible number of galaxies, the properties of which are listed in Tables 7.1 (early & late types together), 7.2 (early-types only) and 7.3 (late-types only).

The top panel of Fig. 7.6 confirms the results found in chapter 6, namely that the clustering strength of the full sample increases gradually with increasing luminosity for galaxies fainter than $M^\star$, before showing a clear, strong increase for galaxies brighter than $M^\star$, with $M^\star$ taken to be $M_{b_J} - 5\log_{10} h \simeq -19.7$, following results obtained in chapter 3 and by Folkes et al. (1999). Furthermore, the projected correlation functions are well described by a power law with a slope that is independent of luminosity. The middle panel of Fig. 7.6 shows the projected correlation function of early-type galaxies for different absolute magnitude ranges. The clustering amplitude displays a non-monotonic behaviour, with the faintest sample having almost the same clustering strength as the brightest sample considered. The significance of this result for the faintest galaxies will be discussed further in the next section. Early-type galaxies with $M_{b_J} - 5\log_{10} h \simeq -19.5$, display weaker clustering than the faint and bright samples. The bottom panel of Fig. 7.6 shows the real space clustering of late-type galaxies as function of luminosity. In this case, the trend of clustering strength with luminosity is much simpler. There is a weak increase in clustering amplitude with luminosity. There also is some evidence that the projected correlation function measured for the brightest sub-set of late-types is steeper than that found for the other late-type samples. In general, in the luminosity ranges for which a comparison can be made, the clustering strength of early-type galaxies is always stronger than that of late-types.

The comparison of the correlation functions measured for different samples is made simpler if we divide the curves plotted in Fig. 7.6 by a fiducial correlation function. The reference sample that we choose to provide this fiducial projected correlation function is the sample of all galaxies that have been assigned a spectral type, with absolute magnitudes in the range $-19.5 \geq M_{b_J} - 5\log_{10} h \geq -20.5$. In Fig. 7.7, we plot the ratio of the correlation functions shown in the panels of Fig. 7.6, to the reference correlation function defined above. The trends reported above for the variation of clustering strength with luminosity and spectral type are now clearly visible in Fig. 7.7, particularly the differ-

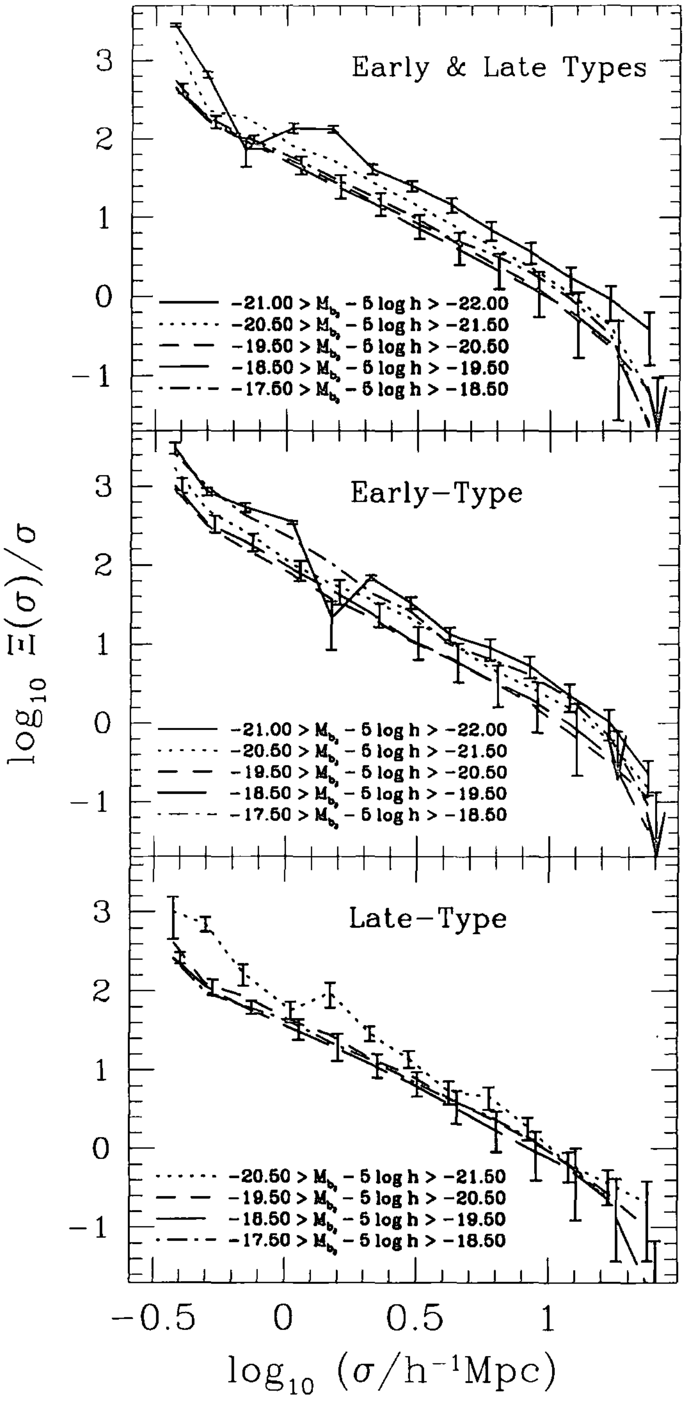


Figure 7.6: The projected galaxy correlation function for samples defined according to luminosity and spectral type. The top panel shows the correlation function measured for all galaxies with a spectral type. The middle panel shows correlation functions for early-types and the bottom panel shows the results for late-types. The absolute magnitude ranges of the samples are indicated in the legend on each panel. The error bars are derived from the 2dFGRS mock catalogues and, for clarity, are only plotted on the correlation function of the $-18.5 \geq M_{b_J} - 5\log_{10} h \geq -19.5$ sample and for the brightest sample in each panel.

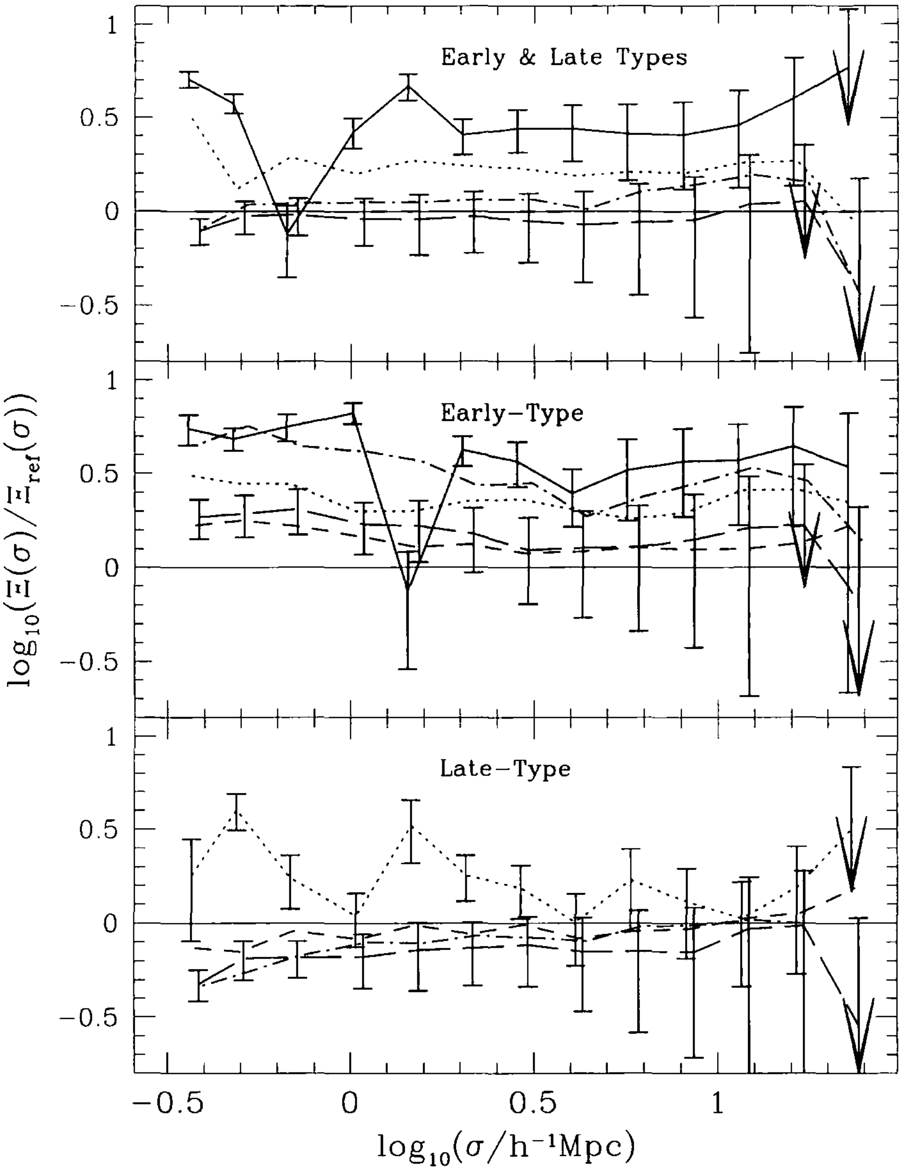


Figure 7.7: The ratio of the projected correlation function measured for galaxies selected by luminosity and spectral type to the projected correlation function of a reference sample. The reference sample consists of all galaxies with an assigned spectral type that lie within the magnitude range $-19.5 \geq M_{b_J} - 5\log_{10} h \geq -20.5$. The top panel shows the ratios for different luminosity bins for all galaxies with a spectral type, the middle panel shows the ratios obtained for early-types and the bottom panel shows the results for late-types. The same lines styles plotted in Fig. 7.6 are used to indicate results for different luminosities. The error bars are from the mock 2dFGRS catalogues, and for clarity, are only plotted on two curves in each panel: the ratio for the sample with $-18.5 \geq M_{b_J} - 5\log_{10} h \geq -19.5$ and the ratio for brightest sample in each panel.

ence in clustering amplitude between early-types and late-types. In the upper and lower panels, the ratios of correlation functions are essentially constant, indicating that a single power slope is a good description over the range of scales plotted. The one exception is the brightest sample of late-type galaxies, which show some evidence of a steeper power law. In the middle panel, the ratios for early-type galaxies shows tentative evidence for a slight steepening of the correlation function at small pair separations, $\sigma < 2h^{-1}$Mpc, which is most pronounced for the brightest sample considered.

7.5.3 Real space correlation length

In the previous subsection, we demonstrated that the projected correlation functions measured for galaxies in the 2dFGRS have a power law form with a slope that varies little as the sample selection is changed, particularly for pair separations in the range $2.0 \leq \sigma/(h^{-1}\,\mathrm{Mpc}) \leq 15.0$. To summarize the trends in clustering strength found when varying the spectral type and luminosity of galaxy samples drawn from the 2dFGRS, we fit a power law form, over the above range of scales. The best fitting values of the parameters of the real space correlation function are determined using Eq. 7.9. We follow a similar approach to the one employed in chapter 6. where we used a simple χ^2 minimisation to extract the best fitting values for parameters in the power law model for the real space correlation function: the correlation length, r_0, and the power law slope, γ. As pointed out in chapter 6, a simple χ^2 approach will not, however, give reliable estimates of the errors on the fitted parameters due to the correlation between the estimates at differing pair separations. We therefore use the mock 2dFGRS catalogues to estimate the errors on the fitted parameters in the following manner. The best fitting values of r_0 and γ are found for each mock individually, using the simple χ^2 analysis. The estimated fractional error is then taken to be the *rms* scatter in the fitted parameters over the ensemble of mock catalogues. The best fitting parameters for each sample are listed in Tables 7.1, 7.2 and 7.3. The results for the correlation length are plotted in Fig. 7.8.

The correlation lengths estimated for the full sample with assigned spectral types (shown by the open squares in Fig. 7.8) are in excellent agreement with the results of chapter 6 (shown by the filled circles). The bright samples constructed in chapter 6 are defined by z_{max} values that exceed the limit of $z_{\mathrm{max}} = 0.15$ enforced upon the samples analysed in this chapter by the PCA used to extract spectral types. The bright samples used in this analysis therefore come from smaller volumes and so the error bars are substantially larger. Thus, a cursory inspection of Fig. 7.8 would give the misleading

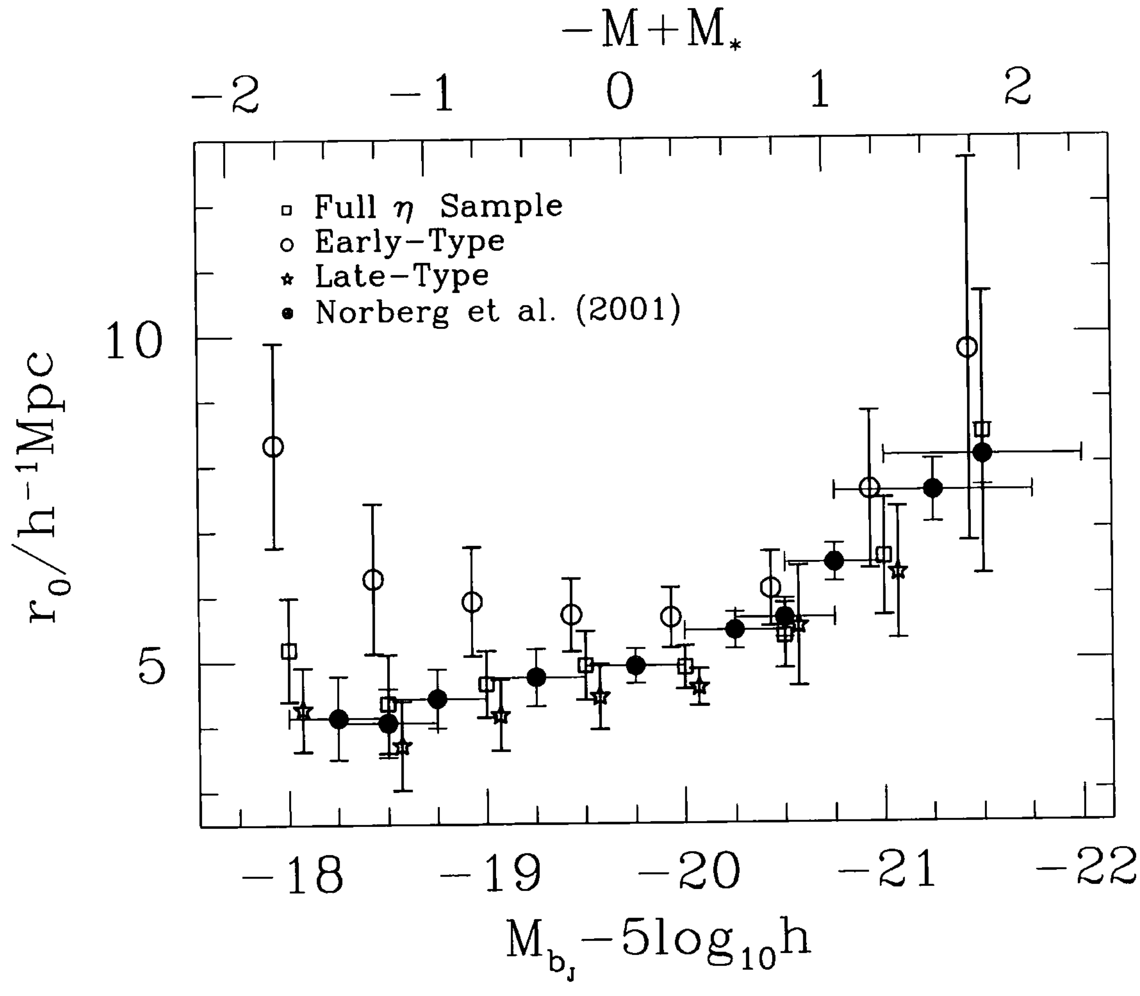


Figure 7.8: The real space correlation length, r_0, as a function of absolute magnitude. The stars show the r_0 values fitted to the projected correlation function of late-type galaxies and the open circles show the best fitting values of r_0 for early-types. The squares show r_0 for the full sample with spectral types. The latter results are in remarkably good agreement with those obtained in chapter 6, which are plotted as filled circles. The horizontal bars on the filled circles show the magnitude range used to define the volume limited samples.

impression that we find much weaker evidence for an increase in correlation length with luminosity. It is important to examine this plot in conjunction with Table 7.2, which reveals that there is significant overlap in the volumes defined by the four brightest magnitude slices, due to a common $z_{\rm max}$ limit. In this case, the error bars inferred from the mocks are pessimistic, as they incorporate cosmic variance, i.e. the variance in clustering signal expected when sampling a fixed volume placed at different, independent locations in the Universe. The volumes containing the four brightest samples listed in Table 7.2 contain long-wavelength fluctuations in common and so are not subject to the same degree of cosmic variance. We have confirmed this hypothesis by examining the scatter between correlation length estimates extracted from two samples in the mock catalogues that have a substantial overlap in volume, and averaging over the ensemble of mock catalogues. We find that in the case where there is significant volume in common, the errors on r_0 can be up to five times smaller than those quoted in the tables.

There is a suggestion of a non-montonic dependence of the correlation length on luminosity for early-type galaxies. However, the error bars on the faintest points should be taken literally here, as these samples are drawn from volumes with little overlap and moreover, the volumes are relatively small compared to those corresponding to brighter galaxies. The projected correlation function of early-type galaxies brighter than $M^\star$ is well fitted by a power law real space correlation function, with a virtually constant slope of $\gamma \sim 1.87$ and a correlation length which increases with luminosity from $r_0 = 5.7 \pm 0.5$ for $M^\star$ galaxies to $r_0 = 9.7 \pm 2.9$ for brighter galaxies ($M_{b_J} - 5\log_{10} h \simeq -21.2$). This represents an increase in the clustering signal by a factor of 2.7, as seen in Fig. 7.7.

The projected correlation functions of late-type galaxies are also consistent with a power law correlation function in real space, with an essentially constant slope. There is a very weak trend of γ increasing with luminosity for late-type galaxies, though at little more than the 1σ level. Ignoring this effect, the fitted slope of the late-type correlation function is $\gamma \sim 1.76$. The correlation length increases with luminosity from a value of $r_0 = 3.7 \pm 0.7$ for faint galaxies ($M_{b_J} - 5\log_{10} h \simeq -18.4$) to $r_0 = 6.3 \pm 1.0$ for bright galaxies ($M_{b_J} - 5\log_{10} h \simeq -20.7$), a factor of 2.5 increase in clustering strength. It should be possible to extend the analysis for late-type galaxies beyond $M_{b_J} - 5\log_{10} h \simeq -21$ when the 2dFGRS is complete.

Fig. 7.8 confirms our earlier conclusion that the clustering of early-type galaxies is stronger than that of late-type galaxies. At $M^\star$, early-types galaxies typically have a real space clustering amplitude that is $1.5 - 1.7$ times greater than that of late-types.

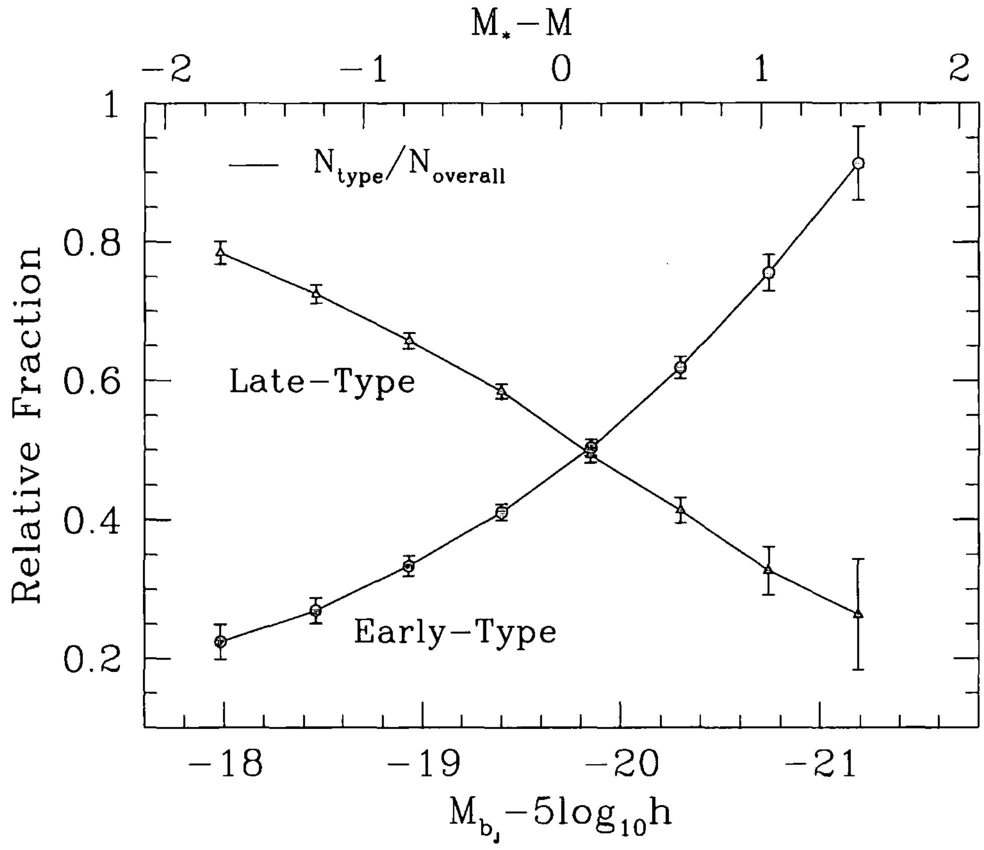


Figure 7.9: The fraction of galaxies in the two broad spectral classes, early-type and late-type, as a function of absolute magnitude. The fractions are derived from the volume limited samples listed in Tables 1, 2 and 3. Note that the fractions need not add up exactly to unity, as for a given magnitude bin, the volumes corresponding to each spectral class do not coincide exactly. The error bars show the Poisson errors on the fractions.

7.6 Discussion

In chapter 6, we used the 2dFGRS to make a precision measurement of the dependence of galaxy clustering on luminosity. The correlation length of galaxies brighter than $M^\star$ was found to increase more rapidly with absolute magnitude than was the case for galaxies fainter than $M^\star$. One of the aims of the present analysis is to understand what phenomena drive this change in clustering strength. In particular, there are two distinct hypotheses that we wish to test. The first of these is that there is a general trend of clustering strength increasing with luminosity, regardless of any other galaxy property. The second hypothesis is that different types of galaxy have different clustering strengths, which, in turn, vary relatively little with luminosity. In this case, the overall sample would show a change in clustering amplitude with luminosity if the mix of galaxy types changed with

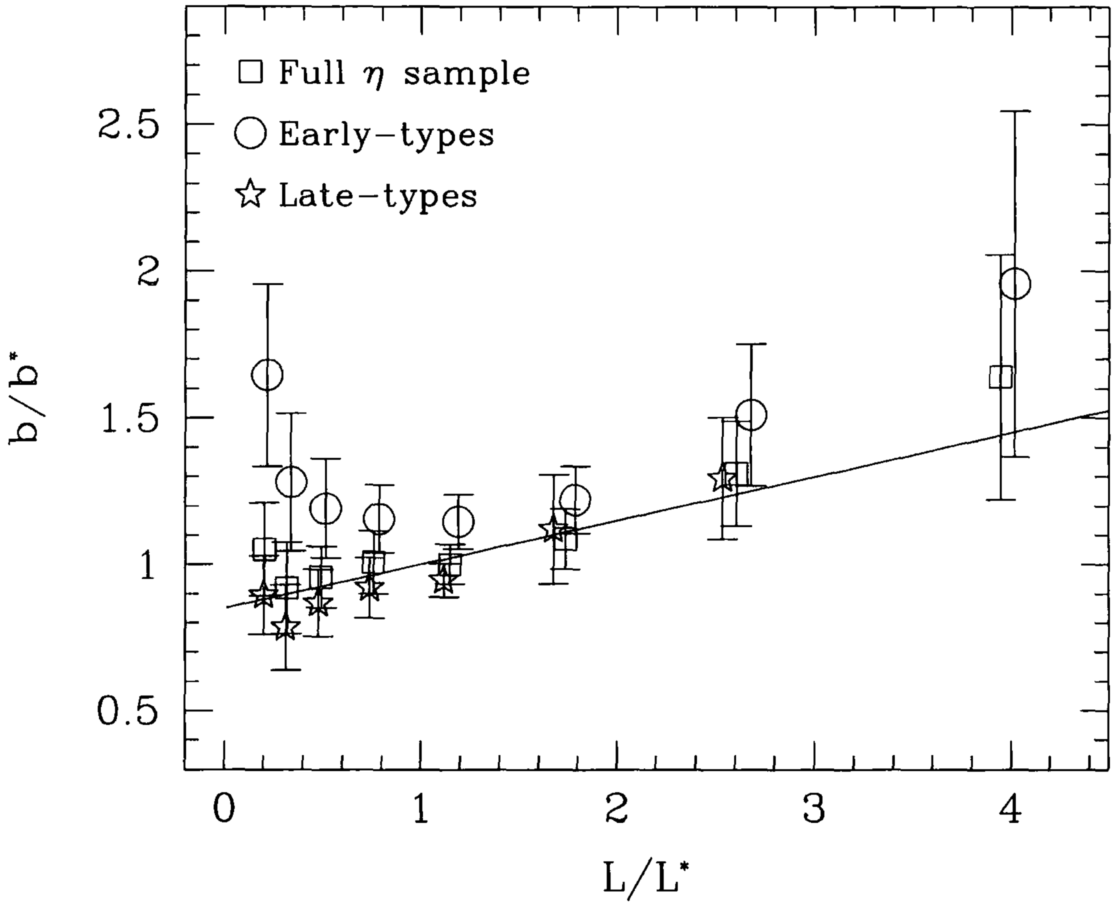


Figure 7.10: The relative bias of the different spectral classes as a function of luminosity, as indicated by the key. The definition of relative bias is given in the text. The reference sample is all galaxies that have been assigned spectral types with absolute magnitudes in the range $-19.5 \geq M_{b_J} - 5\log_{10} h \geq -20.5$. The fiducial luminosity, $L^\star$, is set to $M_{b_J} - 5\log_{10} h = -19.7$. The solid line shows the fit to the results of chapter 6, given by $b/b^\star = 0.85 + 0.15L/L^\star$.

luminosity.

Madgwick et al. (2001) estimated the luminosity function of 2dFGRS galaxies for different spectral classes, and found that the slope of the luminosity function gets steeper and that the characteristic magnitude gets fainter as one moves from early-type to late-types. Another representation of the variation of the luminosity function with spectral class is shown in Fig. 7.9, where we plot the fraction of early and late type galaxies in absolute magnitude bins. The plotted fractions are derived from the volume limited samples listed in Tables 7.1, 7.2 and 7.3; since the volumes for different spectral classes do not coincide exactly, the fractions do not always add up exactly to unity. The mix of spectral types changes dramatically with luminosity; faint samples are dominated by late-types, whereas early-types are the most common galaxies in bright samples. Similar trends were found for galaxies labelled by morphological type in the SSRS2 survey by Marzke et al. (1998).

However, the change in the mix of spectral types with luminosity is not the main effect that drives the increase in the clustering strength of the full sample with luminosity. The results presented in section 7.5 show that the clustering signals for both early and late type galaxies increase with luminosity. Another illustration of this trend is presented in Fig. 7.10, in which we show how clustering varies with luminosity relative to the clustering strength of $M^\star$ galaxies. The fiducial sample is defined as all galaxies with a spectral type that lie in the magnitude range $-19.5 \geq M_{b_J} - 5\log_{10} h \geq -20.5$. For a galaxy sample with best fitting correlation function parameters r_0^i and γ^i, the relative bias is defined by

$$\frac{b^i}{b^\star}(r) = \sqrt{\frac{(r_0^i)^{\gamma_i}}{r_0^{\gamma}} r^{\gamma-\gamma_i}}, \tag{7.10}$$

where r_0 and γ are the best fitting power law parameters for the fiducial sample. The relative bias is computed at a fixed scale, $r = 4.89h^{-1}$Mpc, which is the correlation length of the reference sample given in Table 7.1. The scale dependence in Eq. 7.10 arises if the slopes of the real space correlation functions are different for the galaxy samples being compared; in practice, the term $r^{\gamma-\gamma_i}$ is close to unity for the correlation functions considered. Note that in Fig. 7.10, the ratio $L/L^\star$ is specified relative to $L^\star$ for the full sample, which we take to be set by $M_{b_J} - 5\log_{10} h = -19.7$ following results of chapter 3 and Folkes et al. (1999). For the different spectral classes, the trend of relative bias with luminosity is the same, although there is an offset between the bias factors inferred for early and late types. However, the change in the relative bias with increasing luminosity over the range plotted is much greater than the offset between the relations for the different spectral classes at any given luminosity. The solid line shows the effective bias relation fitted to the results obtained in chapter 6.

We used the 2dFGRS to study the dependence of galaxy clustering on spectral type over a range of a factor of twenty in luminosity. The only previous attempt at a bivariate luminosity-morphology/spectral type analysis of galaxy clustering was performed using the Stromlo-APM redshift survey by Loveday et al. (1995). These authors were able to probe only a relatively narrow range in luminosity around $L^\star$, which is more readily apparent if one considers the median magnitude of each of their magnitude bins (see Fig. 6.3(b)). The clustering strengths measured for the faintest samples considered by Loveday et al. are somewhat weaker than we find, although their results are based on small samples. The scatter between spectral and morphological types shown in Fig. 7.1 precludes a more detailed comparison of the results in this chapter with those obtained in earlier studies.

Chapter 8

Concluding Remarks

In this last chapter, we give a brief overview of the results obtained with the 2 degree Field Galaxy Redshift Survey. We begin with a summary of the contributions made to 2dFGRS Team projects that have not been presented in this Thesis. We then outline the main results obtained within the framework of this Thesis. This chapter ends with an outline of two 2dFGRS based projects that we have already started to undertake.

8.1 Overview of Contributions to 2dFGRS Team Projects

In this Thesis, we have mentioned, on several occasions, that the 2dFGRS project is a large collaboration between several universities in the United Kingdom and Australia*. As a member of such a large collaboration, contributions have often been made to other 2dFGRS projects. In this section, we present these contributions and, at the same time, summarize the results of these projects.

8.1.1 Linear power spectrum and matter content of the Universe

The power spectrum of the galaxy distribution has been determined from the survey using a direct FFT-based technique (Percival et al. 2001). Over wavenumber range $0.02< k <0.15\,h\,\mathrm{Mpc}^{-1}$, the shape of the observed power spectrum is close to that of the linear density perturbations convolved with the window function of the survey. Fitting convolved model power spectra to the 2dFGRS results constrains the shape parameter $\Gamma = \Omega h$ to be 0.20 ± 0.03. This analysis shows that models containing baryons are preferred over models without baryons at the 95% confidence level. This is the first detection of the effect of baryons in the galaxy distribution, and yields an estimate for the baryon fraction $\Omega_b/\Omega = 0.15 \pm 0.07$, assuming scale-invariant primordial fluctuations. This result agrees well with the constraint from clusters of galaxies (White et al. 1993b).

Our contribution to this work was to cross-check the influence on the results of the choice of model used for the selection function. By using different models of the selection

---

*With time, collaborators move around the world with the result that the collaboration now also includes universities in the U.S.A.

function (see chapters 2 and 3), we concluded that the precise choice of model had little influence on the final result, so long as the modelling properly takes into account the shape of the selection function for the SGP and NGP regions beyond the median redshift of the survey, and the analysis only uses survey data observed in 'good conditions' (see §2.3.3 for a precise definition).

8.1.2 Galaxy luminosity function by spectral type

Folkes et al. (1999) and Madgwick et al. (2001) determined the galaxy luminosity function, both overall and as a function of spectral type. Principal Component Analysis was applied to the 2dFGRS spectra, and a linear combination of the first two principal components, η, was been used to classify the spectral type. Going from early-types to late-types, the luminosity functions appear to exhibit a systematic decrease in the characteristic luminosity (from $M^\star_{\mathrm{b_J}} - 5\log h = -19.6$ to -19.0) and a steepening of the faint-end slope (from $\alpha = -0.52$ to -1.43). However there is also evidence that, at the precision afforded by the 2dFGRS sample, the standard Schechter fitting function, when applied to a specific class of galaxies, is no longer an adequate representation of the galaxy luminosity function over the full range of absolute magnitudes probed.

We developed galaxy luminosity function estimators (see chapters 3 and 4), and used these to make independent checks of the results presented in Madgwick et al. (2001). We also refined the survey completeness mask to incorporate incompleteness of spectral type assignment.

8.1.3 Bivariate brightness distribution

The large 2dFGRS sample also allows the generalisation of the luminosity function into the bivariate brightness distribution (BBD) over luminosity and surface brightness (Cross et al. 2001). The BBD derived from the 2dFGRS shows a strong surface brightness–luminosity relation, $M_{b_J} \propto (\,2.4 \pm^{1.5}_{0.5}\,)\mu_e$. The luminosity density is dominated by normal giant galaxies and the peak of the BBD lies away from the survey selection boundaries, implying that the 2dFGRS is complete and that luminous low surface brightness galaxies are genuinely rare over the magnitude range probed.

At the time Cross et al. (2001) carried out their analysis, we were investigating in detail the survey selection function and found discrepant results between the two main strips. As more external checks of the photometry of the SGP region had been carried out at that time, our recommendation was to restrict any analysis only to SGP fields observed in 'good conditions'. Later on with the photometric recalibration of the input catalogue (§2.1.2) , we showed that the photometry of the SGP region was barely modified by the recalibration, whereas the photometry of the NGP region changed more substantially (§2.3.1). Finally, we used the same method as Cross et al. (2001) for estimating the BBD of the combined 2dFGRS-2MASS catalogue in order to confirm that this catalogue does not suffer from surface brightness effects.

8.2 Summary of This Work

We present, in this section, a summary of the work outlined in this Thesis, by emphasizing the context and content of the results rather than repeating the various numerical values we have determined accurately in earlier chapters.

8.2.1 Completeness analysis of the 2dFGRS

First of all, we must point out that it would not have been possible to obtain any of the results presented in this Thesis without the major investment of effort made to implement the statistical tools presented in chapter 2. Whether they map the imperfections of the photometric input catalogue (magnitude limit mask and stellar contamination mask) or the incompleteness of the redshift catalogue (redshift completeness mask and μ-mask), these masks have been heavily used in all the analyses, together with the weighting schemes also outlined in chapter 2. The detailed understanding of the survey involved in implementing these tools has allowed us to make optimal use of the 2dFGRS. The analyses presented in this Thesis represent only a small fraction of the science that will ultimately be extracted from the survey.

8.2.2 Survey selection function and luminosity function

In chapter 3, we studied the galaxy $\mathrm{b_J}$-band luminosity function in great detail. We estimate the 2dFGRS photometric catalogue to be ~88% complete, after a close comparison with deeper and more homogeneous CCD photometry released by the Sloan Digital Sky Survey. Taking into account uncertainties in the photometric calibration, in the adopted $k+e$-correction and in the normalization, we obtained an estimate of the $\mathrm{b_J}$-band galaxy luminosity function which is well fitted by a Schechter function, and which is in very good agreement with the estimate of Blanton et al. (2001), as long as one properly corrects for colour selection and uses the same normalisation method and $k+e$-correction. With this deep investigation of the b_J-band galaxy luminosity function and thanks to detailed mapping of the survey incompleteness done in chapter 2, we are able to model accurately the survey selection function.

8.2.3 Combining 2dFGRS with the near infra-red 2MASS

In chapter 4, we combined the 2dFGRS with the 2MASS extended source catalogue to produce an infrared-selected sample of over 17 000 galaxies with redshifts. After a detailed description of the completeness of the matched 2dFGRS-2MASS catalogue, we used this sample to determine the J and K_S-band galaxy luminosity functions, which are fairly well fit by Schechter functions. From the distributions of $b_J -$ K and $J - K_S$ colours with absolute magnitude and models of the stellar populations, the galaxy stellar-mass function was estimated. Integrated over all galaxy masses, it yields a total mass fraction in stars, representing 0.1% to 0.3% of the critical density, depending on which initial mass function is favoured. These values are consistent with estimates based on the time integral of the observed star formation history of the universe only if dust extinction corrections at high redshift are modest.

8.2.4 Quantifying redshift space distortions

In chapter 5, we provided the first clear detection of the redshift-space clustering anisotropy on large scales that is a key prediction of the gravitational instability paradigm for the growth of structure in the universe. Measurements of this distortion yield a precise estimate of the flattening parameter, β, which depends on the total mass density of the universe, Ω, and the linear bias parameter of the optical galaxies with respect to the mass. Combined with recent measurements of the anisotropies in the cosmic microwave background (Jaffe et al. 2001), this result favours a low-density universe, with $\Omega \approx 0.3$.

8.2.5 Dependence of clustering on luminosity and spectral type

The size of the 2dFGRS sample has allowed investigation of the variation in the strength of galaxy clustering with luminosity (chapter 6) and as a bivariate function of luminosity and spectral type (chapter 7), using the projected two-point correlation function of galaxies in a series of volume-limited samples. The clustering of galaxies in real space is well fitted by a power-law relation over the range $2.0 < (\sigma / h^{-1}\,\mathrm{Mpc}) < 15$.

The exponent of the power-law shows little variation in galaxy samples with luminosities differing by a factor of 40. When the samples are split by spectral type into star-forming galaxies (late-types) and quiescent galaxies (early-types), we find that early-type galaxies have a steeper slope than late-type galaxies. There is also some evidence for a weak increase of the slope for the brightest late-type galaxies.

The clustering amplitude, r_0, shows a clear dependence on luminosity and spectral type. For an average galaxy sample, r_0 increases slowly for galaxies fainter than $M^\star$, but more strongly at brighter absolute magnitudes. This dependence of the correlation length on luminosity is in good agreement with the predictions of the hierarchical galaxy formation models of Benson et al. (2001). These clustering results are well characterised by a linear relation between the relative luminosity, $L/L^\star$, and the relative bias parameter, $b/b^\star$, that gives the amplitude of the correlation function relative to that of $L^\star$galaxies. Split into early and late-type galaxy samples, we find that early-type galaxies are more strongly clustered than late-types, but, more importantly, both classes show a clustering strength that increases with luminosity at approximatively the same rate. This is the first time we have clear evidence that the clustering strength of an average galaxy sample is driven by the intrinsic luminosity of the sample.

8.3 What Next?

Until this point, we have reviewed some of the work done during the past three years. However, with a survey the size of the 2dFGRS, the number of interesting projects is much larger than this. Moreover, the statistical tools developed in chapter 2 have not yet been used to their full extent. As the survey and these tools were devised with some specific projects not yet undertaken in mind, it would be an omission not to mention some of them in this Thesis. Therefore we outline in this last section two very different projects on which we have already started work and which will be the focus of our research with the 2dFGRS data for the coming year.

8.3.1 Groups and clusters of galaxies

The mass function of groups and clusters and the influence of environment on the galaxy population are two important constraints on theories of galaxy formation (Benson et al. 2000b). Moreover, Frenk et al. (1990) and White et al. (1993a) showed that the distribution of cluster dispersions is an important constraint on the amplitude of primordial density fluctuations. Therefore one of the main scientific objectives of the 2dFGRS is to construct the definitive catalogue of groups and clusters of galaxies. This catalogue will be obtained through an extension of the techniques developed by Moore et al. (1993) and those recently discussed by Ramella et al. (1997) and Nolthenius et al. (1997). As the survey will always contain a certain degree of incompleteness, it is necessary to make extensive tests using mock galaxy catalogues which have been constructed to mimic the real survey (Cole et al. 1998) before applying the full methodology to the real data. Moreover, other tests like finding groups in the underlying mass distribution in numerical simulations and seeing how well they are traced by galaxy groups need be carried out. The final group catalogue can, for example, be used to analyze redshift space galaxy clustering (via cluster collapse), to look at clustering of groups and clusters, to do further measurements of the dependence of galaxy luminosity on environment, and to calibrate group finders that operate in 2-D.

8.3.2 Satellite galaxies

The implication that spiral galaxies are imbedded in massive dark haloes, based on the interpretation of their flat rotation curves in the 1970s, provided one of the cornerstones of our present understanding of galaxy formation (eg. Freeman 1970, White & Rees 1978): dark halos of typical galaxies extend far beyond their optical radius and have a mass of at least $10^{12}M_{\odot}$. Such halos are a generic prediction of hierarchical clustering theories of galaxy formation. However, much of this theoretical prejudice is lacking in observational support. There are currently two major ways of probing the mass distribution around isolated galaxies: gravitational lensing (eg. Sheldon et al. 2001) or distribution of satellite galaxies (eg. Zaritsky et al. 1997a). The 2dFGRS is a suitable survey for the second approach, namely detecting satellite galaxies around bright primaries.

The methodology proposed for selecting isolated galaxies is similar to the one described in Zaritsky et al. (1993). Based on preliminary work, we expect to detect, once the 2dFGRS is completed, typically 2500 satellite galaxies, which is $\sim$ 10 times more than any previous compilation (Lorrimer et al. 1994, Zaritsky et al. 1997b). With this sample of satellite galaxies we will be able, in a statistical way, to probe the extent of dark halos around isolated galaxies. A vital issue in this analysis is the understanding of the selection function of the survey, which, if not fully understood, can strongly bias our findings.

The 2dFGRS has already changed our views of the galaxy distribution. Forthcoming works promises to enhance further our understanding of the large scale structure.

Appendix A

Connection between Galaxy Redshift and Cosmology

The aim of this appendix is to give a brief explanation of what a redshift is, and how a galaxy redshift is connected to the expansion of the Universe*. First we introduce the Cosmological Principle and the notion of metric. Then we explain the connection between cosmology and redshift, for finally ending up with a short section on the expansion of the Universe. This appendix is nevertheless very brief. For a nice, enjoyable and more formal introduction to general relativity and cosmology, I can only highly recommend the reader to have a closer look at Weinberg (1972).

A.1 Cosmological Principle and Robertson-Walker Metric

A.1.1 Cosmological Principle

Following observational evidences, cosmologists have adopted a so called 'Cosmological Principle', stating that the Universe is on large scales, to a good approximation, homogeneous and isotropic. This principle is equivalent to saying that there is no preferred direction nor position in the Universe.

A.1.2 Robertson-Walker Metric

From special relativity, we know that time and space coordinates are closely related to each other via a metric, η. It is usually written like this

$$d\tau^2 = -\eta_{\alpha\beta} d\xi^\alpha d\xi^\beta \, , \tag{A.1}$$

---

*This appendix was requested by my examiners after my total failure to explain what a galaxy redshift was... This is a worrying thought, knowing that the aim of this whole thesis is to make best use of the 2 degree Field *Galaxy Redshift* Survey...

with the proper time $d\tau$, the coordinate system ξ^α and the convention of summation on repeated suffixes. By using the equivalence principle, one 'easily' extends special relativity to general relativity, also called general theory of gravitation. In general relativity, time and space coordinates are still closely related to each other via a metric.

Using a comoving coordinate system, it can be shown that the most general space-time metric describing a Universe in which the Cosmological Principle is obeyed is of the form:

$$d\tau^2 = dt^2 - R(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right] , \tag{A.2}$$

where we have used spherical comoving coordinates: the radial coordinate r, the solid angle Ω, and the cosmic standard time t. $R(t)$ is an unknown function of time and k a constant, which by suitable choice of units for r can be chosen to have the value +1, 0, or -1. This metric is know as the Robertson-Walker metric.

For $k = +1$, the proper volume of space is proportional to $R(t)^3$, and hence sets the scale of the geometry of space. As a generalization valid for all values of the curvature parameter k, $R(t)$ can be called the *cosmic scale factor*, but also *expansion parameter* as explained in more detail below.

A.2 Connection between redshift and cosmic scale factor

We explain below how information about the cosmic scale factor $R(t)$ can be obtained through the observation of shifts in wavelength of light emitted by distant sources.

Let us consider the case where a pulse is emitted at time t_1 from a typical galaxy, located at r_1, Ω_1. Given that light follows a null geodesic (ie. $d\tau^2 = 0$), this light pulse will reach the observer, placed at the origin of the coordinate system[†], at a time t_0 given by:

$$\int_{t_1}^{t_0} \frac{dt}{R(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - kr^2}} \div f(r_1) \tag{A.3}$$

As the right hand-side of this equation is only a function of the position of the galaxy, which by definition will not change with time t in comoving coordinates, a second light pulse emitted at a time $t_1 + \delta t_1$ will reach the observer at a time $t_0 + \delta t_0$ given by:

$$\int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{R(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - kr^2}} = \int_{t_1}^{t_0} \frac{dt}{R(t)} \tag{A.4}$$

[†] Thanks to the cosmological principle, this handy location of the observer will not affect the final result!

Hence, for a very small interval of time δt_0 and δt_1 and assuming that $R(t)$ varies very little over this interval, we get:

$$\frac{\delta t_0}{R(t_0)} = \frac{\delta t_1}{R(t_1)} \tag{A.5}$$

and

$$\frac{\nu_0}{\nu_1} = \frac{\delta t_1}{\delta t_0} = \frac{R(t_1)}{R(t_0)} \tag{A.6}$$

where we have introduced the observed and emitted frequencies, ν_0 and ν_1 respectively. This is conventionally expressed in terms of redshift parameter z, defined as the fractional increase in wavelength

$$z \doteqdot \frac{\lambda_0 - \lambda_1}{\lambda_1} \tag{A.7}$$

and hence

$$z = \frac{R(t_0)}{R(t_1)} - 1\,, \tag{A.8}$$

with t_1 the time of emission and t_0 the time of reception.

A.3 Expansion of the Universe

Following the definition of redshift given above, several scenarios can be considered, but two simple cases can be distinguished:

- if most galaxies have a positive redshift value, it implies $R(t_0) > R(t_1)$ and hence the Universe is expanding.
- if most galaxies have a negative redshift value, it implies $R(t_1) > R(t_0)$, and hence the Universe is contracting.

Since Hubble (1929a), we know that we are living in an expanding Universe, as all samples of galaxies covering a sufficiently large volume around us are fully dominated by galaxies with positive recession velocities, or equivalently, the spectrum of nearly all galaxies is redshifted.

Appendix B

Luminosity Function Estimators

Among the various luminosity function estimators present in the literature, the three most commonly used are the $1/V_{max}$ estimator (Schmidt 1968), the Sandage-Tammann-Yahil estimator (STY hereafter; Sandage, Tammann & Yahil 1979) and the stepwise maximum likelihood estimator (SWML hereafter; Efstathiou, Ellis & Peterson 1988). We briefly review here the main characteristics and drawbacks for each them in this appendix. Due to the similarities of the STY and SWML estimators we discuss them together.

Before describing any of the estimators, we would like to draw attention to one particular hypothesis assumed with all the current luminosity function estimators; namely that the galaxy sample on which one applies these methods is described by one single luminosity function. If the sample has a bias, then the estimator will obviously produce a biased estimate of the luminosity function. If the sample contains galaxies which are drawn from several completely different luminosity functions, then these estimators will come out with a mean weighted luminosity function, which might describe the true underlying luminosity function particularly well.

B.1 $1/V_{max}$

The $1/V_{max}$ estimator (Schmidt 1968) is the simplest binned, non-parametric luminosity function estimator. It assumes that for a given luminosity L

$$\begin{aligned} \phi(L)dL &= \sum_{i=1}^{N} \omega_i \frac{W(L - L_i)}{V_{max}(L_i)} && \text{(B.1)} \\ W(L - L_i) &= \Theta(L_i - L + dL/2) - \Theta(L + dL/2 - L_i)\,, && \text{(B.2)} \end{aligned}$$

where the sum is taken over all the galaxies in the sample, ω_i is the weight associated with galaxy i, Θ is the Heaviside unit step function and V_{max} is the maximum volume in which a galaxy of luminosity L_i could have been detected in the survey. V_{max} depends explicitly on the apparent magnitude of the galaxy, its associated magnitude limit and its measured redshift. Strictly speaking V_{max} should be written as $V_{max}(m_i, mlim_i, z_i)$.

Felten (1976) showed that although this estimator is not optimal as it is not independent of clustering (ie it can be biased by local inhomogeneities), the $1/V_{max}$ estimator has the attractive feature of recovering both the shape and the amplitude of the luminosity function, which is not the case for the two maximum likelihood estimators discussed in section B.2. However this can also be a major drawback of this estimator as it is very sensitive to density fluctuations, though they should not really affect the luminosity function estimate of the 2dFGRS as the survey is so deep (except maybe at fainter absolute magnitudes).

B.2 Likelihood Estimators

STY and SWML are both maximum likelihood estimators based on the same general likelihood formula (see Eq. B.4), the difference being the assumption made about the shape of the luminosity function: STY assumes a Schechter luminosity function parametrized by α, $M^\star$ and $\Phi^\star$, whereas SWML parametrizes the luminosity function by series of steps. Let us first give a quick mathematical description of these two estimators and afterwards we discuss them.

The probability per unit luminosity interval that a galaxy of luminosity L_i at redshift z_i is included in a unbiased survey is

$$P_i = \left(\frac{\phi(L_i)}{\int_{L_m(z_i)}^{L_M(z_i)} \phi(L) dL} \right)^{\omega_i} \tag{B.3}$$

where $L_m(z_i)$ is the minimum luminosity that can be seen in the survey at redshift z_i and $L_M(z_i)$ is the maximum luminosity that can be seen in the survey at redshift z_i and ω_i the weight associated with that galaxy. As for the $1/V_{max}$ estimator, $L_M(z_i)$ is dependent upon the galaxy's associated magnitude limit and should therefore, strictly speaking, be written like this: $L_M(z_i, m_i^{lim})$. The likelihood function to be maximized is the product of all these probabilities and can therefore be written as a summation (to an additive constant near):

$$Ln[\mathcal{L}] = \sum_{i=1}^{N} \omega_i \left(Ln\left[\phi(L_i)\right] - Ln\left[\int_{L_m(z_i)}^{L_M(z_i)} \phi(L) dL \right] \right) \tag{B.4}$$

For the STY method, where one assumes that the luminosity function is a Schechter function:

$$\phi(L)dL = \phi^\star \left(\frac{L}{L^\star}\right)^{\alpha} \exp\left(-\frac{L}{L^\star}\right) d\left(\frac{L}{L^\star}\right) \tag{B.5}$$

the likelihood function becomes:

$$\begin{aligned}\mathrm{Ln}[\mathcal{L}] &= \alpha\sum_{i=1}^{N}\omega_i\mathrm{Ln}[L_i] + (\alpha+1)\mathrm{Ln}[L^\star]\sum_{i=1}^{N}\omega_i \\ &- \sum_{i=1}^{N}\omega_i\frac{L_i}{L^\star} - \sum_{i=1}^{N}\omega_i\mathrm{Ln}\left[\Gamma\left(\alpha+1,\frac{L_m(z_i)}{L^\star}\right)\right] \\ &+ \sum_{i=1}^{N}\omega_i\mathrm{Ln}\left[\Gamma\left(\alpha+1,\frac{L_M(z_i)}{L^\star}\right)\right] + \tilde{c} \end{aligned} \tag{B.6}$$

where $\Gamma(x, y)$ is the standard incomplete gamma function and $\tilde{c}$ is the unknown constant.

For the SWML method, where one assumes the luminosity function is represented by N_p steps ($k = 1, N_p$):

$$\phi(L) = \phi_k \quad , \qquad L_k - dL/2 \le L < L_k + dL/2\,, \tag{B.7}$$

the likelihood function becomes:

$$\begin{aligned}\mathrm{Ln}[\mathcal{L}] &= \sum_{i=1}^{N}\omega_i\mathrm{Ln}\left[\sum_{k=1}^{N_p}\phi_k W(\tilde{L}_i - L_k)\right] \\ &- \sum_{i=1}^{N}\omega_i\mathrm{Ln}\left[\sum_{k=1}^{N_p}\phi_k\Delta L H[L_k - L_{min}(z_i)]\right] + \tilde{c}\end{aligned} \tag{B.8}$$

where $\tilde{L}_i$ is the luminosity of galaxy i, $W(L - L_i)$ is given by equation B.2, and $H(x)$ is given by:

$$H(x) = \begin{cases} 0 & x \le -\Delta L/2 \\ x/\Delta L + 1/2 & -\Delta L/2 \le x \le \Delta L/2 \\ 1 & x \ge \Delta L/2 \end{cases} \tag{B.9}$$

As the SWML estimator is able to extract from a dataset a luminosity function with more degrees of freedom than the STY method, the SWML method should give a better estimate of the true luminosity function. The advantage of the STY estimator is that it returns a function described by only three parameters (α, $M^\star$ and $\Phi^\star$) and is therefore very simple to compare with results from other surveys. Moreover, as the functional form used for the STY estimator is already a very good estimate of the galaxy luminosity function, it is clear why it is probably still the most used luminosity function estimator.

Bibliography

The following abbreviations are used in this bibliography:

MNRAS: Monthly Notices of the Royal Astronomical Society
ApJ: The Astrophysical Journal
ApJS: The Astrophysical Journal Supplement
AJ: The Astronomical Journal
A&A: Astronomy and Astrophysics
ARA&A: Annual Reviews of Astronomy and Astrophysics

Allen, C.W. 1973, Astrophysical Quantities, London: Athlone Press, University of London. (3rd edition).

Alonso M. V., et al. , 1993, AJ, 106, 676

Arnouts S., et al. , 2001, A&A, 379, 740

Ballinger W.E., Peacock J.A., Heavens A.F., 1996, MNRAS, 282, 877

Baugh, C.M., 1996, MNRAS, 280, 267.

Baugh C. M., Efstathiou G., 1993, MNRAS, 265, 145

Baugh C.M., Efstathiou G., 1994, MNRAS, 267, 323

Baugh, C.M., Cole, S., Frenk, C.S., 1996, MNRAS, 283, 1361.

Baugh, C.M., Cole, S., Frenk, C.S., Lacey, C.G., 1998, ApJ, 498, 504.

Baugh C.M., et al. , 2001, in preparation

Bean A.J., Ellis R.S., Shanks T., Efstathiou G., Peterson B.A., 1983, MNRAS, 205, 605

Beichman C. A., Neugebauer G., Haling H. J., Clegg P. E., Chester T. J., 1988, IRAS Catalogs and Atlases, Vol 1: Explanatory Supplement (ES). JPL, Pasadena

Benoist C., Maurogordato S., da Costa L.N., Cappi A., Schaeffer R., 1996, ApJ, 472, 452

Benson, A.J., Baugh, C.M., Cole, S., Frenk, C.S., Lacey, C.G., 2000a, MNRAS, 316, 107.

Benson A.J., Cole S., Frenk C.S., Baugh C.M., Lacey C.G., 2000b, MNRAS, 311, 793

Benson, A.J., Frenk, C.S., Baugh, C.M., Cole, S., Lacey, C.G., 2001, MNRAS, 327, 1041

Blair M., Gilmore G., 1982, PASP, 94, 741

Blanton, M., 2000, ApJ, 544, 63.

Blanton, M.R., et al. , 2001, AJ, 121, 2358

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Boyle B. J., Shanks T., Croom S. M., Smith R. J., Miller L., Loaring N., Heymans C.,, 2000, MNRAS, 317, 1014

Bromley, B.C., Press, W.H., Lin, H., Kirshner, R.P., 1998, ApJ, 505, 25.

Bruzual, A.G., Charlot, S., 1993, ApJ, 405, 538

Bruzual, A.G., Charlot, S., 2001, in preparation.

Burles, S. & Tytler, D., 1998, ApJ, 507, 732

Burles, S., Nollet, K.M., Truran, J.N., Turner, M.S., 1999, Phys. Rev. Lett. 82, 4176.

Calzetti, D. 1999, 'Workshop on ultraluminous galaxies: Monsters or Babies', Kluwer in press. (astro-ph/9902107)

Carlberg R.G., et al. , 2000, ApJ, 542, 57

Carpenter, J.M., 2001, AJ, 121, 2851.

Cole, S., Fisher, K.B., Weinberg, D.H., 1994, MNRAS, 267, 785.

Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781

Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, MNRAS, 300, 945

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cole S., et al. (the 2dFGRS team), 2001, MNRAS, 326, 255

Colless M, 1999, Phil. Trans. R. Soc. Lond. A, 357, 105)

Colless M, et al. (the 2dFGRS Team), 2001, MNRAS, 328, 1039

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Loaring N., Miller L., Lewis I. J., 2001, MNRAS, 322, 29

Cross N., et al. (2dFGRS Team), 2001, MNRAS, 324, 825.

Cross N., et al. (the 2dFGRS Team), 2001, in preparation

Cross, N., & Driver, S., 2001, in preparation

da Costa L. N., et al. , 1988, ApJ, 327, 544

da Costa L. N., 1999, 'Evolution of large scale structure : from recombination to Garching', ed. A. J. Banday, R. K. Sheth, L. N. da Costa, Garching, Germany, p.87

Davis, M., Geller, M.J., 1976, ApJ, 208, 13.

Davis M., Huchra J., Latham D. W., & Tonry J, 1982a, ApJ, 253, 423

Davis M., Huchra J., 1982b, ApJ, 254, 437

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Davis, M., Meiksin, A., Strauss, M.A., Da Costa, L.N., Yahil, A., 1988, ApJ, 33, L9.

Davis M., & Newman J. A., 2001, ESO Workshop on "Mining the Sky", Munich, August 2000, ESO Astrophysics Symposia (astro-ph/0104418)

Davis M., Peebles, P.J.E., 1983, ApJ, 267, 465

Dekel A., Lahav O., 1999, ApJ, 520, 24

de Lapparent V., Geller M. J., Huchra J. P., 1986, ApJ, 302, 1L

de Vaucouleurs G., 1958, AJ, 63, 253

de Vaucouleurs G., de Vaucouleurs A., 1964, 'Reference catalogue of bright galaxies', University of Texas Monographs in Astronomy, Austin: University of Texas Press

de Vaucouleurs A., de Vaucouleurs G., 1967, AJ, 72, 730

Dressler, A., 1980, ApJ, 236, 351.

Driver S. P., 1999, ApJ, 526, L69

Efstathiou, G., Ellis, R.S., Peterson, B.A., 1988, MNRAS, 232, 431

Efstathiou G., Moody S. J., 2001, MNRAS, 325, 1603

Ellis R. S., Colless M., Broadhurst T., Heyl J., Glazebrook K., 1996, MNRAS, 280, 235

Evrard, A.E., 1999, 'Evolution of large scale structure : from recombination to Garching', ed. A. J. Banday, R. K. Sheth, L. N. da Costa, Garching, Germany, p.249 (astro-ph/9812377)

Evrard, A.E., et al. (the Virgo Collaboration), 2001, ApJ submitted (astro-ph/0110246)

Feldman H.A., Kaiser N., Peacock J.A., 1994, ApJ, 426, 23

Felten, J. E., 1976, ApJ 207, 700

Ferrara, A., Bianchi, S., Cimatti, A., Giovanardi C., 1999 ApJS, 123, 423

Fisher K. B., Huchra J. P., Strauss M. A., Davis M., Yahil A., Schlegel D., 1995, ApJS, 100, 69

Folkes S.J., et al. (the 2dFGRS Team), 1999, MNRAS, 308, 459

Freedman W.L., et al. , 2001, ApJ, 553, 47

Freeman, K., 1970, ApJ, 160, 811

Frenk, C.S., White, S.D.M., Efstathiou, G.P., & Davis, M., 1990, ApJ, 351, 10

Fukugita, M., Ichikawa, T., Gunn, J.E., Doi, M., Shimasaku, K., Schneider, D.P., 1996, AJ, 111, 1754

Fukugita, M., Hogan, C.J., Peebles, P.J.E., 1998, ApJ., 503, 518

Gardner J.P., Sharples, R.M., Carrasco, B.E., Frenk, C.S., 1996, MNRAS, 282, 1

Gardner J.P., Sharples, R.M., Frenk, C.S., Carrasco, B.E., 1997, ApJL, 480, 99

Geller M. J., & Huchra J. P., 1989, Science, 246, 897

Glazebrook, K., Peacock, J.A., Collins, C.A., Miller, L., 1994, MNRAS, 266, 65

Glazebrook, K., Peacock, J.A., Miller, L., Collins, C.A., 1995, MNRAS, 275, 169

Gnedin N., Ostriker J. P., 1992, ApJ., 400, 108, 459

Governato, F., Baugh, C.M., Frenk, C.S., Cole, S., Lacey, C.G., Quinn, T., Stadel, J., 1998, Nature, 392, 359.

Gunn J. E., Gott J. R.,1972, ApJ, 176, 1

Guzzo, L., Strauss, M.A., Fisher, K.B., Giovanelli, R., Haynes, M.P., 1997, ApJ, 489, 37.

Guzzo, L., et al. , 2000, A&A, 355, 1.

Hamilton, A.J.S., 1988, ApJ, 331, L59.

Hamilton A.J.S., 1992, ApJ, 385, L5

Hamilton A.J.S., 1993, ApJ, 417, 19

Hamilton A.J.S., 1998, "The Evolving Universe" ed. D. Hamilton, Kluwer Academic, p. 185-275 (astro-ph/9708102)

Hamilton A.J.S., Tegmark M., Padmanabhan N., 2000, MNRAS, 317, L23

Hasegawa, T., Umemura, M., 1993, MNRAS, 263, 191.

Hatton S., Cole S., 1998, MNRAS, 296, 10

Hawkins, E., Maddox, S.J., Branchini, E., Saunders, W., 2001, MNRAS, 325, 589.

Hermit, S., Santiago, B.X., Lahav, O., Strauss, M.A., Davis, M., Dressler, A., Huchra, J.P., 1996, MNRAS, 283, 709.

Hoyle, F., Baugh, C.M., Shanks, T., Ratcliffe, A., 1999, MNRAS, 309, 659.

Hubble E., 1929a, Mt. Wilson Communications, n^0 105

Hubble E., 1929b, Proceedings of the National Academy of Sciences, 15, 168

Hubble E.P., 1934, ApJ, 79, 8

Hubble E., Humason M. L., 1931, ApJ, 74, 43

Huchra J. P., Vogeley M S., & Geller M. J., 1999, ApJS, 121, 287

Humason M. L, 1931, ApJ, 74, 35

Iovino, A., Giovanelli, R., Haynes, M., Chicarini, G., Guzzo, L., 1993, MNRAS, 265, 21.

Jaffe A., et al. , 2001, Phys.Rev.Lett., 86, 3475 (astro-ph/0007333)

Jarrett, T.H., Chester, T., Cutri, R., Schneider, S., Skrutskie, M., 2000, ApJ, 119, 2498

Jenkins, A., Frenk, C.S., Pearce, F.R., Thomas, P.A., Colberg, J.M., White, S.D.M., Couchman, H.M.P., Peacock, J.A., Efstathiou, G., Nelson, A.H., 1998, ApJ, 499, 20.

Jenkins, A., Frenk, C.S., White, S.D.M., Colberg, J., Cole, S., Evrard, A.E., Couchman, H.M.P., Yoshida, N., 2001, MNRAS, 321, 372.

Jing Y.P., Mo H.J., Börner G., 1998, ApJ, 494, 1

Johnson, H.L., 1966, ARA&A, 4,193

Kaiser, N., 1986, MNRAS, 219, 785

Kaiser, N., 1987, MNRAS, 227, 1

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795

Kennicutt, R.C., 1983, ApJ, 272, 54

Kennicutt, R.C., 1992a, ApJ, 388, 310.

Kennicutt, R.C., 1992b, ApJS, 79, 255

Kirshner R.P., Oemler A., Schechter P.L., Shectman S.A., 1981, ApJ, 248, L57

Kochanek, C.S., et al. 2001, ApJ, 560, 566

Kron, R.G., 1980, ApJS, 43, 305

Lahav, O., Saslaw, W.C., 1992, ApJ, 396, 430.

Landy S.D., Szalay A.S., 1993, ApJ, 412, 64

Lawrence A., et al. , 1999, MNRAS, 308, 897

Lilly S. J., Le Fèvre O., Crampton D, Hammer F., Tresse L., 1995, ApJ, 455, 50

Lin H, et al. , 1996, ApJ, 464, 60

Lorrimer, S.J., Frenk, C.S., Smith, R.M., White, S.D.M., & Zaritsky, D., 1994, MNRAS, 269, 696

Loveday, J., 1996, MNRAS, 278, 1025

Loveday, J., 2000, MNRAS, 312, 517 80, ApJS, 43, 305

Loveday J., Peterson B. A., Efstathiou G., Maddox S. J., 1992, ApJ, 390, 338

Loveday J., Maddox S.J., Efstathiou G., Peterson B.A., 1995, ApJ, 442, 457

Loveday, J., Tresse, L., Maddox, S.J., 1999, MNRAS, 310, 281.

Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990a, MNRAS, 242, 43

Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990b, MNRAS, 243, 692

Maddox S. J., Efstathiou G., Sutherland W. J., 1990c, MNRAS, 246, 433

Maddox, S.J., Sutherland, W.J., Efstathiou, G., Loveday, J., Peterson, B.A., 1990d, MNRAS, 247, 1

Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1996, MNRAS 283, 1227

Maddox S. J., et al. (the 2dFGRS Team), 2001, forever in preparation

Madgwick, D.S., et al. (the 2dFGRS team), 2001, MNRAS, in press (astro-ph/0107197)

Madgwick, D.S., Lahav, O., & Taylor, K., 2000, 'ESO Astrophysics Symposia', Springer-Verlag in press (astro-ph/0010307)

Marzke R. O., Huchra J. P., Geller M. J., 1994, ApJ, 428, 43

Marzke, R.O., da Costa, L.N., Pellegrini, P.S., Willmer, C.N.A., Geller, M.J., 1998, ApJ, 503, 617

Maurogordato, S., Lachieze-Rey, M., 1991, ApJ, 369, 30.

Meiksin A.A., White M., 1999, MNRAS, 308, 1179

Metcalfe N., Fong R., Shanks T., 1995, MNRAS, 274, 769

Mobasher, B., Sharples, R.M., Ellis, R.S., 1993, MNRAS, 263, 560

Moore, B., Frenk, C.S., & White, S.D.M., 1993, MNRAS, 261, 827

Mould J.R., et al. , 2000, ApJ, 529, 786

Nolthenius, R., Klypin, A. A.,& Primack, J. R., 1997, ApJ, 480, 43

Norberg P., et al. (the 2dFGRS Team), 2001, MNRAS, 328, 64 (chapter 6)

Outram P.J., Hoyle F., Shanks T., 2001, MNRAS, 321, 497

Padilla, N.D., Merchan, M.E., Valotto, C.A., Lambas, D.G., Maia, M.A.G., 2001, ApJ, 554, 873.

Peacock, J.A., 1997, MNRAS, 284, 885.

Peacock J.A., Dodds S.J., 1996, MNRAS, 280, L19

Peacock, J.A., Smith, R.E., 2000, MNRAS, 318, 1144

Peacock, J.A., et al. (the 2dFGRS Team), 2001, Nature, 410, 169.

Peebles, P.J.E., 1980, The Large Scale Structure of the Universe, Princeton.

Peebles, P.J.E., 1993, Principles of Physical Cosmology, Princeton.

Pen U.-L., 1998, ApJ, 504, 601

Percival W.J., et al. (the 2dFGRS Team), 2001, MNRAS, 327, 1297

Perlmutter S., et al. , 1999, ApJ 517, 565

Persic, M., Salucci, P., 1992, MNRAS, 258, 14P

Peterson B. A., Ellis R. S., Efstathiou G., Shanks T., Bean A. J., Fong R., Zen-Long Z., 1986, MNRAS, 221, 233

Phillipps, S., Shanks, T., 1987, MNRAS, 229, 621.

Phillipps, S., Davies, J., Disney M. 1990, MNRAS, 242, 235.

Pimbblet K. A., Smail I., Edge A. C., Couch W. J., O'Hely E., Zabludoff A. I., 2001, MNRAS, 327, 588

Ramella, M., Pisani, A., & Geller, M. J., 1997, AJ, 113, 483

Ratcliffe A., et al. , 1996, MNRAS, 281, L47

Renzini, A., Voli, M., 1981, A&A, 94, 175

Riess A. G, et al. , 1998, AJ, 116, 1009

Robinson L., Wampler J., 1977, Astronomical Applications of Image Detectors with Linear Response, IAU Colloq. 40, ed. Duchesne & Lelievre, p.14

Salpeter, E.E., 1955, ApJ. 121, 61

Salucci, P., Persic, M., 1999, MNRAS, 309, 923

Sandage A., Tammann G. A., 1975, ApJ, 196, 313

Sandage, A., Tammann, G.A., Yahil, A., 1979, ApJ, 232, 352

Sandage A., Tammann G. A., 1990, ApJ, 365, 1

Santiago, B.X., Strauss, M.A., 1992, ApJ, 387, 9.

Sargent W. L. W., Turner E. L., 1977, ApJ, 212, L3

Saunders W., et al. , 1991, Nature, 349, 32

Saunders W., et al. , 2000, MNRAS, 317, 55

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Schmidt, M., 1968, ApJ 151, 393

Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1

Shanks, T., 1990, 'Galactic and extragalactic background radiation', Dordrecht, Kluwer Academic Publisher, 269

Shectman S., & Hiltner W. A., 1976, PASP, 88, 960

Shectman S. A., et al. , 1996, ApJ, 470, 172

Sheldon, E.S, et al. , 2001, ApJ, 554, 881

Slipher V. M., 1917, The Observatory, 40, 304

Somerville R., Primack J. R., 1999, MNRAS, 310, 1087

Somerville, R.S., Lemson, G., Sigad, Y., Dekel, A., Kauffmann, G., White, S.D.M., 2001, MNRAS, 320, 289.

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17

Steidel, C.C., Adelberger, K.L., Giavalisco, M., Dickinson, M., Pettini, M., 1999, ApJ, 519, 1

Storrie-Lombardi, L. & Wolfe. A., 2000, ApJ, 543, 552.

Stoughton C., et al. , (The SDSS Collaboration), AJ, 123, 485

Strauss M.A., Willick J.A., 1995, Phys. Reports, 261, 271

Strömberg G., 1925, ApJ, 61, 353

Szalay A.S., Matsubara T., Landy S.D., 1998, ApJ, 498, L1

Szapudi, I., Branchini, E., Frenk, C.S., Maddox, S., Saunders, W., 2000, MNRAS, 318, L45.

Szokoly, G.P., Subbarao, M.U., Connolly, A.J., Mobasher, B., 1998, ApJ, 492, 452

Taylor K., et al. , 1997, SPIE, 2871, 145

Tegmark, M., Bromley, B.C., 1999, ApJ, 518, L69.

Tonry, J., & Davis, M., 1979, AJ, 84, 1511

Verde L., Heavens A.F., Matarrese S., Moscardini L., 1998, MNRAS, 300, 747

Vettolani G. et al. , 1997, A&A, 325, 954

Weinberg, S., 1972, 'Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity', John Wiley & Sons, New York.

White, S.D.M., & Rees, M.J., 1978, MNRAS, 183, 341

White, S.D.M., Davis, M., Efstathiou, G., Frenk, C.S., 1987, Nature, 330, 351.

White, S.D.M., Efstathiou, G.P.,& Frenk, C.S., 1993a, MNRAS, 262, 1023

White, S.D.M., Navarro, J.F., Evrard, A.E., Frenk, C.S., 1993b, Nature, 366, 429

Willmer, C.N.A., Da Costa, L.N., Pellegrini, P.S., 1998, AJ, 115, 869.

Woosley, S.E., Weaver, T.A., 1995, ApJS, 101, 181

Wright, E.L., 2001, ApJ, 556, L17.

Yasuda N., et al. (The SDSS Collaboration), 2001, AJ, 122, 1104

Yee H. K. C., Ellingson E., & Carlberg R. G., 1996, ApJ, 102, 269

Yee H. K. C., et al. , 2000, ApJS, 129, 475

York D., et al. , 2000, AJ, 120, 1579

Zaritsky, D., Smith, R., Frenk, C.S., & White, S.D.M., 1993, ApJ, 405, 464

Zaritsky, D., Smith, R., Frenk, C.S., & White, S.D.M., 1997a, ApJ, 478, L53

Zaritsky, D., Smith, R., Frenk, C.S., & White, S.D.M., 1997b, ApJ, 478, 39

Zehavi, I., et al. (the SDSS Collaboration), 2001, ApJ submitted (astro-ph/0106476)

Zucca E., et al. , 1997, A&A, 326, 477