Large-scale structure in the distant universe and the build-up of the passive population

William G. Hartley

The University of Nottingham

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

Feb 2010

Supervisors:	Dr. Omar Almaini
	Dr. Frazer R. Pearce
Examiners:	Prof. Seb Oliver *(University of Sussex)*
	Prof. Alfonso Aragón-Salamanca *(University of Nottingham)*
Submitted:	28 February 2010
Examined:	tbc
Final version:	tbc

Contents

List of Figures

List of Tables

Abstract

In this thesis I use the UKIDSS Ultra-Deep Survey to investigate the evolution in clustering of the galaxy population across the redshift range $0 < z < 3$. Selected in the K-band, the sample does not suffer as greatly from biases against passive and dusty objects that plague optically-selected samples, and so represents the first analysis of a relatively unbiased galaxy sample above $z = 1$. The unique combination of area (0.77 deg.^{2}) and depth ($K_{AB} < 24$) allows the analysis of both rare, intrinsically bright objects and faint objects, reaching $\sim L^*$ at $z = 3$. The UDS data are complemented by similarly deep optical data from Subaru and mid-infrared data from Spitzer. These data have enabled accurate photometric redshifts to be computed, which are used extensively throughout this thesis.

The combined optical and infrared data are used to firstly select passive and star-forming samples over a wide range in redshift $1 < z < 2.5$. The results of their number counts and clustering properties motivate a more in-depth study of the respective clustering strengths of passive and star-forming galaxies. The main result of this thesis is that passive galaxies are found to be more strongly clustered than star-forming galaxies to at least $z = 1.5$, irrespective of rest-frame K-band luminosity. Furthermore, within either the passive or star-forming sample, luminosity segregation is found to be very mild. These results therefore indicate that even at fixed stellar mass, passive galaxies are more strongly clustered. Clustering strength is directly related to the mass of the typical host dark matter halo. Galaxies which are observed to be passive at $z < 1.5$ are therefore hosted by more massive halos than their star-forming counterparts. This finding is consistent with downsizing in galaxy formation. Downsizing is also found in the clustering strengths of the star-forming sub-samples. The correlation lengths of star-forming galaxies smoothly decline towards $z = 0$. Galaxies of a given K-band luminosity or stellar mass are therefore found in less massive halos in the low-redshift universe. Passive galaxies on the other hand appear to have constant r_0 values over $0 < z < 1.5$. If confirmed, this may indicate a preferred mass of dark matter halo for the transition from star-forming to passive. Above $z = 1.5$ the clustering strengths of passive and star-forming galaxies are found to be more similar. If this trend continues it suggests that they may become equal within $2 < z < 3$. This would therefore constitute the epoch over which the passive/star-forming bimodality becomes distinct.

These results are then examined more closely in a marked-correlation function analysis and tested against the one of the latest theoretical galaxy formation models. Due to difficulties in defining a suitable quantity for 'passivity' from photometric data, the results from the marked correlation functions are largely inconclusive at this stage. However, the potential in their future use is demonstrated. My results highlight two known issues in the semi-analytic galaxy formation model: the overproduction of high-redshift sources and red satellites. In addition I offer tentative evidence for further possible discrepancies, in clustering properties at high-redshift.

Acknowledgements

First and foremost I would like to thank Katrina, not only for her endless support throughout the time I've spent working towards producing this thesis, but because without her I wouldn't have even got as far as starting. I also have a deep debt of gratitude to my supervisors, Omar and Frazer, and also Seb. Their scientific guidance has, of course, been instrumental in producing this work, but I have learnt many other lessons along the way as well. I hope I can someday pass on a similar standard of supervision to others. Everyone at the astronomy and particle theory group at Nottingham, past and present, I'd like to thank for creating such a great atmousphere to work in - and for some memorable moments. To my family and friends, thanks for being interested in my work, or just being there at all. I also wish to offer my apologies, for the six months or so of recent neglect; I'll make it up to you all. Of these, a last special mention. Thanks to LEP for letting me use his house as a hotel in the last few frantic days of writing!

Published work

Much of the work in chapter 2 has been previously presented in a paper (paper I), while much of chapter 3 will feature in a forthcoming paper (paper II).

I Hartley W. G., Lane K. P., Almaini O., Cirasuolo M., Foucaud S., Simpson C., Maddox S., Smail I., Conselice C. J., McLure R. J., Dunlop J. S.,2008, "The clustering and abundance of star-forming and passive galaxies at $z \sim 2$", MNRAS, 391, 1301.

II Hartley W. G., Almaini O., Cirasuolo M., Foucaud S., Simpson C., Conselice C. J., Smail I., McLure R. J., Dunlop J. S., Chuter R. W., Maddox S., Lane K. P., Bradshaw E. J.,2010, "The evolution of galaxy clustering since $z = 3$ using the UKIDSS Ultra Deep Survey: the divergence of passive and star-forming galaxies", MNRAS, *submitted*.

The vast majority of the work presented in this thesis was performed by the author, with advice from the paper coauthors listed above. Where the material presented is the result of more collaborative work, this is mentioned in the relevant chapter.

Chapter 1

Introduction

The study of extragalactic astronomy began with Hubble's discovery that so-called spiral nebulae were in fact systems independent of the galaxy in which we live. He followed up this astounding observation by classifying these galaxies into what is often termed the 'Hubble tuning fork diagram'. This classification system, shown in figure 1.1, not only remains in use today, but many of the questions that resulted from Hubble's work are yet to be answered.

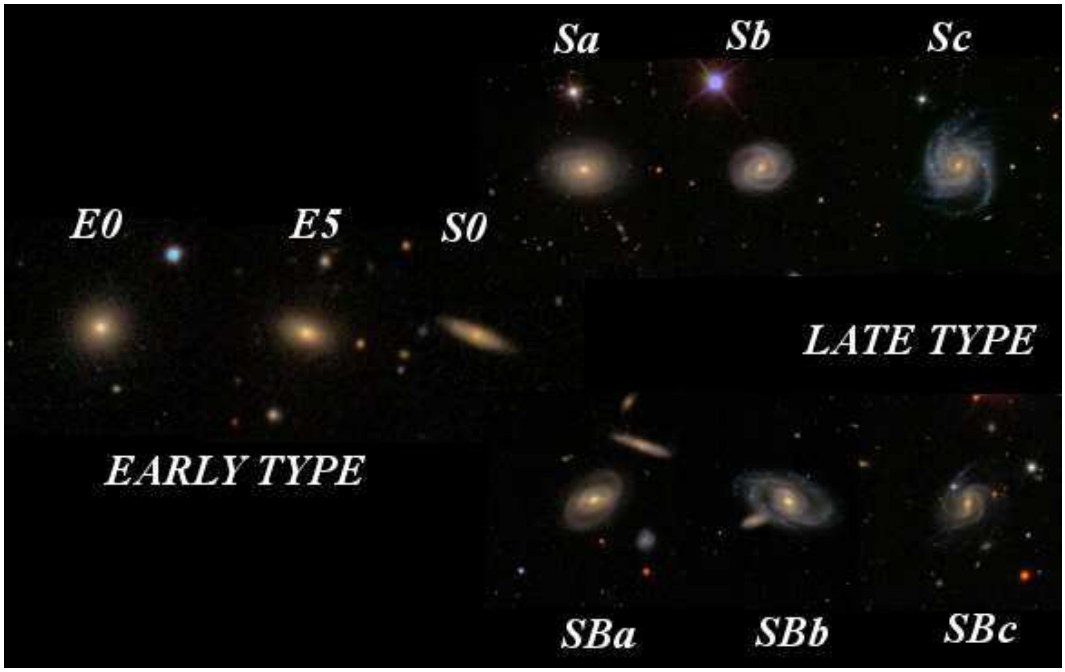

Figure 1.1. Hubble's tuning fork classification diagram, constructed here using galaxies taken from the Sloan digital sky survey (SDSS) and classified by users of Galaxy Zoo Lintott et al. (2008).

The way in which Hubble constructed his classification diagram implied that galaxies evolved from one end to the other. But although he labelled the elliptical and lenticular (S0) end 'early type' galaxies, with the Sb and Sc end being 'late type' galaxies (Hubble 1936), he was careful to stress that there was no evidence to suggest such evolution. We now believe that evolution in this fashion is almost certainly incorrect, though the precise evolutionary path of galaxies remains largely

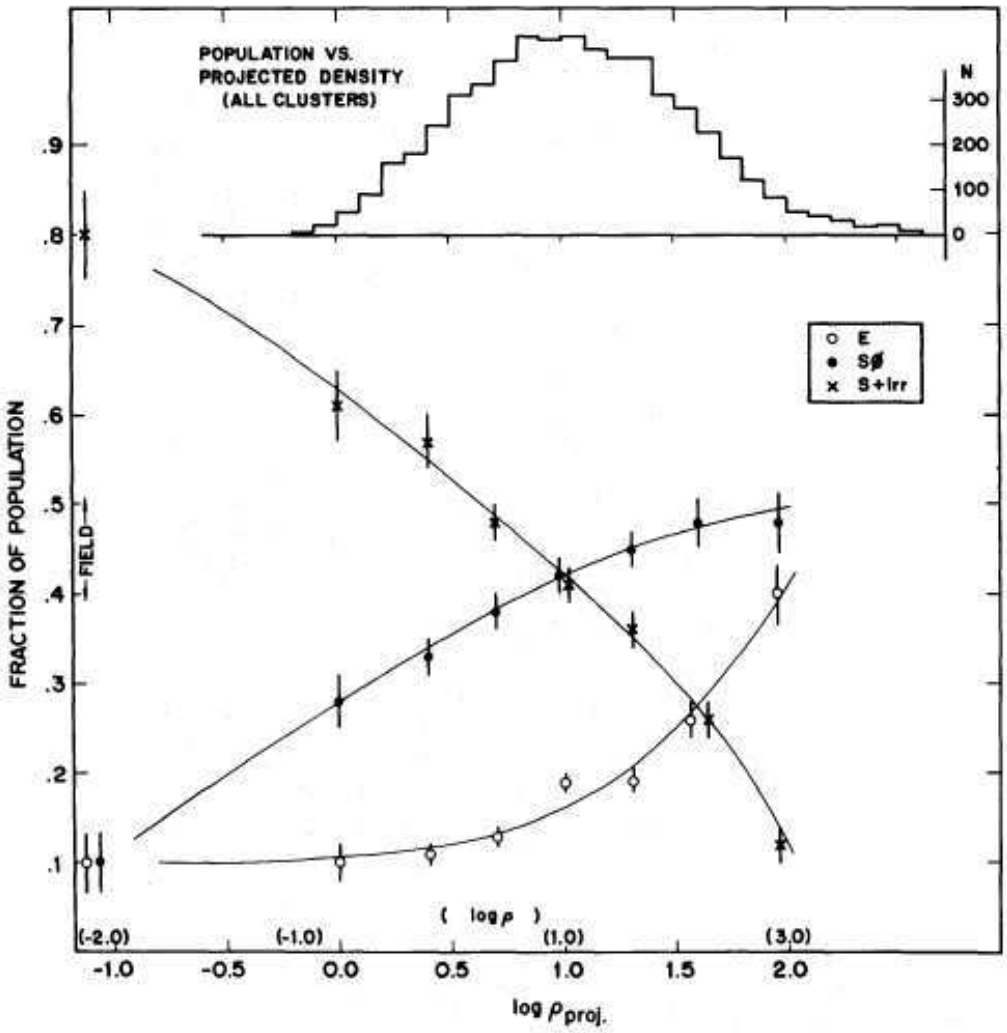

Figure 1.2. The morphology-density relation, reproduced from Dressler (1980). From the outskirts to the centres of clusters there is a smooth increase in the fractions of S0 and elliptical morphological types. These observations have since been extended to encompass smaller group and field environments, and can equally be applied to colour: red galaxies dominating high density environments and blue galaxies areas of lower density.

a mystery.

Even during Hubble's time, it was apparent that the distribution of early-type galaxies in particular is not uniform across the sky. Rather, they tend to group together and are found in regions of greater galaxy density (Abell 1965). This latter point was quantified by Dressler (1980), who showed that for rich clusters of galaxies the local number density of galaxies was related to morphological type. This work is reproduced as figure 1.2, and has since been extended to more general environments (e.g. Postman & Geller 1984; Bamford et al. 2009; Skibba 2009). The trend for elliptical galaxies to dominate the densest regions, while being sub-dominant overall, became known as the morphology-density relation and was an early piece of evidence for different evolution mechanisms in spiral and elliptical galaxies.

Despite intriguing correlations such as this one, the use of morphologies alone in investigating galaxy evolution is limited. For instance, bulge-dominated spirals can easily be mis-classified as ellipticals, particularly at greater distances. Furthermore, the volume of imaging data now at our disposal makes visual classification of full samples impractical at best. Innovative techniques, such as the Galaxy Zoo project (Lintott et al. 2008), and computer-defined morphologies (e.g. CAS,GINI, Conselice 2003; Lotz et al. 2004) enable scientific studies of morphology even with very large samples. However, even through the use of such techniques, morphologies of galaxies at high redshift are highly uncertain and other properties are often preferred (e.g. Mancini et al. 2010).

Through the study of other properties of galaxies, such as their colours, physical sizes and spectral signatures, it became apparent that there were important correlations between these quantities and galaxy morphology (Faber & Jackson 1976; Tully & Fisher 1977; Guzman et al. 1993; Jorgensen et al. 1996). Early-type galaxies lie almost exclusively at the red end of optical colour distributions, where 'colour' is defined as the difference in magnitudes of an object observed using two broadband filters. In fact, when a fair sample of galaxies in the present-day universe is taken, the colour distribution has a curious bimodality (Kauffmann et al. 2003; Balogh et al. 2004). Upon adding additional information, such as the luminosity of the galaxies in the sample, two rather distinct populations were found.

Figure 1.3, from Skibba & Sheth (2009), shows the distribution of galaxies in the Sloan Digital Sky Survey (SDSS, York 2000). Immediately apparent are the two regions described above. The tight red sequence in the upper region of the diagram is dominated by morphologically early type galaxies, particularly at the brighter end. Only around 1/3rd of the galaxies on this red sequence have Sa or later type morphologies (e.g. Wolf et al. 2005), despite the vastly greater number of spiral galaxies in general. The blue population on the other hand show a much broader range of colours, and in contrast to the red sequence there is not as strong a correlation between colour and luminosity. This colour-magnitude relation is observed to be in place to at least $z = 1$ (Bell et al. 2004), but begins to disappear by $z = 2$ (Cirasuolo et al. 2007).

The light we receive from a galaxy is the integrated light of the stellar population that it contains. The colours of the galaxies can therefore be understood in the context of stellar evolution. The most massive stars emit more strongly at bluer wavelengths, but these stars are short-lived. Despite their short life, they are very bright and only a modest amount of young, massive stars are required to cause a galaxy to appear blue (Trager et al. 2000; Ellis et al. 2001). A blue colour therefore indicates the presence of ongoing star-formation, though the bulk of the stellar mass in the galaxy may be old. Low mass stars contribute relatively little to the galaxy light and so a red galaxy's light is typically dominated by evolved, red giants (though the effects of dust are a complicating factor). The precise age of an evolved star has little effect on its colour and this is a major reason why the red sequence is observed to be so tight. Bower et al. (1992) and Stanford et al. (1998) used the lack of scatter in the red sequence to infer that the majority of star formation in these galaxies must have been either at very similar times (if the stars are presumed to be reasonably young) or were formed in the early ($z > 2$) Universe. With no physical reason for so many stars to be formed simultaneously and at low redshift, they each concluded that the stellar populations were likely to be very old.

Spectroscopy has confirmed these findings, by identifying emission lines in blue

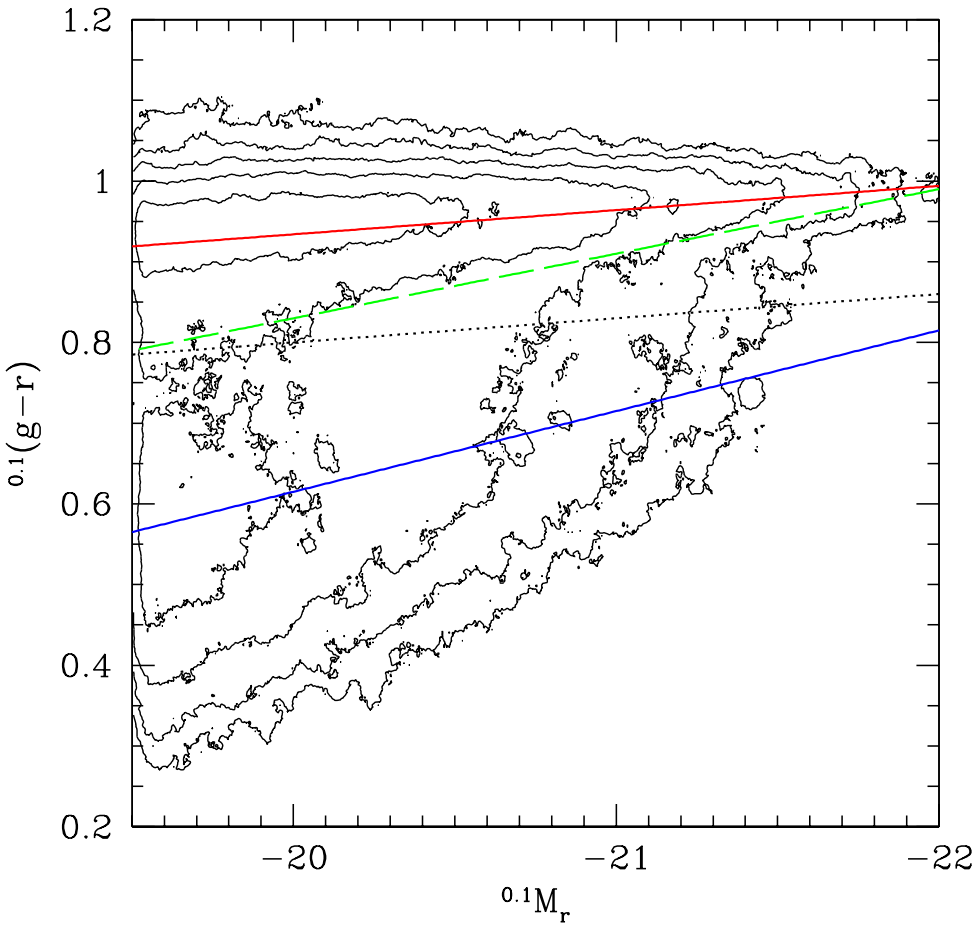

Figure 1.3. This colour-magnitude diagram of galaxies from the Sloan digital sky survey (SDSS) exhibits a bi-modal galaxy population: a tight 'red sequence' and a 'blue cloud', with a more varied range of colours. This figure, taken from Skibba & Sheth (2009), represents almost 100,000 galaxies in the low-redshift Universe. The '0.1' superscript was used by Skibba & Sheth (2009) to highlight the fact that the quantities are k-corrected to $z = 0.1$.

galaxy spectra and the presence of a distinct break at 4000Å ($D_n(4000)$) in the spectra of elliptical galaxies on the red sequence. Strong emission lines, such as the forbidden [OII] line are usually a result of the ionising flux from massive, young stars and are sensitive enough to the level of current star formation that they can be used in estimates of the star-formation rate (SFR) of a galaxy. The break observed in the vast majority of elliptical galaxies is also present in star-forming galaxies, but to a much lesser degree. It is caused by the co-incidence of the balmer break at 3646Å and metal absorption lines, particularly the calcium H and K lines. Lower mass stars emit very little flux beyond the balmer break, and so the strength of this break is dependent on both the age of the stellar population and the metalicity of the galaxy.

It is therefore apparent that elliptical galaxies contain, on average, much older

stellar populations than spiral galaxies. What is not clear however is why this should be the case. What mechanism could cause a galaxy to become passively evolving, with an elliptical morphology, as opposed to star-forming and spiral in structure? In order to tackle this question, we must first understand how the mass of a galaxy is assembled. The current most widely accepted model is that of hierarchical assembly.

1.1 Hierarchical assembly

The stellar component of a galaxy is only a fraction of its total mass. In addition to the stars and gas, each galaxy contains weakly interacting dark matter. On scales larger than the visible extent of a galaxy, this dark matter component dominates the gravitational interaction and is also responsible for the large-scale structure that we observe in the Universe.

Over the last decade our understanding of the development of structure throughout cosmic history has taken a significant step forward. The evolution of structure is dependent on the parameters which describe the cosmology of the Universe. Observations such as large-scale galaxy and cluster distributions (Romer et al. 1994; Cole et al. 2005), primordial temperature fluctuations in the cosmic microwave background (Spergel et al. 2007), supernovae distance estimates (Reiss et al. 1998; Perlmutter et al. 1999) and the abundance of deuterium (White et al. 1993; Burles et al. 2001) are sensitive to different cosmological parameters. The availability of high quality data in each of these areas has enabled a convergence of our cosmological model on a relatively small region of parameter space (Spergel et al. 2007).

The set of values on which the data converge is called the 'concordance cosmology' and forms an important basis upon which the theory of galaxy evolution is built. According to the concordance cosmology the Universe is made up of approximately 70% dark energy, responsible for the accelerated expansion of the Universe. The remaining 30% is mostly a non-baryonic form of matter termed 'dark matter'; only around 4% of the density of the Universe is made up of the baryons from which stars are made (Spergel et al. 2007). It is important that the results of this thesis can be compared with previous work and the results of numerical models. The set of cosmological parameters that were favoured before the latest WMAP results are therefore used throughout this thesis: $H_0 = 100h = 70$ kms^{-1}Mpc^{-1}; $\Omega_m = 1 - \Omega_\Lambda = 0.3$; $\sigma_8 = 0.9$; $n_s = 1$ (Spergel et al. 2003).

Equipped with this information, cosmologists have used both analytic and computational N-body techniques to describe in detail how the structures that we find around us were formed (e.g. Mo & White 2002; Smith et al. 2003; Springel et al. 2005). The dark matter density fluctuations of the very early Universe were approximately Gaussian and the very highest peaks of the density distribution were the first to collapse under self gravity and de-couple from the Hubble flow Peebles (1980); Coles & Lucchin (1995). Their assembly is entirely determined by gravity. Over time these small, collapsing dark matter halos accrete more material from the surrounding Universe, growing in mass. Also under gravity, they merge with other halos, building ever more massive structures (Lacey & Cole 1993). This mass growth of dark matter halos, with the smallest objects being the first to form, is termed 'hierarchical assembly'. Merging of halos is on-going even today and is seen in clusters of galaxies such as Abell 2218. Figure 1.4, reproduced from Lacey & Cole (1993), shows schematically the hierarchical assembly process. Time runs from top to bottom, with t_0 the time at observation and t_f the formation time. The width of

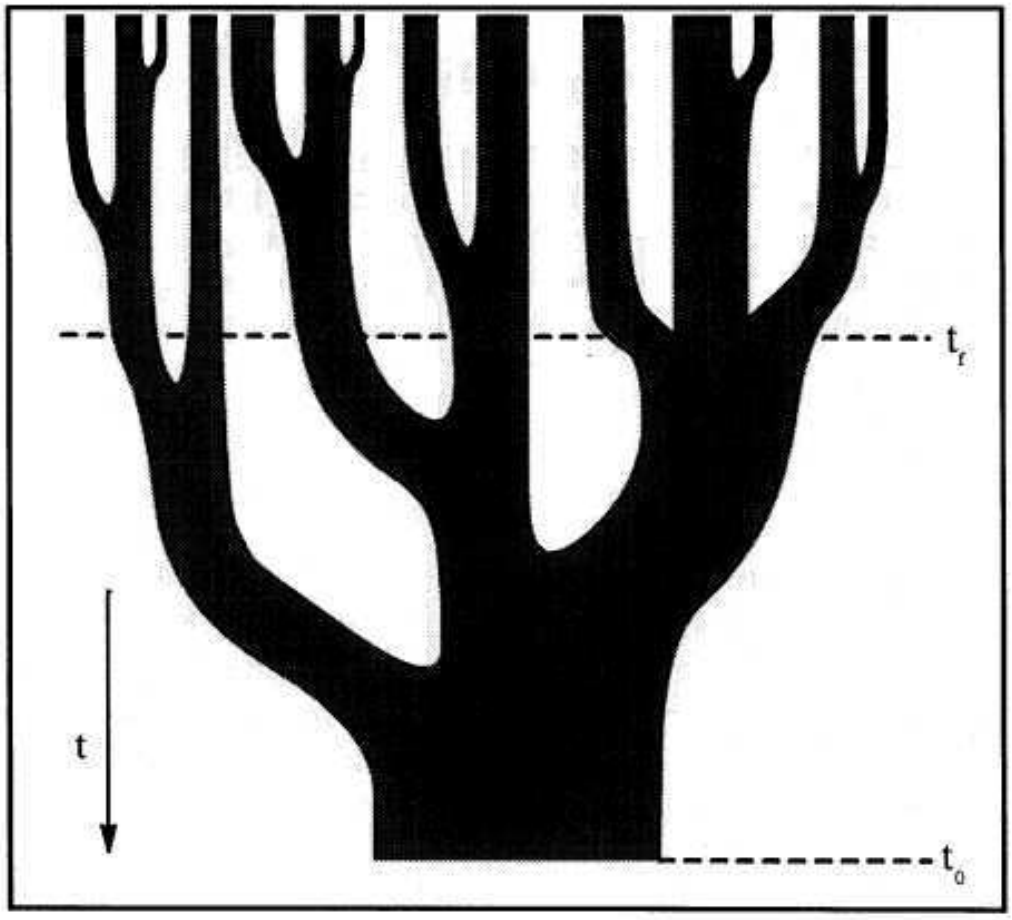

Figure 1.4. This schematic diagram of the build up of a halo's mass through a hierarchical assembly is taken from Lacey & Cole (1993). The width of the branches represent the masses of the halo and its progenitor pieces at any given time. The formation time, t_f, is taken here to be the earliest time when one of the progenitors contained half of the final halo mass.

each branch represents the mass within that halo and the formation time is defined as the earliest time when one of the progenitors contained half of the final halo's mass.

From dark matter N-body models, such as the Millennium simulation (Springel et al. 2005), and analytical theory (Press & Schechter 1974) the abundance of dark matter halos as a function of mass (the halo mass function) and their statistical spatial distributions is predicted for any set of cosmological parameters. Predictions of how the mass function and clustering properties evolve with time was a crucial step in the development of the halo model (e.g. Cooray & Sheth 2002) (a more detailed description of which is given in Chapter 5).

1.2 Galaxy formation

The dark matter component, although being dominant in structure formation is unfortunately not the one that we directly observe. All current observations are of radiation from the baryonic component of the Universe. However, to first ap-

proximation the baryonic mass traces that of the dark matter. The basic principles behind the modern understanding of galaxy formation in a hierarchical universe were laid out by White & Rees (1978). Under the influence of gravity, gas collects in dark matter halos where it cools through thermal Bremsstrahlung emission and line cooling. The cold gas condenses into large molecular clouds, inside of which stars are formed, following a mass distribution: the stellar initial mass function (IMF). This mass function, derived from observations in the local Universe, is generally assumed to be universal. Though some time-evolving mass functions have been used in the literature (e.g. Baugh et al. 2005), there is currrently no compelling evidence for such evolution.

As more massive dark matter halos can gather a greater reservoir of gas, one might naively expect that the mass of a halo and the stellar mass or luminosity of the hosted galaxy are proportional. However, the abundance of galaxies as a function of their absolute magnitude (luminosity function) and the mass function of dark matter halos are very different. In order to align the two, the conversion of gas into stars must be suppressed in both the most massive and the least massive halos.

Figure 1.5, taken from Benson et al. (2003), illustrates this difficulty. The data points show the K-band galaxy luminosity functions of Cole et al. (2001) from the 2-degree field galaxy redshift survey (2dfGRS), Kochanek et al. (2001) from the 2micron all sky survey (2MASS) and Huang et al. (2003). The short-dashed line is the halo mass function multiplied by a constant mass-to-light ratio, chosen to fit the knee in the luminosity function. The luminosity function can be fit by a Schechter function (Schechter 1976):

$$\Phi(L)dL = \Phi_* \left(\frac{L}{L_*}\right)^{\alpha} exp\left(-\frac{L}{L_*}\right)(dL/L_*), \tag{1.1}$$

while the halo mass function is closer to a simple power law. Either side of L^* the mass-to-light must change, i.e. the star-formation must be strongly suppressed.

This mass dependence of the mass-to-light ratio indicates that there are physical processes that are not being accounted for in our simplistic assumption of galaxy formation. Star formation must be quenched, which in turn requires the removal of some of the star-forming gas for both low and high galaxy masses (Benson et al. 2003). One way to do so is to heat the gas and thus prevent it from cooling and forming stars.

1.2.1 Quenching star-formation

Supernovae and an active galactic nucleus (AGN) are two possible mechanisms that are able to output tremendous amounts of energy. Indeed, both of these phenomena are invoked in galaxy formation models as feedback mechanisms (e.g. Bower et al. 2006; Croton et al. 2006; De Lucia & Blaizot 2007). The near simultaneous explosions of a number of supernovae, as might be expected in regions of high star-formation rate, are thought to create hot bubbles of gas. These bubbles are energetic enough to expel significant amounts of gas from low mass galaxies. In addition, such gas outflows are observed, a famous example being Arp 220 (Arribas et al. 2001). Models utilising supernovae feedback of this sort are able to match the faint end of the luminosity function reasonably well (e.g. Bower et al. 2006). However, in massive galaxies the potential well is too deep for the gas to escape and it simply cools back onto the galaxy.

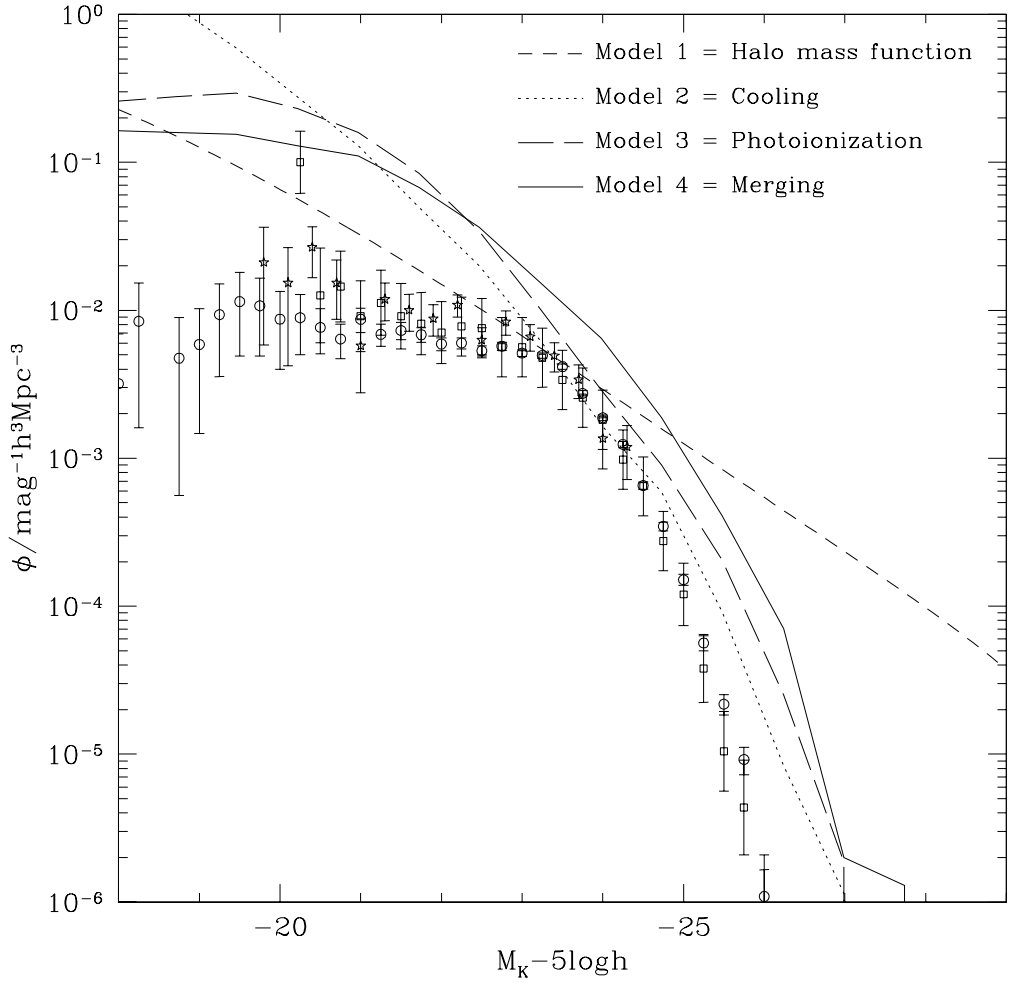

Figure 1.5. The $z = 0$ K-band luminosity functions of Cole et al. (2001), Kochanek et al. (2001) and Huang et al. (2003). Also plotted is the halo mass function which has been normalised to fit the knee of the luminosity functions by a constant mass-to-light ratio (short-dashed line). Credit: Benson et al. (2003)

In order for theoretical models to match the the bright end of the luminosity function an even greater source of energy is required: the energy output from an AGN (e.g. Croton et al. 2006). There is certainly enough energy produced by AGN to unbind the gas from even massive galaxies and, once expelled, to keep it heated (Silk & Rees 1998; Fabian 1999; Nesvadba et al. 2008). However, how this energy might couple to the gas remains unclear. There are two 'modes' of AGN activity which are observed: a bright quasar mode and a radio mode. In the quasar mode it is thought that the AGN could cause momentum-driven winds, far in excess of those typically seen for star-burst galaxies (Fabian 1999; King 2005). High velocity outflows have been observed from post-starburst galaxies of $\sim 1000\mathrm{kms}^{-1}$ by Tremonti et al. (2007), which were interpreted by the authors as likely to have been driven at least in part by an AGN. Further evidence has been found by Reeves et al. (2003), who observed a strong outflow of X-ray bright gas. Perhaps the most compelling

evidence, however, is that of the effects of radio mode feedback. Fabian et al. (2006) observed shocks in the intra-cluster medium of the Perseus cluster which were caused by the radio lobe cavities created by the AGN of the central galaxy. It is plausible that small shocks such as these could prevent cooling of the gas (Fabian et al. 2006).

In addition to being required to help match the bright end of the luminosity function, there is further evidence for the influence of an AGN on its host (Magorrian et al. 1998). Gebhardt et al. (2000) estimated the mass of the spheroidal component and central dark object of 36 early type galaxies, and found a significant correlation between the two. The central dark object is assumed to be a super-massive black hole (SMBH), of the sort that could power an AGN. This relation between black hole mass and the properties of the bulge implies a common origin of the two. An AGN is therefore a prime candidate for the missing mechanism needed to end star formation in a massive galaxy.

A merger between two galaxies should trigger a galaxy-wide star-burst, forming a galaxy with elliptical morphology (Toomre & Toomre 1977). It is thought that at these moments the SMBH will be able to accrete vast amounts of mass and shine as a bright quasar (Silk & Rees 1998). During this violent phase of a galaxy's evolution the gas that would otherwise be available for forming stars is heated and expelled from the galaxy. In this way it is thought that an AGN can be responsible for terminating the star-formation of a galaxy and also provides an explanation of how the central parsec of a galaxy can be so tightly correlated with the entire bulge component (Fabian 1999). However, the majority of the evidence described in this section is indirect and we still await the vital observation to confirm their role.

1.2.2 Downsizing

Observational evidence for the early termination of star formation in massive elliptical galaxies has been in existence for some time, longer than the quasar-based galaxy formation models described above. Cowie et al. (1996) used a nearly complete spectroscopic sample of galaxies to investigate the B-band and K-band luminosity functions. They also derived a time scale for the stellar mass of a galaxy to double from the rest-frame U-K colour. They observed that the bright end of the luminosity function evolved very little over the $0 < z < 1.6$ range which they probed. Evolution at the faint end, however, was significant. They defined galaxies which had a doubling time of less than 10^{10} years as 'in formation'. The critical result was that at low redshift, despite 10^{10} years being a significant fraction of the age of the universe, only the fainter galaxies were in formation, while at high redshift the sub-set of galaxies in formation included progressively brighter objects. This finding showed that more massive galaxies form their stars earlier than lower mass objects. Cowie et al. (1996) coined the term 'downsizing' to describe this behaviour. Following on from this work, many other authors have found that their results also imply downsizing in galaxy evolution (Bell et al. 2004; Kodama et al. 2004).

Downsizing initially seems at odds with the highly successful hierarchical structure formation: the smallest constituents are the first to form and they only later merge to form more massive systems. However, the two can be reconciled. If the merging galaxies contain very little remaining gas, having been through a quasar phase, then the star-burst associated with merging cannot occur. A merger of this sort is often referred to as a 'dry merger' (a thorough description can be found in Hopkins et al. 2010). As previously stated, Bower et al. (1992) showed that the lack of significant scatter in the red sequence implies a narrow redshift range over which

the stars in red sequence galaxies were formed. In a dry merger the resulting stellar population would be one with a very high formation redshift, exactly as found in observational studies.

Measurements of the evolution of the bright end of the luminosity function (Cirasuolo et al. 2010) show that massive galaxies only grow in stellar mass by a factor ~ 2 over the interval $0 < z < 1$. Dry merging is suspected to be responsible for much of this growth (e.g. Hopkins et al. 2010), but due to the difficulty in obtaining consistent measurements of merger rates at both high and low redshift, the precise rate of growth due to mergers remains unknown. What is certain, however, is that much of the stellar mass in the brightest (most massive) galaxies was already in place at high redshift ($z > 2$) (e.g. Drory et al. 2004; Cimatti et al. 2006; Zirm et al. 2007).

A second aspect of downsizing, is that massive, passive galaxies were almost fully assembled at $z > 1$. Their host halos, however, continue to grow in mass at $0 < z < 1$. This second apparent contradiction can also be rationalised. Mergers between dark matter halos do not imply an immediate merger between the central galaxies hosted by them. The dynamical timescale is higher in more massive halos (Binney & Tremaine 1987) and N-body simulations predict that mergers within massive clusters are rare (e.g. Ghigna et al. 1998). It is therefore plausible that the central galaxies may remain distinct long after the halos have merged. So, while the halos hosting them grow rapidly, the central galaxies of massive clusters do not (Foucaud et al., submitted).

1.2.3 Open questions

In order for massive, passive galaxies to already be in place at such an early epoch, they must have undergone very rapid star-formation. Galaxies detected at sub-mm wavelengths have been suggested to have star-formation rates of up to $1000 M_{\odot}$per year (Hughes et al. 1998). Such tremendous rates could certainly build up the stellar mass of the brightest galaxies and thus these sub-mm galaxies are identified as the progenitors of low redshift massive ellipticals (Smail et al. 1997). However, whether these galaxies continue to form stars at such a rate for a significant fraction of their history, or whether it is a short burst is as yet unknown. It is therefore unknown if these sources can account for all of the progenitors of massive ellipticals. Previous studies have identified a substantial population of passive galaxies at $z \sim 1$ (e.g. Cimatti et al. 2006), but optical datasets are unable to probe higher. The progenitors of low-redshift massive ellipticals are therefore only partly known and are deserving of much greater study.

The star-formation density of the Universe is thought to peak around $2 < z < 3$ (Madau 1995), and to study the galaxy population at these distances in an unbiased way requires deep infrared surveys. Data sets with the combination of both the depth to observe typical galaxies and the volume to find significant samples of massive objects are only now becoming available. We can now, for the first time, push our understanding of the general galaxy population to $z = 2$ and beyond. How are galaxies at high redshift distributed? Is a galaxy in a more massive halo more likely to become passive earlier, or is it solely dependent on the stellar mass of the galaxy itself? These are just two of the questions that we can now begin to tackle, and throughout this thesis I have made substantial steps in answering them. Without a doubt this is a very exciting time to be involved in extragalactic astronomy.

1.3 Techniques important to this work

In this section I shall introduce some of the techniques important to the work contained in this thesis.

1.3.1 Galaxy clustering

Much of this thesis concerns the results of galaxy clustering. Clustering is the statistical description of the spatial distribution of a galaxy population and is one of the oldest and most well-studied techniques in astronomy. There are a number of techniques that can be used to study the clustering of a sample, with the most common being the 2-point correlation function.

The 2-point correlation function, $\xi(r)$, is defined by the joint probability of finding two galaxies in infinitesimal volume elements, δV separated by a distance, r, (Groth & Peebles 1977; Peebles 1980):

$$\delta^2 P = n_g^2[1 + \xi(r_{12})]\delta V_1 \delta V_2, \tag{1.2}$$

where n_g is the galaxy space density. $\xi(r)$ must be estimated numerically from the data. This is performed by constructing a random catalogue within the same volume and with the same selection function as the galaxy catalogue. An estimator based on the counts of pairs of data or random points can then be used to find $\xi(r)$. The simplest choice for such an estimator is

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1. \tag{1.3}$$

Where DD and RR are pairs of data (galaxies) and randoms (randomly placed points), normalised by the total number of pairs summed over all separations. If $\xi(r) = 0$ the galaxy distribution is random.

Galaxies form within dark matter halos and the correlation function of a galaxy sample is intimately related to the typical mass of those halos: the more clustered a sample is, the more massive the typical dark matter halos which host them. This relationship comes about because massive dark matter halos are biased with respect to the overall dark matter distribution, forming from the densest peaks of primordial dark matter distribution. This direct link between a measurable quantity and a fundamental property of the unmeasurable dark matter hosts has helped maintain clustering as one of the avenues of greatest interest in extragalactic astronomy.

However, a computation such as the 2-point correlation function requires that we know the 3-dimensional positions of the galaxies in our sample. Typically, their angular positions on the sky are accurately known, but the radial distances between them are not. We therefore define, by analogy, the angular 2-point correlation function, $w(\theta)$:

$$\delta^2 P_{ang} = n_\Omega^2[1 + w(\theta_{12})]\delta\Omega_1 \delta\Omega_2, \tag{1.4}$$

which is the joint probability of finding two galaxies in solid angle elements, $\delta\Omega$, with an angular separation, θ, on the sky. The equivalent estimator, using pair counts at a given angular separation rather than physical separation, is

$$w(\theta) = \frac{DD(\theta)}{RR(\theta)} - 1. \tag{1.5}$$

In practice, though, this estimator is often insufficiently accurate. In cases where $w(\theta)$ is small the variance is overestimated and there is an additional bias introduced by edge effects (Landy & Szalay 1993; Hamilton 1993). Throughout this work I use the estimator suggested by Landy & Szalay (1993):

$$w(\theta) = \frac{DD - 2DR + RR}{RR}. \tag{1.6}$$

N_{DD} and N_{RR} have been replaced by DD and RR respectively, and a further quantity, DR, has been introduced. This quantity is the number of data-random cross pairs, again normalised by the total number possible. Using this estimator the variance is reduced to the expected Poisson level and the bias arising from field edges is removed. This is particularly important in pencil-beam survey strategies.

Despite this improved estimator for $w(\theta)$, there remains a small bias introduced in all but all-sky surveys by the finite field size. This bias is an integral constraint due to the limited number of galaxy pairs. If a sample is clustered, i.e. has positive $w(\theta)$ at some small scale, then $w(\theta)$ must become negative at some other (larger) scale. This is corrected by adding a constant to $w(\theta)$. This constant depends on the form of $w(\theta)$ and is formally the double integral of $w(\theta)$ over the area of the field. However, if $w(\theta)$ is fit by a power-law such that

$$w(\theta) = A(\theta^{-\delta} + IC), \tag{1.7}$$

then, following Roche et al. (2002), this constant can be estimated using the random-random pair counts as

$$IC = \frac{\Sigma RR(\theta).\theta^{-\delta}}{\Sigma RR(\theta)}. \tag{1.8}$$

One of the most attractive aspects in using clustering for studying galaxy evolution is the link with the underlying dark matter properties. $w(\theta)$ is in fact the projection of the three dimensional correlation function on the sky. This simple fact means that when comparing two galaxy samples selected at the same epoch, the galaxies in the sample with greater amplitude (quantity A in equation 1.7) are on average hosted by more massive halos.

However, it is possible to do better than such a simple qualitative comparison. Limber (1957) derived the projection from the real-space correlation function to $w(\theta)$ and, following Peebles (1980), we can invert the Limber equation to deproject $w(\theta)$ and find the characteristic clustering scale length, r_0:

$$r_0 = \left[\frac{cA}{H_0 H_\gamma}\left(\frac{(\int_0^\infty n(z)dz)^2}{\int_0^\infty n(z)^2 x(z)^{1-\gamma} P(\Omega_0, z)F(z)dz}\right)\right]^{1/\gamma}. \tag{1.9}$$

This deprojection requires knowledge of the redshift distribution, $n(z)$, but if it is accurately known then the recovered scale length is accurate to $\sim 10\%$ (Simon 2007). In addition to the redshift distribution, the slope of the real-space correlation function must be known ($\gamma = -\delta - 1$), and a cosmology must be assumed. The remaining quantities are (e.g. Magliocchetti & Maddox 1999): the speed of light in a vacuum, c; co-moving distance at redshift z, $x(z)$; $F(z) = 1$ for the assumed cosmology ($\Omega_M = 1 - \Omega_\Lambda = 0.3$, $h = 0.7$, $\sigma_8 = 0.9$); $H_\gamma = 3.68$ when $\gamma = 1.8$ (as is assumed in the work presented in this thesis); and $P(\Omega_0, z) = \Omega_0^{1/2}[(1+z)^3 + \Omega_0^{-1} - 1]^{1/2}$.

The derived quantity, r_0, is related to $\xi(r)$ in the following way:

$$\xi_{\rm deproj}(r) = \left(\frac{r}{r_0}\right)^{-\delta-1}, \tag{1.10}$$

where the subscript, deproj, represents that the real space correlation function is a single power law deprojected from $w(\theta)$, rather than estimated directly. Further details will be given in the relevant sections as appropriate.

1.3.2 Colour selection

In order to study the evolution of galaxies we require samples of galaxies at different epochs which we can use to compute statistics such as the auto-correlation function. In other words, we need to know the redshifts of large numbers of galaxies which can be very faint in the optical bands. Unfortunately obtaining spectra of optically faint galaxies is extremely difficult, especially those with weak or absent emission lines. Infrared multi-object spectrographs are currently few and heavily over-subscribed, while using optical spectrographs is expensive at best and impractical in the worst cases. In addition there are biases arising from the lack of 100% completeness in redshift determination that need to be accounted for.

These difficulties in obtaining representative samples at high redshift has led to the development of several photometric-based redshift selection criteria. Deep photometry is far cheaper than spectroscopy, and the simplest of these criteria require only two bands (a single colour). Examples of these single-colour selection techniques are the extremely red object (ERO) criterion (McCarthy et al. 1992; Cowie et al. 1994), requiring $I-K>5$, and the distant red galaxy (DRG) criterion (Franx et al. 2003), using two near-infrared bands: $J-K>2.3$ (Vega).

Both of these selections were designed to take advantage of the sharp decline in flux received from evolved galaxies blue-ward of 4000Å (rest frame). EROs are expected to lie in the range $1<z<2$, where the lower limit corresponds to the redshift at which the 4000Å break enters the I-band. The upper limit could theoretically be higher, but in practice completeness issues restrict how high in redshift such sources can be found. The DRG selection works in much the same way, but the choice of filters moves the selection to higher redshift. DRG sources are found to be predominantly at $z>2$,though the brightest DRGs tend to be at $z<2$.

Though these techniques were designed to find evolved galaxies at high redshift, it became apparent after follow-up observations that there were many dusty, star-forming contaminants (Cimatti et al. 2002a). Pozzetti & Mannucci (2000) noted that, although the $I-K$ colours of such dusty galaxies and evolved galaxies are similar, the overall shape of their respective SEDs are not. Using a band intermediate between the I and K-bands (J-band) they proposed a second colour criterion that could approximately separate the two sub-populations. This second delimiter was $J-K=0.36(I-K)+0.46$, with passive EROs (pERO) bluer in $J-K$ than the star-forming EROs (sERO). Figure 1.6, from Pozzetti & Mannucci (2000) shows a dusty star-forming and a passive galaxy SED at $z=1.5$, together with the I, J and K filters that they used.

Despite this success, neither the ERO nor the DRG selection are able to provide a complete census of galaxies over their respective redshift ranges. The most populous, 'normal', galaxies are missed and additional sets of criteria are required to identify

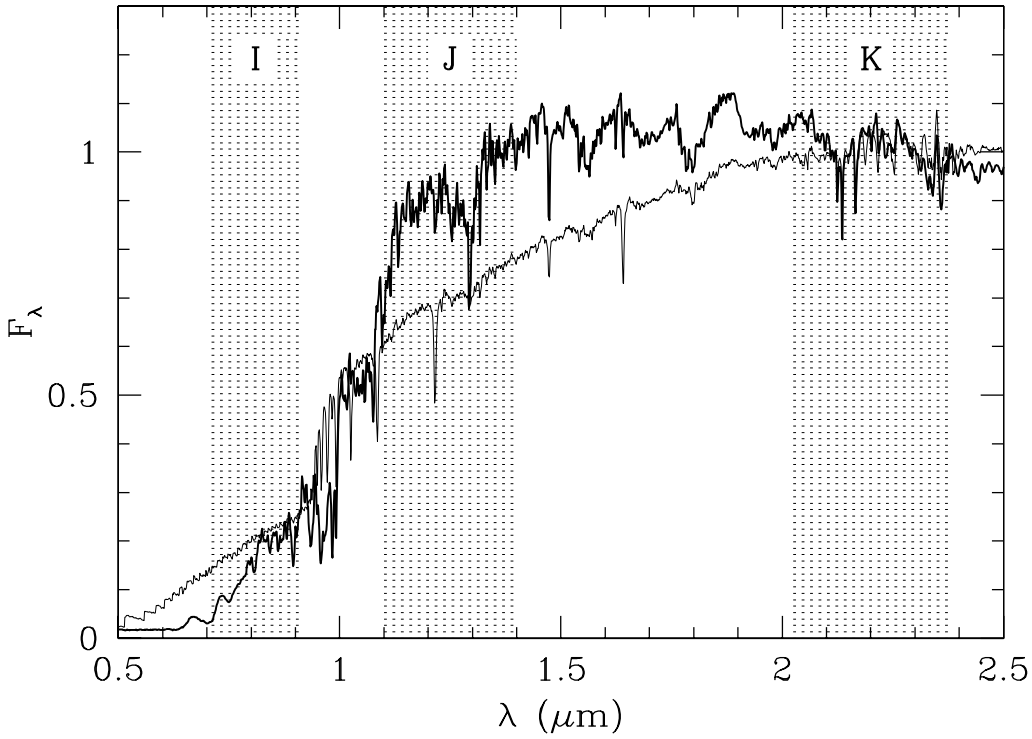

Figure 1.6. SEDs for an old galaxy and a dusty star-forming galaxy at $z = 1.5$. The $I - K$ colours are very similar, but the two can be distinguished by the addition of J-band information.

them. Steidel et al. (1996, 2003) used three broadband filters to select galaxies which 'drop-out' of the images taken using the shortest wavelength filter, in the ultra-violet. Using follow-up spectroscopy, they showed that the drop-outs were galaxies at very high redshifts, $z > 3$. The UV filter is shortward of the Lyman limit, a region of the SED past which even star-forming galaxies emit very little flux. This 'drop-out' technique has since been applied to work at lower redshifts (BM/BX Erb et al. 2003; Adelberger et al. 2004), and to probe the very high redshift universe (Bunker et al. submitted, McLure et al. 2009).

Comparing samples selected by two such very different techniques is troublesome. The effective redshift ranges are likely to be different and there may also be overlap between the samples. A more satisfactory approach would be one akin to the passive and star-forming ERO definitions of Pozzetti & Mannucci (2000), but without being biased against unobscured star-forming galaxies. This desire motivated the two-colour selection proposed by Daddi et al. (2004), commonly called the BzK selection after the three filters that are required. Using the near spectroscopically-complete K20 survey (Cimatti et al. 2002), Daddi et al. (2004) defined two boundaries which together simultaneously select star-forming and passive galaxies over a broad range in redshift ($1.4 < z < 2.5$). Furthermore, the boundaries were arranged such that the effects of dust on inclusion in one of the samples were minimised. This selection technique is described in more detail in Chapter 2, where it is used to investigate the clustering properties of high redshift passive and star-forming galaxies.

1.3.3 Photometric redshifts

The photometric redshift technique (Baum 1962), like the colour-selection technique described above, is used to obtain redshifts from broadband photometry. It therefore shares many of the advantages of colour selections, such as being able to probe deeper than the current spectroscopic limits. But unlike the simple colour-selections described above, there is not a pre-determined redshift range that the target galaxies are assumed to lie within.

In recent years photometric redshifts ("photo-zs") have become increasingly popular and surveys are routinely designed with the technique in mind. The technique is of particular importance for high redshift galaxy populations, such as those studied in this thesis, as obtaining spectra for such sources is extremely expensive. However, even at the more modest redshifts probed by the SDSS (York 2000) and COMBO-17 (Meisenheimer & Wolf 2002) surveys, photometric redshifts have been invaluable.

Obtaining a photometric redshift for a galaxy involves minimising the χ^2 of the available photometry over a large suite of galaxy templates. Once a template and redshift combination have been ascertained, further quantities of interest can be computed, making the technique extremely valuable. However, a detailed set of templates will not produce accurate redshifts if the number of photometric points is small. This is also true if the photometry are poor or if the object is very faint. While not all survey areas can have quite the photometric coverage of COMBO-17, biases introduced by the redshift errors must be considered while interpreting scientific results.

Figure 1.7 shows a best fit template for an example set of galaxy photometry taken from the SXDS / UKIDSS UDS survey (see below). The lower panel shows the distribution of χ^2 values for the best fit from the template set at each redshift. Clearly the galaxy is at very high redshift, $z \sim 4$. But this is not the only possible solution. There is a secondary solution, strongly disfavoured in this example, at $z \sim 0.5$. The error on a photometric redshift is taken to be the point either side of the χ^2 minimum where $\chi^2 = \chi^2_{min} + 1$ throughout this thesis. In some cases, however, the true redshift may be that of the secondary solution, or there are simply no solutions which match the true redshift. These cases are termed 'catastrophic errors' and were one of the misgivings people had about the technique when it was first used.

In the example shown the galaxy has very little flux short-ward of 5000Å , and could have easily been identified by a drop out technique. Strong features such as a Lyman break or 4000Å break allow only a small tolerance in the preferred redshift so, of the faint galaxy population, passive galaxies and galaxies above $z = 3$ have among the most reliable photometric redshifts. The converse must also therefore be true however. The galaxies most difficult to obtain photometric redshifts for are modestly star-forming galaxies in the range $1 < z < 3$; the majority of sources studied in this thesis.

Photometric redshift codes are typically trained and tested on spectral data, as shown in Figure 1.8. Therein lies one of the biggest problems: spectra are expensive and difficult to obtain at high redshift, which is why photometric redshifts are desired in the first place. Furthermore, the spectra of high redshift galaxies that are available are often of peculiar objects, intrinsically bright objects or of a biased sample. An ESO large programme, the UDSz, which is targeting a few thousand K-selected, $z_{phot} > 1$ galaxies will go someway to addressing this issue. However, even in this case the use of an optical spectrograph required compromises during object selection.

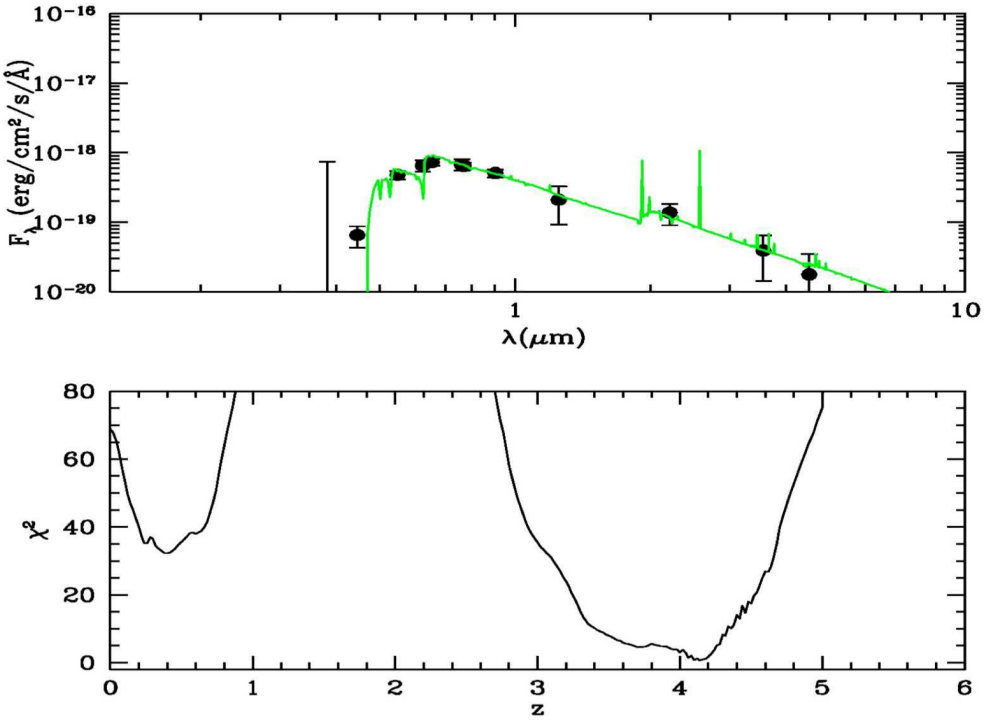

Figure 1.7. The photometric redshift technique consists of minimising the χ^2 of the available photometric data over a large suite of templates. The upper panel shows an example of the photometry of a galaxy with the best fitting template. The lower panel shows the χ^2 distribution with redshift.

There are photometric redshift codes publicly available (e.g. Bolzonella et al. 2000) which are relatively easy to set up and run. However, the default parameters are unlikely to be suitable for all applications, and in reality a great deal of work and experience is required before the resulting redshifts are robust. It is partly for these reasons that, despite being of crucial importance throughout this thesis, the photometric redshifts used in this work were not computed by myself. All redshifts were computed by Dr Cirasuolo (Royal Observatory Edinburgh and UK Astrophysical Technology Centre) exclusively for the UDS working group.

1.4 Primary data set: the UKIDSS UDS

The study of galaxy evolution at high redshifts requires highly advanced survey data. The flux received from a source falls off as the square of the distance, so we need deep data from which to extract the very faint sources of interest. In addition, due to the redshift effect data taken at optical wavelengths was emitted by distant sources ($z > 1$) in the ultraviolet. Passively evolving or very dusty objects at high redshift will therefore be extremely difficult, if not impossible, to detect in optical survey data. To circumvent this problem we require data to be taken at near infrared wavelengths ($1 - 2.5\mu$m).

The history of infrared astronomy is much shorter than optical astronomy. Until

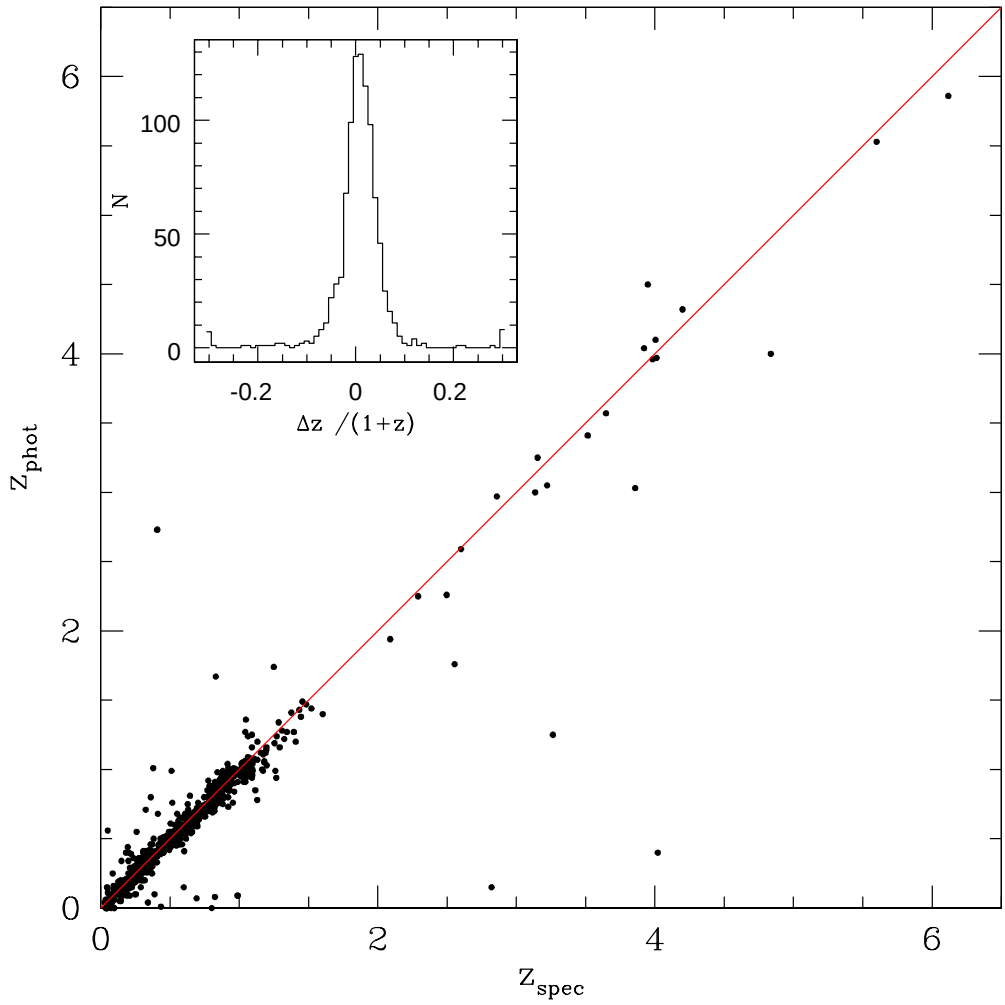

Figure 1.8. Photometric redshifts are trained and tested on spectroscopic data. In this case, from the UKIDSS UDS, the agreement is extremely good ($\Delta z/(1+z) \sim 0.035$), and there are very few catastrophic errors. However, the vast majority of spectroscopic redshifts are below $z = 1$.

recently infrared detectors were simple bolometers, recording the power of photons collected from a telescope pointing. However, the US military commissioned the development of CCD-like detectors at near-infrared wavelengths, and these were adapted for use in astronomy shortly after. Near infrared survey data, and science results utilising that data, quickly followed the introduction of such detectors. However, these data sets were still unsuitable for addressing the questions I shall tackle in this thesis. The surveys conducted were either wide area and shallow, such as the 2-micron all-sky survey, or very deep 'pencil-beam' surveys (e.g. FIRES, Franx et al. 2003). In order to study the evolution of structure formation and passively evolving galaxies at high redshift we need a data set which combines both deep data at near infrared wavelengths and a wide area. Such data sets are only now becoming available and throughout this thesis I make extensive use of data from the currently leading panoramic near-infrared survey, the UKIDSS Ultra Deep Survey.

The construction of a new wide-field camera (WFCAM, Casali et al. 2007) for the United Kingdom Infrared Telescope (UKIRT) prompted a call for proposals in 2001 to make use of the new instrument. Among the successful applications was the UKIRT Infrared Deep Sky Survey (UKIDSS, Lawrence et al. 2007. UKIDSS comprises five sub-surveys, with the deepest of them being the Ultra-Deep Survey (UDS, Almaini et al. in prep.). UKIDSS data acquisition began in 2005 and has already seen six data releases, beginning with an early data release of the first few hours of data (EDR, Dye et al. 2006). Each semester thereafter has produced a data release, though for the UDS these are only once a year as the field is only visible during a single semester. Further details of the exact data sets used will be given in the appropriate sections throughout this thesis, however here I shall briefly describe the current status of the UDS.

The UDS is a single mosaic of four WFCAM pointings, forming a contiguous square field of 0.88×0.88 degrees. The original science case required data in the J, H and K-bands to depths of $J = 26$, $H = 25.4$, and $K = 25$ in the AB magnitude system. The most recent release, in September 2009, was the DR5. The 5σ, point source depths in an aperture of $2''$ diameter in AB magnitudes are $J = 24.0$, $H = 23.7$ and $K = 24.0$. These depths make the UDS the deepest contiguous near infrared field over an area of ~ 1 square degree.

Through the use of such deep, wide-field infrared data it is possible to probe the same range of galaxy properties at high redshift that have been well studied in the low-redshift Universe. However, near infrared data alone are limited. Aside from the DRG selection (see section 1.3.2), near infrared photometry cannot tell us the redshifts of the galaxies in the survey. The co-incidence with data at other wavelengths, particularly optical wavelengths, is crucial and was one of the main considerations in the choice of field for the UDS.

The field selected for the UDS was the Subaru-XMM Deep Field (SXDF) (Sekiguchi et al. 2005). The optical data, taken with Suprime cam on the Subaru telescope, are extremely deep reaching $B_{AB} = 28.4$, $V_{AB} = 27.8$, $R_{AB} = 27.7$, $i'_{AB} = 27.7$ and $z'_{AB} = 26.7$ (3σ, $2''$ Furusawa et al. 2008). The field is also covered by ultra-violet data from CFHT Megacam and *Spitzer* data from a legacy survey reaching 5σ depths of 24.2 and 24.0 (AB) at 3.6μm and 4.5μm respectively (SpUDS, PI:Dunlop). These data allow the construction of highly accurate photometric redshifts (see section 1.3.3) which are of critical importance to the rest of this thesis. In addition the field is covered by X-ray data from XMM-Newton, Sub-mm data taken with SCUBA for the SCUBA Half A Degree Survey (SHADES, Coppin et al. 2006) and 1.4GHz radio data from the VLA to $12 - 20\mu$Jy.

1.5 Thesis structure

My thesis concerns the origin and evolution of the galaxy bimodality, and in particular the build-up of the passive galaxy population. I will address the question of where in the Universe passive galaxies are found at a given epoch, in terms of the mass of dark matter halos which host them. I shall also determine the same information for star-forming galaxies and will therefore be able to trace through cosmic history the dominance of passive or star-forming systems in the more massive structures of the universe.

I begin in chapter 2 with an analysis of colour-selected populations of passive and star-forming galaxies at high redshift, using the unique UDS data set to make

new findings regarding their properties. The shifts in colour required to match the original selection definition are computed on a galaxy-by-galaxy basis, as these shifts depend slightly upon the galaxy SED. I then compute their K-band number counts and angular clustering properties and deproject the clustering measurements to find their correlation lengths, r_0. Using this information I place the two populations in the overall context of galaxy evolution and motivate the main investigation of this thesis, which features in chapter 3.

In chapter 3 I make use of photometric redshifts and properties derived from them, namely rest-frame K-band luminosity, rest-frame $U - B$ colour and star-formation histories, parameterised by an age and exponential decay rate. I split the UDS galaxy sample into sub-samples in both redshift and K-band luminosity and compute their correlation lengths. These sub-samples are then further split by colour and I use the star-formation histories to define a passive and a star-forming sample. Correlation lengths for each of the red, blue, passive and star-forming sub-samples are found and I plot the passive and star-forming measurements together with clustering evolution tracks for dark matter halos of fixed mass. I find that at $z < 1.5$, red galaxies are more strongly clustered than blue galaxies and that passive galaxies are the most highly clustered of the red sample. While red and blue correlation lengths become indistinguishable above $z = 1.5$, passive and star-forming samples remain distinct, possibly converging at $z > 2$. An interpretation of these results follows the measurements.

In chapters 4 and 5 I attempt to gain a more detailed picture of the properties of the UDS galaxies. A technique closely related to the auto-correlation function, the marked-correlation function, is introduced and applied to the UDS data in chapter 4. The results of this technique are in broad agreement with those of the previous chapter (chapter 3). As the work is preliminary the results are unable to expand greatly on the earlier chapters at this stage. However, the work presented demonstrates the potential of the the technique as a productive avenue of further investigation. Finally, the observational data are compared to one of the current leading semi-analytic models. A mock observation from the model is treated in the same way as the real data in chapters 2 and 3. I use the data to highlight two known issues with the model and give tentative evidence for two further potential issues.

Chapter 2

The statistical properties of BzK-selected galaxies

In this chapter I use a 2-colour selection technique to study the statistical properties of galaxies at high redshift. Much of the work presented in this chapter features in a paper written by me, Hartley et al. (2008). In addition to writing the paper I contributed the clustering analysis, whilst the sample selection, number counts and luminosity function were performed by co-authors of the paper. In the preparation of the paper I independently performed the sample selection, based on the filter transforms, and computed the number counts.

Since publication, it has come to light that the filter transforms used in Hartley et al. (2008) are incompatible with a further field with data of comparable depth to the data used in the paper (McCracken, private communication). This new field was observed using filters more closely matched with those used in the original definition of (Daddi et al. 2004). In response to these new data, I revisit the selection method to improve our match with the original BzK definition and hence the analysis I present in this chapter differs slightly from that of the published work. I have re-computed the number counts and clustering properties for this thesis, but not the luminosity function (which was not my work).

2.1 Introduction

There is growing evidence to support the view that the most massive objects in the Universe were the first to assemble and complete their star formation (Kodama et al. 2004; Thomas et al. 2005; De Lucia et al. 2004; Bundy et al. 2006; Stott et al. 2007); this phenomenon has become known as *downsizing* (Cowie et al. 1996). A number of key issues remain unresolved however. There are indications that the build up of the galaxy colour bimodality occurs around $z \sim 2$ (e.g. Cirasuolo et al. 2007), but the precise evolutionary path from the distant Universe to the present day is still undetermined. The mechanism that terminates the major episode of star-formation is poorly understood (Benson et al. 2005), and it is now suggested that massive galaxies undergo significant size evolution from $z \sim 2$ to the present day (e.g. Cimatti et al. 2008 and references therein).

Galaxy clustering is an important tool for investigating these populations, since the amplitude of clustering on large scales (> 1Mpc) can provide a measurement of the dark matter halo mass (Mo & White 1996; Sheth & Tormen 1999). In principle, therefore, one can relate galaxy populations from the distant past to the present day

by tracing the evolution of dark matter halos within the context of a framework for structure formation.

A full exploration of these issues will require large spectroscopic surveys of IR-selected galaxies over a representative volume of the distant Universe. Such surveys are currently at an early stage, so recent work has focused on the photometric colour selection of passive vs. star-forming galaxies at the crucial $z \sim 2$ epoch. The two key methods to date are the BM/BX selection (Erb et al. 2003), which is an extension of the Lyman-break dropout technique, and the BzK technique (Daddi et al. 2004) which is based on K-band selection. Recent studies have shown that the BM/BX technique is reasonably efficient at selecting actively star-forming galaxies at $z \sim 2$, but largely misses the passive galaxy population at these epochs (Quadri 2007; Grazian et al. 2007). In contrast, the BzK method appears to be the most complete of the broad-band techniques for the selection of both star-forming and passive galaxies (Daddi et al. 2004; van Dokkum 2006; Grazian et al. 2007).

An alternative is to use a larger set of filters over a range of wavelengths and infer a galaxy's redshift from comparison of the calculated magnitudes with a set of templates (Cirasuolo et al. 2007 and references therein). This work makes use of such photometric redshifts and in principle one can derive the stellar age of a galaxy using such templates. However due to the uncertainties in determining stellar ages from the template fits, and the desire for comparison with the literature, the bulk of the present analysis is based on simple BzK selection.

2.1.1 BzK colour selection

The BzK selection technique was developed by Daddi et al. (2004) to take advantage of forthcoming deep, near-infrared surveys, made possible by progress in detector technology. Using the nearly spectroscopically complete K20 survey (Cimatti et al. 2002), it was designed empirically to preferentially select star-forming and passive galaxies in the redshift range $1.4 < z < 2.5$. The name, BzK, is derived from the 3 filters used: the B, z and K_s-broadband filters on the VLT instruments, FORS1 & 2 and ISAAC. The technique comprises two colour-colour selection criteria:

$$\begin{aligned} (z-K)-(B-z) &> -0.2 \qquad \mathrm{Star-forming\ (sBzK)} \\ \text{and}\ \ (z-K)-(B-z) &< -0.2 \\ \text{with}\ \ (z-K) &> 2.5 \qquad \mathrm{Passive\ (pBzK)} \end{aligned}$$

Based on initial K-band selection, in principle it does not suffer the same heavy biases caused by dust or the age of the stellar populations. The vector in the $(B-z)-(z-K)$ plane caused by dust extinction is parallel to the sBzK selection boundary. So not only are dusty sources identified due to selection in the K-band, but the classification as high or low redshift star-forming galaxies is also relatively robust against biases due to dust. Figure 2.1 shows the selection criteria and two sets of galaxy SEDs in the $(B-z)-(z-K)$ plane. The first set are those of Coleman et al. (1980) as updated in the publicly available photometric redshift code, *Hyperz* (Bolzonella et al. 2000); the second are a series of synthetic templates from Tremonti et al. (2004) using the stellar population synthesis code of Bruzual & Charlot (2003).

The BzK selection technique was tuned to classify galaxies with $K_s < 20$ and as such has not been tested for fainter galaxy samples. However, the empirical and

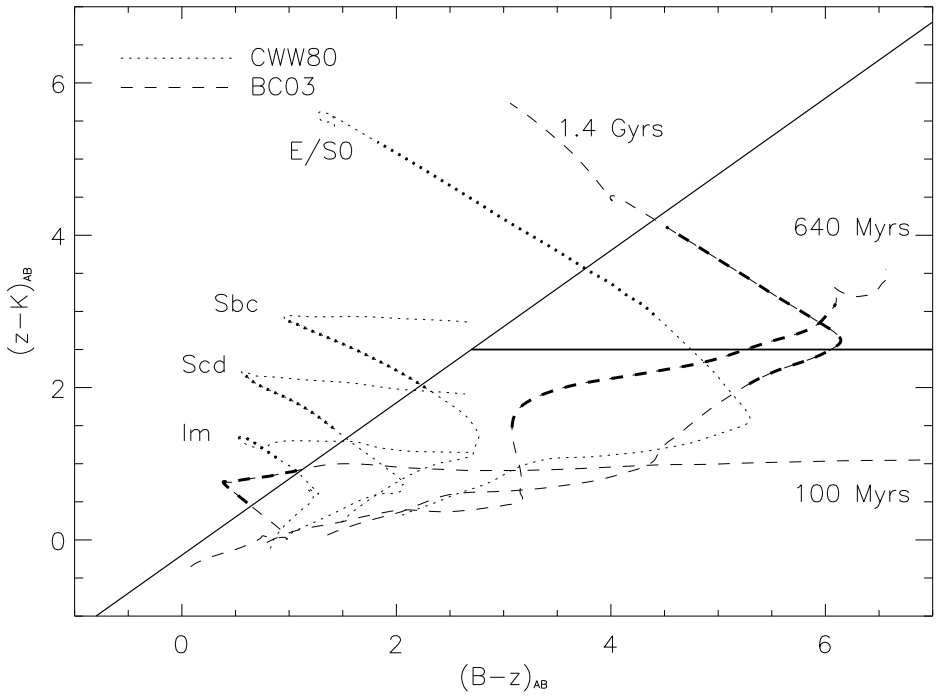

Figure 2.1. $(B-z)$ vs $(z-K)$ colour-colour plot of template SEDs evolved with redshift. Four empirical templates are shown from Coleman et al. (1980) together with three synthetic templates from the stellar population synthesis code of Bruzual & Charlot (2003). Bold sections indicate the regions of these tracks which have $1.4 < z < 2.5$. Also plotted are the selection boundaries of Daddi et al. (2004).

synthetic templates support the selection regions and so it is reasonable to use them, with the appropriate degree of caution, with fainter galaxy samples.

2.1.2 Previous BzK work

The K20 survey used by Daddi et al. (2004) in their selection definition covers an area of only 52arcmin^2, limiting the number of galaxies they selected to be at $z > 1.4$ to 55 objects. Nevertheless, they were able to estimate the space density of $z > 1.4$ galaxies with $K < 20$ (Vega) as $\sim 10^{-4}$Mpc^{-3}. This value is consistent with these objects being the progenitors of low redshift massive elliptical galaxies. The majority of these galaxies were found to be rapidly star-forming galaxies. The small area prevented Daddi et al. (2004) from being able to say anything about the relative number densities of the star-forming and passive samples.

Recent work has shown that the number counts of these star-forming and passive BzK-selected populations differ markedly (Kong et al. 2006; Lane et al. 2007), with star-forming galaxies being far more abundant at all magnitudes studied. In addition the deeper of the fields studied by Kong et al. (2006) exhibited a change in slope for the ERO and pBzK selections around $K_s = 19$.

Lane et al. (2007) used the Early Data Release (EDR) from the UKIDSS UDS to investigate the number counts and overlap in colour-selected galaxies, including those selected by the BzK technique. They found that the flattening in pBzK number counts observed by K06 extends to become a plateau at the EDR depth

of $K_{AB} < 22.5$ and noted that if this were to continue to turn over it could imply an absence of high-redshift passive galaxies at low luminosities.

The abundance of sBzK galaxies allowed Kong et al. (2006) (henceforth K06) to perform a detailed clustering analysis, segregated by limiting K-band luminosity in the range $18.5 < K_{vega} < 20$. They found that the clustering of sBzKs is strongly dependent on K-band luminosity. Hayashi et al. (2007) studied the clustering of sBzKs over a smaller area (180 arcmin^2) but to much greater depth ($K_{AB} < 23.2$) and confirmed this strong luminosity dependence. The sBzK population therefore appear to inhabit a range of halo masses, and are therefore likely to be the progenitors of a wide range of present day galaxies.

There have been fewer studies of *passive* galaxy clustering at high redshift, in part because of their relatively low surface density. K06 found early evidence that pBzKs are more strongly clustered than sBzKs, although the limited depth and area of this survey (320arcmin^2 to $K_{\rm Vega} = 20$) gives rise to significant statistical uncertainly, particularly given the relatively low surface density of pBzKs at these magnitudes (0.38 arcmin^{-2}). Blanc et al. (2008) measured the clustering and number counts of BzK-selected galaxies to the same depth as K06, but over a much wider area (0.71 deg^2 total between two fields). Their BzK clustering amplitudes are smaller than those of K06 but consistent at the 1σ level. Analysis of deeper survey data over a similar area is required to confirm these results, and we present such an analysis in this work.

2.2 Data set and sample selection

In this work the UDS DR1 release (Warren et al. 2007) is used, which reaches 5σ ($2''$) depths of $K_{AB} = 23.5$ and $J_{AB} = 23.7$. For details of the completeness estimation, image stacking, mosaicing and catalogue extraction procedures see Foucaud et al. (2007) and Almaini et al. (in preparation). In addition to these data, the deep B, V, R, i', z' imaging from Subaru Suprimecam was used, with limiting depths of $B_{AB} = 28.4,\ V_{AB} = 27.8,\ R_{AB} = 27.7,\ i'_{AB} = 27.7$ and $z'_{AB} = 26.7$ (see Chapter 1, Section 1.4 and Furusawa et al. 2008). The UDS and Subaru survey areas are not entirely coincident, which reduces the usable area to 0.63 deg^2 in this analysis.

K-band sources ($> 5\sigma$) from the UDS were used as the primary catalogue, with optical magnitudes extracted directly from the Subaru imaging data after careful matching of the Subaru and UDS astrometric frames. Aperture magnitudes were then extracted using a 2-arcsec diameter. This procedure was performed by Dr. Chris Simpson (Liverpool John-Moores University). Only sources outside the contaminated halos of bright stars were used. In addition, the long exposure times used during the Subaru observations cause many of the stars in the field to be saturated. These stars must be excluded as their colours are likely to deviate from the main stellar sequence. In this present analysis, and unlike Hartley et al. (2008), the stars saturated in the K-band were also excluded.

Saturated stars may be defined by inspection of the difference in magnitudes of a source in two different aperture diameters. Figure 2.2 shows the K-band $2'' - 3''$ against $2''$ aperture magnitudes. The tight track at the bright end of Figure 2.2 is the stellar track. Saturation becomes important beyond the point at which this track turns upward, approximately $K < 12$ (Vega) in Figure 2.2. A conservative limit for saturation was taken in each band.

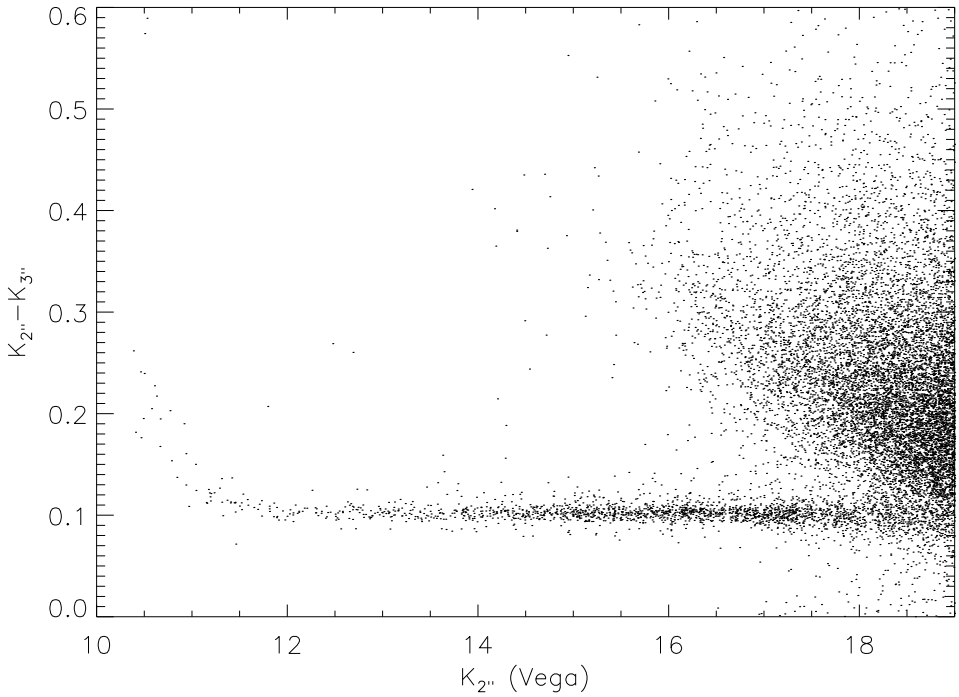

Figure 2.2. The difference in K-band magnitudes using $2''$ and $3''$ apertures. This plot provides a robust estimate of the point at which saturation effects become important, roughly $K < 12$ (Vega) in this case. A diagram of this sort also provides a reasonably clean way to separate bright ($K < 18$) stellar objects from the rest of the sample.

2.2.1 Construction of the BzK colour-colour diagram

The original BzK selection technique was introduced in Daddi et al. (2004) based on photometry in the Bessel B-band, Gunn z-band and K_s-bands as defined on the FORS1, FORS2 and ISAAC instruments at the Very Large Telescope (VLT). They published the colours of the stars in their field (the K20 field, Cimatti et al. 2002) with the intention that they could be used to replicate their selection using other filter sets. In Hartley et al. (2008) we shifted our colours to these filters by using the stellar track of K06. K06 had performed a detailed colour correction to Daddi et al. (2004) using published stellar spectra from Lejeune et al. (1997) and Pickles (1998). The colours computed by integrating the two filter response sets over these spectra should provide a translation between the two data sets.

As both the K06 and UDS data sets produce clear stellar loci with an obvious 'knee', matching the two should provide the same correction to the Daddi et al. (2004) colours for our filter set. This was the approach taken in Hartley et al. (2008). However, it has since become apparent that our implementation was incorrect. Stars saturated in the K-band were not excluded, resulting in a redder $z - K$ colour for the fit to the main branch of the stellar locus. It should also be noted that the assumption of a constant colour shift in the $(B - z)$ vs $(z - K)$ plane is incorrect. In general the required translation from one filter set to another depends on the SED of the source. For these reasons the selection of pBzKs and sBzKs here differ slightly from those of Hartley et al. (2008).

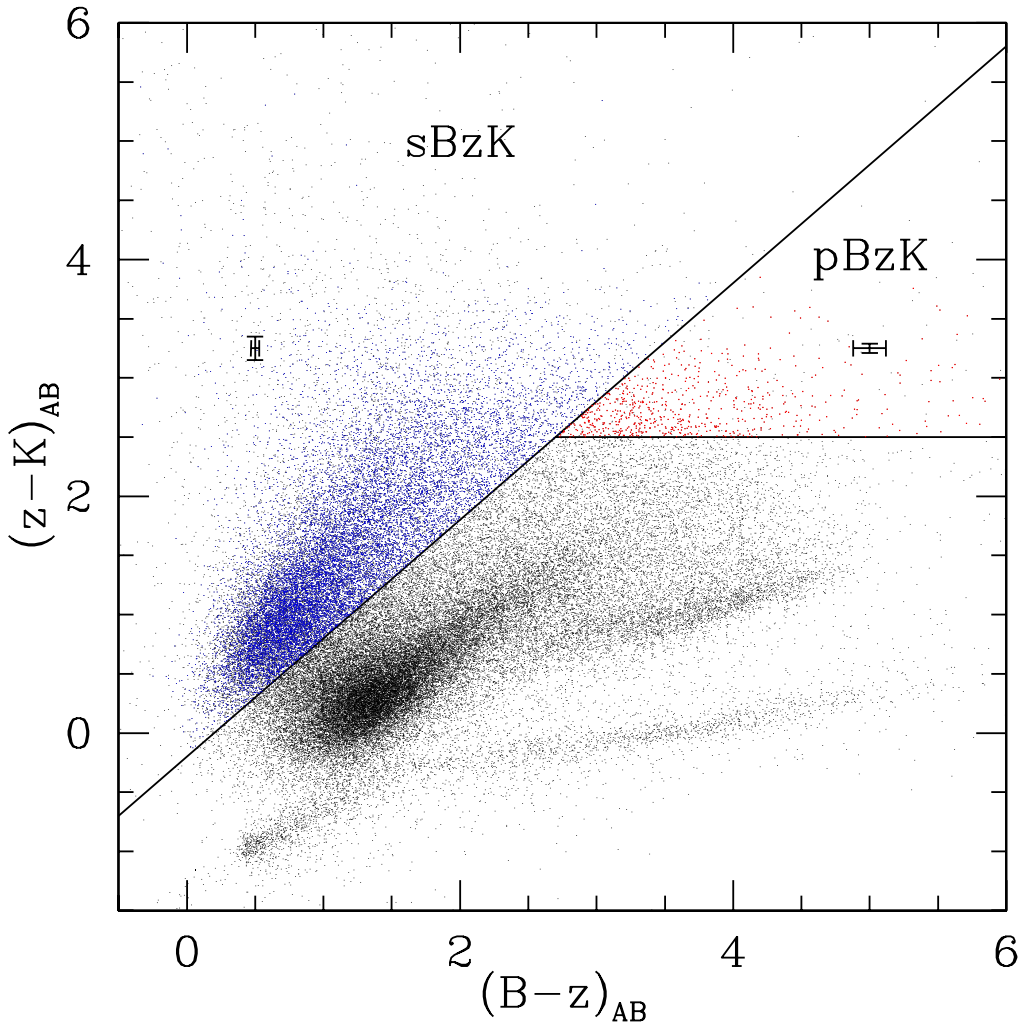

Figure 2.3. BzK colour-colour diagram for sources in the UDS DR1. The sBzK and pBzK selection regions are marked accordingly. Also visible are the passive galaxy track identified in Lane et al. (2007), and the stellar locus. The error bars shown are the mean error-bars for the galaxies in each of the BzK selection regions.

In Figure 2.1 evolutionary tracks with redshift in $B-z$ and $z-K$ colours are shown for the SEDs of Coleman et al. (1980) and various synthetic templates using Bruzual & Charlot (2003) stellar population synthesis models. The Coleman et al. (1980) templates have been extended into both the ultra-violet and infrared, again using Bruzual & Charlot (2003) models (*Hyperz* user manual). For each template, both UDS and Daddi $B-z$ and $z-K$ colours were computed. In this way SED-dependent colour shifts in the $(B-z)$ vs $(z-K)$ plane are sampled, albeit non-uniformly. For each source in the UDS DR1 sample described above, colour shifts were computed as the mean of the template-derived colour shifts within a cell of side 1 magnitude in colour, centred on the source. Contrary to the colour-corrections given in Hartley et al. (2008), where there is a shift of ~ 0.2 in $z-K$ colour, the correct $z-K$ shift is very small (< 0.1 mags), while that in $B-z$ is substantial (~ 0.27 mags).

With these corrected colours, the standard BzK definitions of Daddi et al. (2004) were used to construct several BzK samples. All K-band sources with $K_{AB} < 23.5$ were used for BzK selection unless both the B and z'-band magnitudes were fainter than the 3σ limits measured on those images ($\sim 1.5\%$ of sources outside of contaminated regions), since these cannot be constrained within the BzK plane. Of these, 230 objects had no detection in either band and each of these objects was visually inspected and classified as spurious (148 objects) or possibly a real source (82 objects). Figure 2.4 shows how the sample completeness is affected by the limiting z'-band magnitude. The z' data are deep enough that the samples are highly complete at all magnitudes, but there is a small incompleteness introduced at $K > 22.5$. To the limit of $K < 23.5$ that is imposed, the sample becomes incomplete in $z'-K$ above $\simeq 2.94$.

The samples for which statistical properties are computed are described below.

1. Samples of sBzKs and pBzKs using the Daddi-shifted colours defined above. These samples contain 19098 and 1990 objects with median photometric redshifts $z = 1.77$ and 1.66 respectively.

2. A pBzK sample using the shifted colours and a selection criterion $z-K > 2.4$. The original selection was based upon very few passive objects and it is prudent to test how robust the results are under slight a variation of the selection criteria. In particular the $z-K$ criterion is a concern, as for two of the synthetic templates, part of the intended redshift range ($1.4 < z < 2.5$) is below the traditional selection limit (Figure 2.1). This sample selects 2396 objects at median photometric redshift $z = 1.59$.

3. An additional pBzK sample as above, but including all objects with z and B band magnitudes greater than the 3σ limits and those without optical detections which were classified as 'possibly real' as described above. This sample includes the objects that have uncertain photometry and those that would otherwise be excluded due to incompleteness in the optical bands. This sample constitutes a 'worst case' sample. It contains 2790 objects with median photometric redshift $z = 1.85$.

4. An sBzK and pBzK sample defined for galaxies using the un-shifted UDS colours. These samples select 21821 and 1389 objects, with median photometric redshifts $z = 1.64$ and 1.67 respectively and provide a useful check as to how sensitive the sample properties are to the exact definitions.

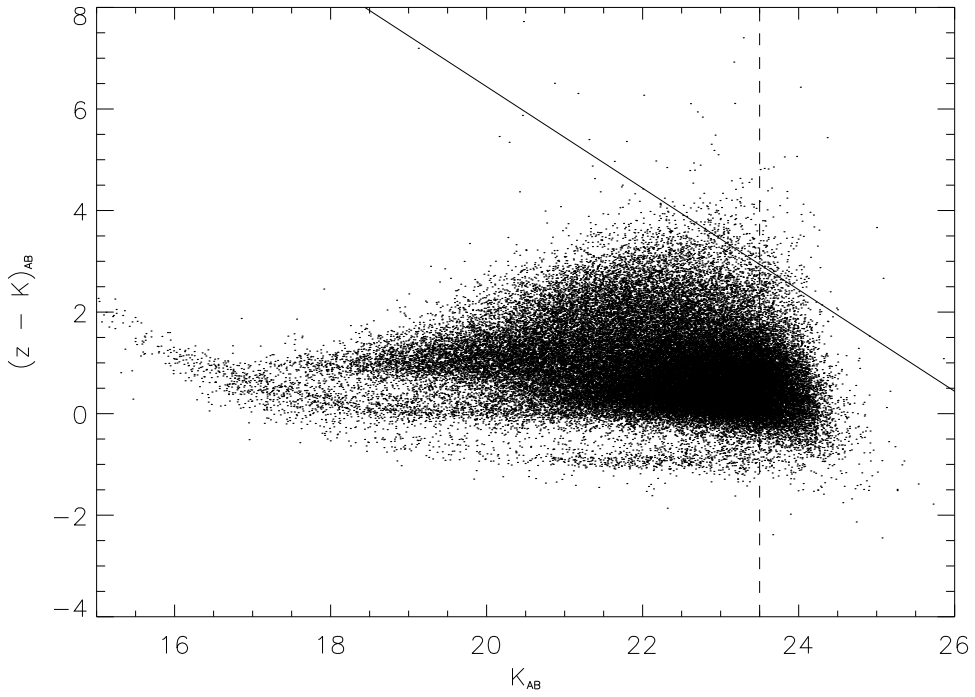

Figure 2.4. $(z-K)_{AB}$ colour versus K-band magnitude. The dashed line is our completeness limit of $K_{AB} = 23.5$ for the full sample, while the solid line shows how the completeness limit in z affects $(z-K)$ colour. We begin to be incomplete above $(z-K) > 2.94$. The impact of this incompleteness is discussed in the text.

2.3 Statistical properties of BzK galaxies

In this section I present the number counts and clustering of pBzKs and sBzKs to compare with and extend beyond previous measurements. Beginning with number counts I will show that the turn-over expected by Lane et al. (2007) (henceforth L07) is indeed likely, but that slightly deeper data still are required to unequivocally confirm its existence. Following this, I compute angular clustering measurements and de-project them to find the real space correlation length, r_0, finding that the passive objects are significantly more strongly clustered than star-forming objects.

2.3.1 Number counts

The differential K-band number counts are illustrated in Figure 2.5, and tabulated in Table 1, for both the full sample of galaxies and those selected as passive and star-forming BzKs. There is a steep rise in the counts of star-forming sBzKs toward fainter magnitudes. In contrast, pBzK number counts exhibit an apparent flattening at $K \sim 21$, consistent with the findings of Kong et al. (2006) and L07. Since these galaxies are sampled over a relatively narrow redshift range, this may imply that the passive population consists largely of luminous galaxies at this early epoch (see Section 4).

To investigate the reality of the turn-over, it should be noted that the galaxies around and beyond this feature have very faint B-band magnitudes, with a sub-

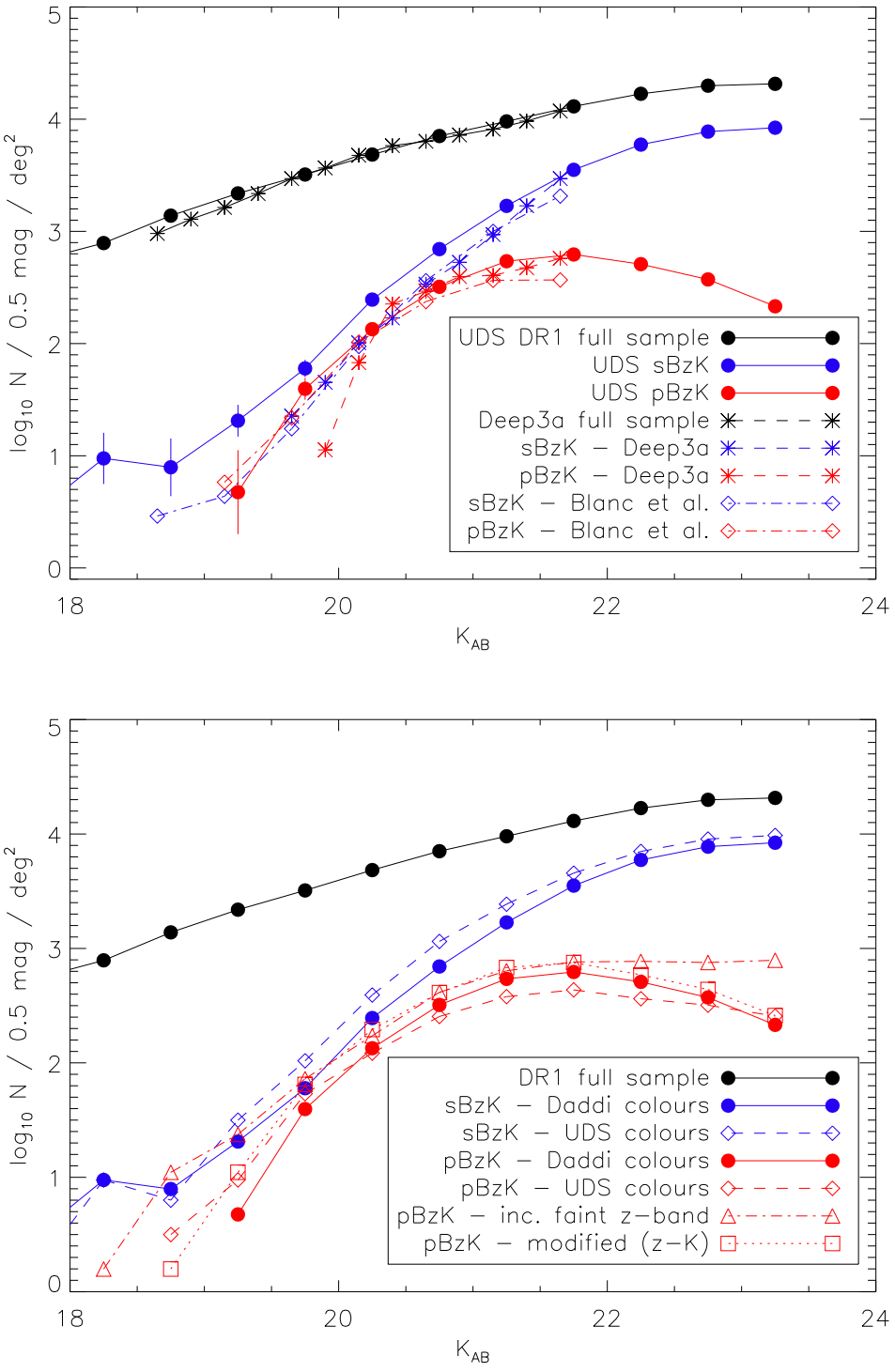

Figure 2.5. Differential number counts of passive and star-forming BzK galaxies with apparent total K-band magnitude, compared with the full-sample of K-selected galaxies. The error-bars shown are standard Poisson errors on the counts. The upper panel compares the UDS number counts, with colours corrected to those of Daddi et al. (2004), alongside some values from the literature (Kong et al. 2006; Blanc et al. 2008). The plateau identified in previous works (K06, L07, Blanc et al. 2008) and subsequent turn-over prior to our magnitude limit are apparent in the pBzK number counts. The lower panel shows number counts for the various samples defined in section 2.2.1. Numerical values are reproduced in table 2.1.

Table 2.1. Differential Number Counts in log(N/deg^2/0.5mag) bins for sBzK and pBzK in the UDS DR1.

K bin centre	all sources	±	sBzK	±	pBzK	±	Worst case pBzK	±
14.25	0.98	$^{+0.12}_{-0.17}$	-	-	-	-	-	-
14.75	1.15	$^{+0.10}_{-0.14}$	-	-	-	-	-	-
15.25	1.60	$^{+0.06}_{-0.07}$	-	-	-	-	-	-
15.75	1.80	$^{+0.05}_{-0.06}$	-	-	-	-	-	-
16.25	2.11	$^{+0.04}_{-0.04}$	-	-	-	-	-	-
16.75	2.40	$^{+0.03}_{-0.03}$	-	-	-	-	-	-
17.25	2.55	$^{+0.02}_{-0.02}$	0.2	$^{+0.3}_{-0.2}$	-	-	-	-
17.75	2.74	$^{+0.02}_{-0.02}$	0.5	$^{+0.2}_{-0.4}$	-	-	-	-
18.25	2.90	$^{+0.02}_{-0.02}$	1.0	$^{+0.1}_{-0.2}$	-	-	0.2	$^{+0.3}_{-0.2}$
18.75	3.14	$^{+0.01}_{-0.01}$	0.9	$^{+0.1}_{-0.2}$	-	-	1.0	$^{+0.1}_{-0.2}$
19.25	3.338	$^{+0.009}_{-0.009}$	1.31	$^{+0.09}_{-0.1}$	0.7	$^{+0.2}_{-0.3}$	1.37	$^{+0.08}_{-0.1}$
19.75	3.507	$^{+0.007}_{-0.008}$	1.78	$^{+0.05}_{-0.06}$	1.60	$^{+0.06}_{-0.07}$	1.86	$^{+0.05}_{-0.05}$
20.25	3.684	$^{+0.006}_{-0.006}$	2.39	$^{+0.03}_{-0.03}$	2.13	$^{+0.04}_{-0.04}$	2.24	$^{+0.03}_{-0.03}$
20.75	3.850	$^{+0.005}_{-0.005}$	2.84	$^{+0.02}_{-0..02}$	2.51	$^{+0.02}_{-0.02}$	2.62	$^{+0.02}_{-0.02}$
21.25	3.980	$^{+0.004}_{-0.004}$	3.23	$^{+0.01}_{-0.01}$	2.73	$^{+0.02}_{-0.02}$	2.80	$^{+0.02}_{-0.02}$
21.75	4.114	$^{+0.004}_{-0.004}$	3.549	$^{+0.007}_{-0.007}$	2.79	$^{+0.02}_{-0.02}$	2.88	$^{+0.02}_{-0.02}$
22.25	4.226	$^{+0.003}_{-0.003}$	3.773	$^{+0.006}_{-0.006}$	2.71	$^{+0.02}_{-0.02}$	2.89	$^{+0.02}_{-0.02}$
22.75	4.299	$^{+0.003}_{-0.003}$	3.889	$^{+0.005}_{-0.005}$	2.57	$^{+0.02}_{-0.02}$	2.88	$^{+0.02}_{-0.02}$
23.25	4.315	$^{+0.003}_{-0.003}$	3.924	$^{+0.005}_{-0.005}$	2.33	$^{+0.03}_{-0.03}$	2.90	$^{+0.02}_{-0.02}$

stantial fraction (62%) below the 3σ detection limit in this band ($B_{lim} = 28.4$). A non-detection in the B-band does not affect their classification, however, and merely pushes them red-ward in $B - z$ colour in the BzK diagram illustrated in Figure 2.1. The more worrying class are the objects which are below the 3σ limit in z' *and* B, which cannot be assigned BzK classification (164 objects). Figure 2.4 shows that even within the K-band limit, the samples become incomplete above $(z - K) \simeq 3$. The effect of this incompleteness is studied by considering the extreme case in which all of the objects with B and z'-band magnitudes fainter than the 3σ limits (28.4 and 26.7 respectively) are pBzKs. In this 'worst case' sample (sample 3 of section 2.2.1) the number counts still exhibit a plateau, as before, but with the absence of a turn-over.

At the conservative limit of $K_{AB} < 23$ the completeness should be $> 95\%$ for compact galaxies (Almaini et al., in prep.), so the feature is unlikely to be due to K-band incompleteness unless there is unprecedented size-evolution in the pBzK population towards fainter magnitudes compared to the general population. The extreme compactness of passive galaxies at these redshifts observed by Cimatti et al. (2008) suggests that of the two classes, the sBzKs should suffer more from such incompleteness. There is no evidence from their number counts to suggest that they are adversely affected in this way, so we expect that the pBzKs are likewise unaffected below this magnitude.

Photometric errors are another source of concern. Adjusting the boundaries for pBzK selection by the mean photometric errors for sources close to the boundaries (0.1 magnitudes), it can be seen that such errors had no noticeable impact on the decline in the pBzK population. Additionally, since the pBzK region of the colour-

colour plane is sparsely populated, photometric errors are likely to scatter fainter objects *into* pBzK selection, rather than the opposite. The faint pBzK counts are thus likely to be slightly *overestimated* because of photometric errors.

The conclusion is that the flattening in the pBzK population is likely to be real and the subsequent turn-over is probable but not certain. The turnover corresponds to an absolute magnitude $M_K = -23.6$ at $z = 1.4$, which is close to the value of L^* determined for the K-band luminosity function at these redshifts (Cirasuolo et al. 2007). Taken at face value, these results are suggestive of a sharp decline in the number density of pBzK galaxies towards lower luminosities, consistent with expectation from downsizing. As correctly pointed out in Grazian et al. (2007) and predicted by Daddi et al. (2004), this is not necessarily equivalent to a true decline in the number density of passive galaxies, since the efficiency and completeness of the pBzK technique is largely untested at such faint magnitudes. In particular, there is some evidence that passive galaxies can also be found among the redder sBzK galaxies (at large values of $z - K$), with a possible incompleteness as high as 34% Grazian et al. (2007). Randomly adding an additional 34% of galaxies to the pBzK sample from this part of the diagram is indeed sufficient to remove the apparent turn-over, since these are predominantly the fainter sBzK galaxies.

The sBzK number counts are significantly higher than those of K06, Blanc et al. (2008) and Imai et al. (2008) (not plotted) at $K < 21$; while the pBzK number counts are very similar. The UDS field is more than 6 times larger than that of the K20 survey and 4 times larger than the one used in Imai et al. (2008) and should therefore provide a better estimate of the bright end of the number counts. However, the combined field used in Blanc et al. (2008) is of a similar size to the UDS. Cosmic variance is a possible explanation for the difference in number counts in this case, but there may remain some differences due to the colour transformations and differing photometric choices.

2.3.2 Clustering

The sBzK and pBzK samples were selected to a limit of $K_{AB} < 23$. This conservative magnitude limit was adopted to ensure the minimum contamination by spurious sources. Regions of the survey area which contained the bleed regions from bright stars, obvious cross-talk artifacts and noisy edges were masked out. The correlation function was then computed using the Landy & Szalay (1993) estimator, introduced in Chapter 1, Section 1.3.1:

$$w(\theta) = \frac{DD - 2DR + RR}{RR}, \tag{2.1}$$

using 10 times more randoms than data, which were restricted to the same unmasked region as the data. The number of randoms is required to be vastly more than the data to ensure that the shot noise error arising from them is negligible. The errors were computed by 100 bootstrap resamplings and an estimate of the cosmic sample variance (see section 2.5.1), added in quadrature. Poisson errors are a relatively poor estimate of the variance in the data, while bootstrap error estimates are much closer to the true values, especially at small scales (e.g. Norberg et al. 2009). The errors found from the bootsrap analysis are 2-3 times greater than Poisson errors, though the sample variance is the dominant source of error in these measurements (approximately a factor 2 greater than the bootstrap estimates). The resulting $w(\theta)$ values were then corrected for the integral constraint by an additive constant,

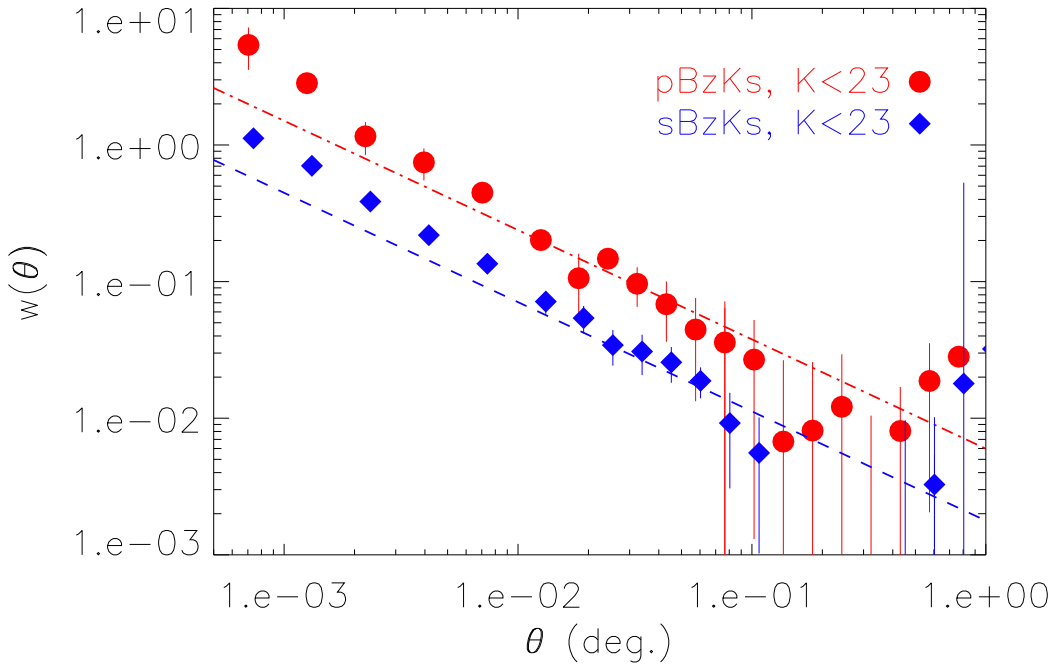

Figure 2.6. The angular correlation function for the BzK-selected galaxy samples. The best-fitting power laws are shown, with slopes fixed to the fiducial value of $\delta = 0.8$. The pBzK galaxies are very strongly clustered, much more so than sBzKs, indicating that they occupy the most massive dark matter halos at their epoch.

$$C = \frac{\Sigma RR(\theta).\theta^{-\delta}}{\Sigma RR(\theta)}. \tag{2.2}$$

A single power law of the form $w(\theta) = A.\theta^{-\delta}$ was fit to the corrected data over the separation range $0.01 - 0.2$ degrees, fixing the slope to the fiducial $\delta = 0.8$ and minimising the χ^2. The errors on the measurements were found by splitting the field into 4 and computing the variance of $w(\theta)$ (see below). In order to fit the clustering reliably, we wish to avoid small-scale excesses due to multiple galaxy occupation of a single halo. The lower bound was chosen to correspond to ~ 0.9Mpc (co-moving) at z$\sim$ 2 for this reason. The upper bound is a conservative estimate of the limit to which our data have enough signal for a reliable fit to be obtained.

For sBzK galaxies the amplitude was found to be $A = 1.78^{+0.05}_{-0.05} \times 10^{-3}$ (deg.$^{0.8}$) and for the pBzK population $A = 5.99^{+0.52}_{-0.48} \times 10^{-3}$ (deg.$^{0.8}$). The clustering amplitude of the pBzKs is inconsistent with the clustering of sBzK galaxies at the 3-sigma level. Figure 2.6 shows the clustering measurements corrected for the integral constraint for the sBzK and pBzK galaxies.

The correlation function of each sample, but the star-forming sample in particular, exhibit a deviation from power-law behaviour at small scales. Though it is not relevant to the aims of the present work (other than as a possible source of error on the power-law fit), this departure contains information about how many galaxies are hosted by a typical dark matter halo. This is a point of interest in its own right that

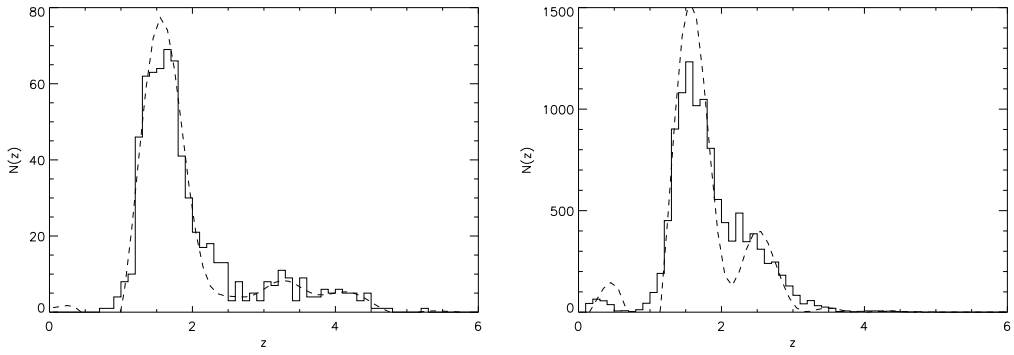

Figure 2.7. Photometric redshift distributions for passive (left) and star-forming (right) BzK-selected galaxies (solid line histogram). The over-plotted dashed lines are distributions with photometric errors de-convolved using a Weiner filter (see section 2.5.3).

shall be covered more extensively later in this thesis (see Chapter 5 Section 5.1.3).

2.3.3 De-projection

The real space clustering and projected clustering are linked by the relativistic Limber equation (Limber 1954). If the redshift distribution of a sample is known, the Limber equation can be inverted and the correlation length, r_0, can be calculated in a robust manner (Peebles 1980; Magliocchetti & Maddox 1999)

To estimate the redshift distribution we use photometric redshifts based on the Subaru and UDS bands previously introduced, with the addition of Spitzer data taken as part of the SWIRE survey (Lonsdale 2003). The method is described fully in Cirasuolo et al. (2007) and an overview of the technique given in Chapter 1 Section 1.3.3. It consists of minimising the χ^2 of a galaxy's photometry over synthetic galaxy templates. The resulting redshift distributions are shown in figure 2.7. The photometric redshift distribution was then used directly in the inverted Limber's equation during the calculation of r_0. The values obtained in this manner are $r_0 = 16.0^{+0.8}_{-0.7}h^{-1}$ Mpc and $9.1^{+0.6}_{-0.6}h^{-1}$ Mpc for $K < 23$ pBzKs and sBzKs respectively. The quoted errors are due to the error in the fit to the clustering amplitude which are propagated through the Limber's inversion and therefore take into account the shot noise and cosmic variance.

The photometric redshifts are subject to errors ($\sigma/(1+z) = 0.095$ and 0.105 for pBzKs and sBzKs respectively), however, and assuming that the true redshift distribution is highly peaked, these errors are likely to broaden the measured distribution. A broader distribution will result in a larger inferred clustering scale length. It is therefore important that we take such errors into account. This is done so by assuming the errors are Gaussian, and then deconvolve the errors from the redshift distribution using a Fourier-based Wiener filter. A more detailed description of this method is provided in section 2.5.3. The resulting de-convolved redshift distributions are shown in figure 2.7. The scaling lengths recovered using the corrected redshift distribution are as follows: $15.2^{+0.7}_{-0.7}h^{-1}$ Mpc and $8.4^{+0.5}_{-0.6}h^{-1}$ Mpc for pBzKs and sBzKs respectively.

It should be noted, however, that the photometric redshift code used in this work is largely untested for pBzK and sBzK galaxies. These results will be refined in the

Sample	A_{fit}	A+1σ	A−1σ	r_0	$r_0 + 1\sigma$	$r_0 - 1\sigma$	$\bar{z}_{err}$
$pBzK_{UDS}$	5.02e-3	5.77e-3	4.26e-3	14.3	15.4	13.0	0.222
$pBzK_{Daddi}$	5.99e-3	6.50e-3	5.51e-3	15.2	15.9	14.5	0.237
$pBzK_{z-k>2.4}$	5.36e-3	5.93e-3	4.84e-3	13.5	14.3	12.8	0.229
$pBzK_{worstcase}$	5.61e-3	6.10e-3	5.07e-3	16.8	17.6	15.9	0.217
$sBzK_{UDS}$	2.00e-3	2.23e-3	1.77e-3	8.0	8.5	7.5	0.266
$sBzK_{Daddi}$	1.78e-3	1.82e-3	1.74e-3	8.4	8.9	7.8	0.274

Table 2.2. Clustering amplitudes and de-projected correlation lengths for the pBzK and sBzK samples described in section 2.2.1.

$K_{AB,lim}$	N	r_0	$r_0 + 1\sigma$	$r_0 - 1\sigma$
23.0	13395	8.4	8.9	7.8
22.5	8237	8.7	9.5	7.9
22.0	4303	9.1	10.1	8.0
21.5	1964	14	16	12
21.0	829	23	26	21
20.5	336	29	35	23
20.0	155	77	86	67

Table 2.3. Values of r_0 for sBzKs in the UDS by limiting K_{AB}-band magnitude along with the number of objects brighter than the relevant limit. As for the full samples, the error ranges are derived from estimates of the shot noise and cosmic variance.

future by the use of an ongoing ESO Large Programme using VIMOS and FORS2 on the VLT to target one sixth of the UDS DR1 galaxies with photometric redshifts > 1. The clustering amplitudes and derived de-projected correlation lengths for each of the samples defined in Section 2.2.1 are given in Table 2.2.

In addition to the relative clustering of star-forming and passive BzK-selected galaxies, the dependence of limiting K-band magnitude on the clustering of the star-forming galaxies can also be investigated. Figure 2.8 and Table 2.3 show the values for sBzKs with varying limiting magnitude. These results confirm previous claims for a dependence of sBzK clustering on apparent magnitude (Hayashi et al. 2007). The dependence is much stronger at magnitudes of $K < 22$, indicating a strong correlation between halo mass and sBzK magnitude for these objects. In each case the redshift distribution appropriate for the sample was used during deprojection as these can change greatly between the brightest and faintests subsamples. This change in redshift range sampled may also contribute to the strong evolution seen at brighter magnitudes. Such effects are tackled in the following chapter.

2.4 BzKs within $1.4 < z_{phot} < 2.5$

The BzK selection was defined by Daddi et al. (2004) to isolate galaxies in the redshift range $1.4 < z_{phot} < 2.5$. Clearly from figure 2.7, and as expected by Daddi et al. (2004), there are contaminating objects from outside of this range. In particular, the selection is only tested for galaxies with $K < 20$ (Vega). The deeper sample used in this work may therefore contain significant numbers of galaxies of some other population. Using the photometric redshift information it is possible to assess how successful the BzK selection technique is in reproducing the clustering of

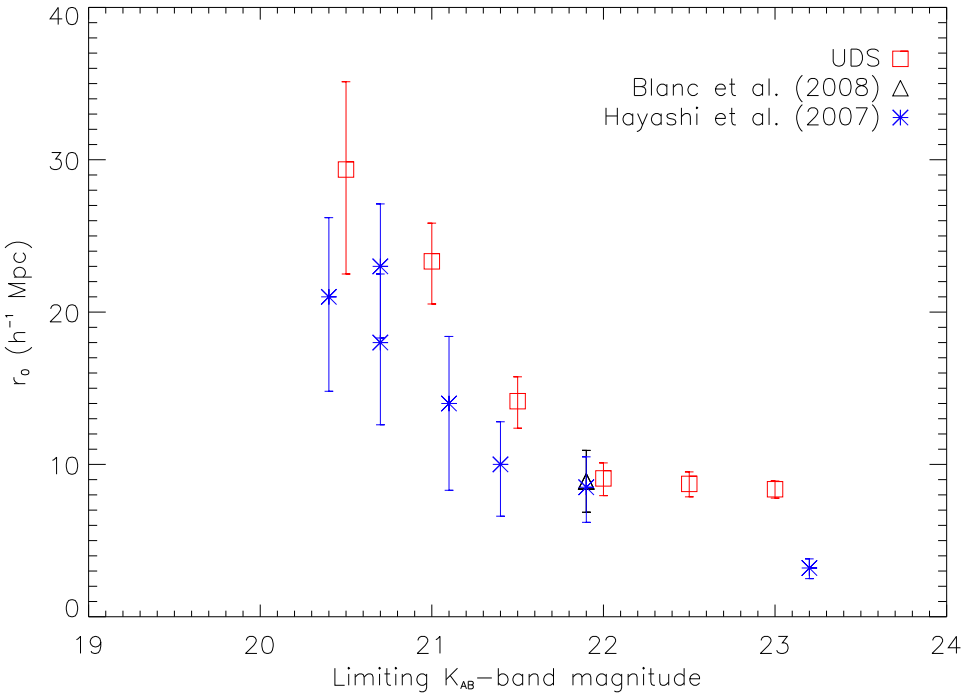

Figure 2.8. The dependence of clustering strength of sBzK-selected galaxies on limiting K_{AB}-band magnitude. The measurements from this work (open squares) are shown together with literature values from Blanc et al. (2008) and Hayashi et al. (2007) (open triangle and asterisks respectively). The values presented here confirm the magnitude dependence of sBzK clustering strength.

objects within the desired range.

Following the same method outlined for the full samples, clustering amplitudes and correlation lengths for pBzKs and sBzKs which have $1.4 < z_{phot} < 2.5$ were computed. The angular correlation functions for these samples are shown in Figure 2.9. The correlation lengths were found to be $11.8^{+1.0}_{-1.0}h^{-1}$ Mpc and $7.0^{+0.5}_{-0.5}h^{-1}$Mpc for pBzKs and sBzKs respectively. These values are slightly lower than those for the full samples and indicate that the high redshift tails in the full sample are at least as highly clustered as those within the $1.4 < z_{phot} < 2.5$ range. However, the conclusions drawn from the full sample are still valid, namely that the pBzK galaxies are significantly more strongly clustered than the sBzK galaxies.

2.5 Error budget

Robust estimation of errors is always highly important, especially in clustering measurements. Fortunately, uncertainties in clustering statistics have been studied extensively (e.g. Hamilton 1993; Landy & Szalay 1993; Bernstein 1994; Simon 2007). However, this is not to say that they are straightforward. In this section two of the more problematic sources of error are described, together with the strategies employed to tackle them.

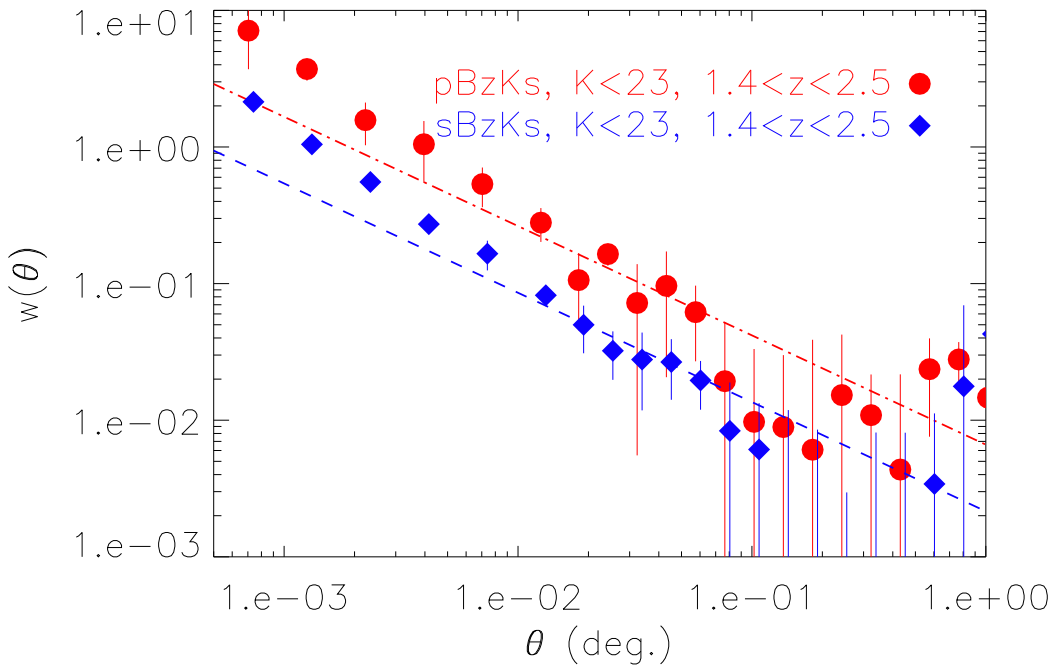

Figure 2.9. The angular correlation function for BzK-selected galaxy samples within $1.4 < z_{phot} < 2.5$. The symbols and lines have the same meanings as those in 2.6.

2.5.1 Cosmic Sampling Variance

The term 'cosmic variance' (or cosmic sampling variance) refers to the fact that our observable Universe is only a sub-volume of the larger Universe. The power spectrum of density fluctuations includes modes with scales comparable to the size of the observable Universe. Although the amplitudes of these long modes are small, there are possible biases in statistics such as number counts and clustering that cannot be accounted for, irrespective of the survey volume. However, the term is often also used to refer to the sample variance within the observable Universe, but on scales comparable to and larger than the survey volume. It is in this sense that the term is used here.

While the UKIDSS UDS is currently the widest survey of its depth, the effects of cosmic sampling variance may still be problematic. Considering the pBzK sample used here in particular, the clustering scale length, $r_0 = 15.2\mathrm{h}^{-1}$Mpc, is more than a third of the angular extent of the survey at the median redshift of the sample. Taking account of cosmic sampling variance is highly important for such a sample.

In order to do this we would ideally require additional fields of similar areas and depths as the UDS. Unfortunately, as previously stated, the UDS is currently unique. We can, however, obtain an estimate by splitting the UDS field into four quadrants and computing the clustering for each quadrant separately. This would provide us with an estimate of the variance for a field one quarter the size of the UDS field. There is an additional caveat to this procedure in that the four sub-fields are connected and not entirely independent. However, as the variance is sensitive to the field size (Somerville et al. 2004), this method is likely to provide a slight

over-estimate of the variance across fields of the size of the UDS.

Finally, it is important to note that the error estimate derived from this sampling variance should include the shot-noise error. However, as the field is only being split into four, the variance estimate will not be greatly reliable. To ensure that the errors are indeed over-estimated rather than under-estimated, the sample variance is added in quadrature with the results of a bootstrap resampling in the present analysis.

2.5.2 Redshift errors

While the errors in pair counting can be significant, they are at least problems which can be dealt with in a fairly standard and robust way. However, there is a further source of uncertainly in this work which is extremely difficult to treat, and even more difficult to treat in a way which is consistent with previous work. As an avenue of study, the photometric redshift technique is still in relative infancy. Many of the biases and systematics are now known, but there is yet to be a standard proven way to correct for them in a study such as the clustering of galaxies.

In this present work, the central result does not rely on the redshift determination for either the classification into samples or clustering measurement. However, the redshift information is of great value in interpreting the result. With this additional information it is possible to recover the real-space correlation length and obtain an estimate of the typical mass of the host dark matter halos. We therefore require an understanding of the possible effects that uncertainties in the redshifts might have. The photometric redshift technique is used when we do not have spectral information such as emission or absorption lines. The discrete photometric data points of a sample galaxy are therefore fit to the continuum of a template galaxy (see Chapter 1, Section 1.3.3). Clearly the technique will be more successful if the sample galaxy in question has a strong feature such as a clear 4000Å break. Such galaxies are identified with comprising an old stellar population and hence will likely be in the pBzK sample if they are at a suitable redshift.

However, variation in the SEDs of sample galaxies and the finite width of the filters used to observe them allow a certain degree of tolerance in the acceptable solutions for a galaxy's redshift. In the case of an old, burst-like galaxy SED it is reasonable to approximate this uncertainty as a Gaussian error, with width given by the redshifts with $\Delta\chi^2 = 1$. For galaxies without strong features, such as the majority of star-forming galaxies, the situation can be significantly poorer.

In addition to this uncertainty in the redshift solution around a minimum, there is a further possible error caused by a galaxy having two minima, i.e. a secondary solution. Colloquially, the cases where the true redshift is very different from the photometric solution are called 'catastrophic errors'. These can occur when the secondary solution is correct, or if none of the preferred solutions are correct. The number of such catastrophic outliers is minimised by using a large number of filters and by using appropriate templates and training sets. Unfortunately there are currently relatively few galaxies with spectroscopic redshifts at $z > 1$ in the UDS field. This particular limitation must therefore be kept in mind when considering the interpretation of these results.

2.5.3 Weiner filter

In this section the Fourier method used to deconvolve photometric redshift errors from the n(z) distribution is described. The errors in the photometric redshift dis-

tribution are assumed to be Gaussian and it is also assumed that the measured distribution is simply the convolution of these errors with the true redshift distribution. Under these assumptions it should be possible to deconvolve the errors from the measured distribution, using the fact that a convolution is simply a multiplication in the Fourier domain. Dividing the Fourier transform of the measured distribution by that of the Gaussian errors, should provide an estimate of the true distribution. Given the discrete nature of the redshift distribution, a discrete Fourier transform is preferred. The terms are calculated in the standard way, with n(z) reflected such that only the real part remains:

$$\begin{aligned} a_0 &= \tfrac{1}{L}\int_{-L}^{L} n(z)dz \\ a_i &= \tfrac{1}{L}\int_{-L}^{L} n(z)\cos\left(\tfrac{i\pi z}{L}\right)dz \end{aligned} \tag{2.3}$$

where L is the upper limit of the redshift distribution. The discrete Fourier transform of the true redshift distribution should be each term of the transform of the redshift distribution, divided by the corresponding term of the transform of the redshift errors:

$$a_{i,true} = \frac{a_{i,data}}{a_{i,error}} \tag{2.4}$$

In practice, a relatively large number of terms in the discrete Fourier transform are required to reproduce the redshift distribution accurately, and as the the Fourier transform of a Gaussian is also a Gaussian, small levels of high frequency noise can be amplified greatly to give a spurious result. One way to avoid such a problem is to use a Wiener filter:

$$W(k) = \frac{1}{H(k)} \times \left(\frac{H(k)^2}{H(k)^2 + \frac{1}{SNR}}\right). \tag{2.5}$$

Where H(k) is the Fourier transform of the Gaussian errors and SNR the signal to noise ratio. In the limit of SNR being infinite, this filter tends to $1/H(k)$ as in the simple deconvolution above. SNR is estimated by fitting the power spectrum of the redshift distribution to a function of the form $a \times 10^{-b.k} + c$, with c identified as being the noise level. The resultant redshift distributions are shown in Figure 2.7 and it is these which are used in the inversion of the Limber equation (section 2.3.3).

2.5.4 Incompleteness

The completeness of the sample at the magnitude cut is estimated to be $> 95\%$, which suggests that up to 5% of the intended galaxy sample may be missing. If these missing galaxies are a fair sample of the overall population then there would be no impact upon the results. In terms of DD and RR pair counts in any given bin of separation, each would be multiplied by a constant factor which would cancel out leaving a robust value for $w(\theta)$. However, as incompleteness is magnitude dependent, slightly deeper consideration is required.

The pBzK sample exhibits a probable turn-over in their number counts at faint magnitudes, and at the very least a plateau. Removing 5% of the overall faint galaxy sample will therefore have little impact upon the selection of these objects

when compared with the star-forming sample, where the number counts continue to rise at fainter magnitudes. Furthermore, it could well be expected that extended objects with relatively low surface brightness are the objects that would suffer the most from incompleteness. As passive galaxies are found to be compact at $z > 1$ (e.g. Zirm et al. 2007), it is the star-forming population that are of most concern.

In the low-redshift Universe it is observed that there is a mild luminosity dependence on the clustering strength of star-forming galaxies (Zehavi 2005). The redshift distribution of the sBzK sample is highly peaked, so if the sBzK sample truly is missing some of its fainter members, then we may expect the majority of these to be of lower intrinsic luminosity than the parent sample. Following this reasoning the recovered value for r_0 would be higher for the incomplete sample than that of the (unobtainable) complete sample. This source of error is therefore highly unlikely to be responsible for the difference in clustering observed for passive and star-forming galaxies. However, it is possible that this spurious increase in clustering strength could be responsible for the disagreement with the results of Hayashi et al. (2007) (figure 2.8). Quantitatively, Zehavi (2005) find that for a change in limiting luminosity of 0.5 magnitudes near L^*_R, the corresponding value for r_0 changes by $\sim 0.5\mathrm{h}^{-1}$Mpc (with a much weaker dependence at sub-L^*_R limiting luminosities). The effect of incompleteness in this case is likely to be much smaller than a change of 0.5 magnitudes in limiting luminosity and hence this value constitutes a strict upper limit. This source of error is insufficient to explain differences between the samples or the discrepancy with Hayashi et al. (2007), however robust determination of the uncertainties is both difficult and time consuming and as such is not included in the quoted error values.

A final concern is that of how the variation of incompleteness across the field may impact upon the results. As the UDS field is a mosaic of four pointings, each with four cameras, there is a non-negligible variation in the depth of the image. Any incompleteness in the samples is likely to affect the shallower regions more severely, a facet that is not taken into account by the placement of the random objects. The effect of this inhomogenous incompleteness would be to create an artifical void in the distribution of the faintest members of the sample, potentially increasing the clustering signal. However, removing a further 50% of the faintest ($22 < K < 23$) galaxies from the sBzK sample within a region of $\sim 1/16^{\mathrm{th}}$ of the survey area was only sufficient to boost the amplitude by a factor of ~ 1.25, comparable to the errors derived from the bootstrap analysis. The true distribution of missing galaxies is likely to be more evenly spread than this example, and the systematic error therefore smaller.

2.6 Discussion

Hartley et al. (2008) contained an analysis of the luminosity function of the BzK-selected galaxy samples. As stated at the beginning of this chapter, this analysis has not been re-performed with the updated selections. However, the samples used in this work are very similar to those used in the original analysis of Hartley et al. (2008). Furthermore, the number counts and clustering measurements produce consistent results. For illustration only, it is therefore appropriate to include the luminosity functions and associated analysis, performed by Dr Cirasuolo (Royal Observatory Edinburgh) and Dr Lane (University of Oxford). Following this section, an interpretation of the results is provided.

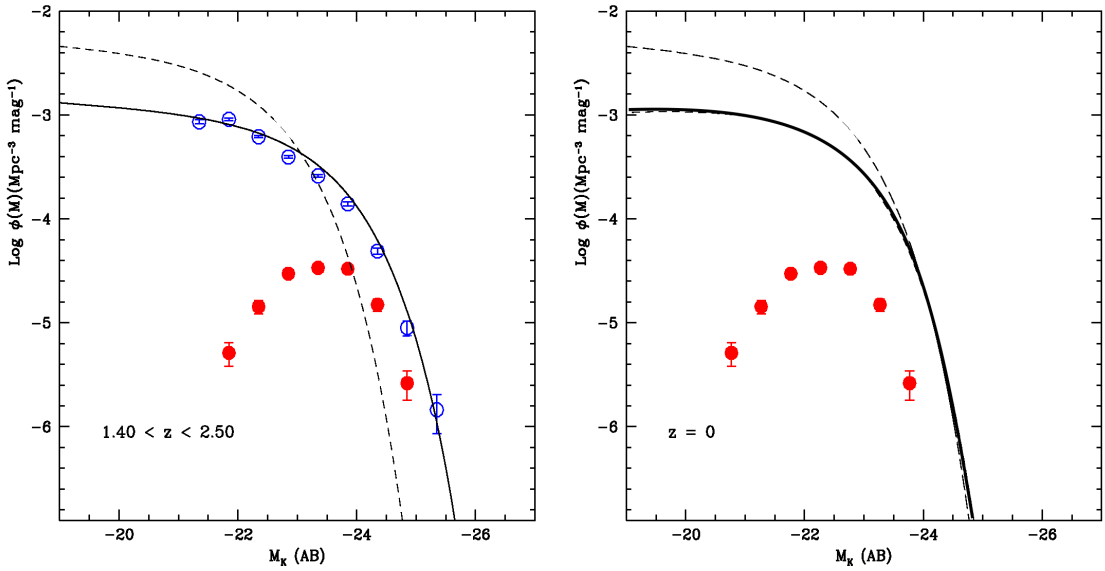

Figure 2.10. *Left*: Luminosity functions for pBzK and sBzK samples (filled and open circles respectively), plus all K-selected galaxies to a limit of $K < 23$ (restricted to galaxies with photometric redshifts in the range $1.4 < z < 2.5$, solid line). The sBzKs are representative of the overall population in the range $1.4 < z < 2.5$, while the pBzKs are found to be exclusively bright objects. Also shown is the $z = 0$ luminosity function for K-selected galaxies from Kochanek et al. (2001) (dashed line). *Right*: The dashed line shows the same luminosity function for $z = 0$ galaxies as in the left hand panel. The solid curve shows the $z = 0$ luminosity function for early type galaxies from Kochanek et al. (2001), while the points are the $z = 0$ luminosity function for pBzKs under the assumption of passive evolution (see text).

2.6.1 Luminosity function

A luminosity function was constructed for our BzK-selected galaxies in the same way as detailed in Cirasuolo et al. (2007), by using the $1/V_{max}$ method (Schmidt 1968). Figure 2.10 shows the luminosity function for the BzK galaxies with photometric redshifts in the range $1.4 < z < 2.5$ (points with error-bars), compared with all K-selected galaxies in the same range (solid line). Also plotted is the $z = 0$ luminosity function from Kochanek et al. (2001). It is clear that sBzK galaxies sample a wide range in luminosity, while the pBzK population are dominated by bright galaxies with $M_K > -23$. At this epoch, however, the bright end of the luminosity function is still dominated by star-forming objects, consistent with Cirasuolo et al. (2007) who found that the galaxy colour bimodality is weak at these redshifts. The pBzK galaxies are nevertheless likely to be among the most massive systems, since the mass-to-light ratio will be significantly lower for the actively star-forming systems.

Under the assumption that the pBzKs passively evolve to the present day (with minimal merging) one can estimate their contribution to the bright end of the present-day luminosity function. The implied evolution is modelled from $z \sim 1.60$ (the median value for pBzKs) to $z = 0$ assuming the spectral evolution models of Bruzual & Charlot (2003). For simplicity we chose a solar metalicity model and a Salpeter initial mass function (Salpeter 1955) and assume only passive evolution from initial formation bursts at redshifts $z_f = 3$ and $z_f = 10$. This results in a Johnson K-band absolute magnitude evolution in the range $0.96 < \Delta M_K < 1.21$.

Taking the mean value we can estimate the minimum number of present day bright ($K < -23.4$) galaxies that could previously have been pBzKs. Under our assumptions we find that 2.5% of such galaxies can be explained by passively evolved pBzKs. It is assumed that the remainder is made up of the descendants of bright sBzKs and merger remnants from within and between the two classes.

2.6.2 Interpretation

The number counts, clustering and luminosity function of galaxies at $z \sim 2$ selected using the BzK selection criteria have been presented. The pBzK galaxies show a marked flattening in their number counts which cannot be explained by the effects of incompleteness, and a probable turn over at faint magnitudes. Given the highly peaked nature of their redshift distribution, the number counts imply a lack of faint passive galaxies at their redshift. We should conclude then, at this epoch, it is generally luminous, massive galaxies that are undergoing passive evolution. This is consistent with the down-sizing scenario, in which the most massive galaxies are formed first and are the first to evolve onto the red sequence (Kodama et al. 2004).

The angular clustering of the passive galaxy sample is very strong, approximately 4 times the amplitude of the sBzK population. This is in part due to their relatively narrow redshift distribution (Kong et al. 2006), but a large difference remains after de-projection to the real-space correlation length. Correlation lengths, $15.2^{+0.7}_{-0.7}h^{-1}$Mpc and $8.4^{+0.5}_{-0.6}h^{-1}$Mpc were found for the pBzK and sBzK galaxies respectively. This value for the correlation length of sBzKs is almost twice that found by Hayashi et al. (2007), who used a sample with similar limiting magnitude but with smaller field size (180 arcmin^2, compared with ~ 2250 arcmin^2 for the UDS). The present sample consists of more than 10 times the number of sBzKs than that of Hayashi et al. (2007) and in addition the empirical redshift distribution used here makes fewer assumptions regarding their redshift distribution. Further data sets in independent fields reaching depths of $K_{AB,lim} \sim 23$ are required to fully account for the effects of cosmic sample variance. However, this discrepancy remains a puzzle at present.

This present analysis also confirms that the scale-length for the clustering of sBzK galaxies is dependent on apparent magnitude, consistent with the work of Hayashi et al. (2007) (figure 2.8). This dependence is far more significant at magnitudes below $K_{AB} \sim 22$ indicating a strong correlation between such galaxies and the mass of the hosting halo. In addition to this magnitude dependence there is a clear enhancement at small scales in the sBzK clustering. This enhancement is indicative of multiple sBzK galaxies occupying a single dark matter halo. The scale at which this turn-off occurs can provide an indication of the size, and hence mass, of the hosting dark matter halos. The turn-off occurs at ~ 0.01 deg, which corresponds to ~ 0.3 Mpc at z= 1.65. Halos of this size have masses between 10^{11} and 10^{12}M$_\odot$(Mo & White 2002). A fuller consideration of this enhancement within the halo occupation distribution framework will be presented later in this thesis. The host halo mass can also be derived from comparing the de-projected clustering scale length with models of dark matter halo clustering evolution. Using models based on Mo & White (2002) (see section 5.1.2) the correlation length implies a typical halo mass of $\sim 6\times10^{11}$M$_\odot$.

Applying the model of clustering evolution to the pBzKs we can qualitatively conclude that they are inhabitants of the most massive halos at their epoch (in excess of 10^{13}M$_\odot$); halos which will eventually become massive groups and clusters by the present day. The evolutionary path that pBzKs take with redshift will be

the subject of future work, using further colour selection techniques.

The conclusions based on the r_0 measurements are strengthened by their luminosity functions. Even under the strict assumption of purely passive evolution, the descendants of pBzKs occupy the bright end of the luminosity function. Such galaxies are group and cluster dominant galaxies in the local universe. The brightest ($K < 21$) sBzKs have clustering scale lengths comparable to, or greater than, that found for the pBzKs. This finding indicates that such galaxies will also become group and cluster dominant galaxies by $z = 0$. The implication is that we are indeed witnessing the epoch at which the build up of the red sequence begins.

2.7 Conclusions

In this chapter I have used a colour-colour selection technique which, despite only having been proposed in 2004, already has a body of literature utilising it. Using angular clustering measurements, de-projected to find real space correlation lengths, and galaxy number counts I have shown that passive galaxies occupy more massive dark matter halos than star-forming galaxies at their median redshift of $z \sim 1.6$. Together with dark matter halo evolution models, this result implies that pBzKs will evolve to become group and cluster-centre galaxies by $z = 0$.

However, the clustering of the sBzK sample is strongly dependent on the limiting K-band magnitude for $K < 22$. In fact, the brightest of the sBzKs are even more strongly clustered than the pBzKs. Inspection of the luminosity functions of pBzK and sBzK samples shows that pBzKs are an exclusively bright galaxy sample. The question therefore remains whether the longer correlation length of the pBzKs is simply due to their mean luminosity, rather than because they are passive. To answer this question requires a much more in depth study, and this study is provided in the next chapter.

Chapter 3

The clustering evolution of passive and star-forming galaxies over $0 < z < 3$

This chapter further develops the work of the previous chapter by performing selections of passive and star-forming galaxies in a more detailed and physcially motivated way. The clustering properties are computed for sub-samples over the range $0 < z < 3$. Much of the work presented in this chapter features in a paper submitted on the 23rd of November, 2009. All magnitudes in this chapter are given in the Vega magnitude system unless otherwise stated.

3.1 Previous work on the clustering of bimodal galaxy populations

Over the last few years the clustering technique has been applied to complete samples of optically-selected galaxies, separated by luminosity and colour. Zehavi (2005) computed the angular auto-correlation functions of red and blue volume-limited sub-samples of galaxies from the Sloan Digital Sky Survey (SDSS; York 2000) for separations $< 10h^{-1}$Mpc and redshifts, $z < 1$. They found that the population of galaxies red in $g-r$ colour are more strongly clustered than those with bluer colours ($r_0 = 5.7$ and $3.6h^{-1}$Mpc for red and blue populations respectively). However, they show that this difference remains when the samples are further broken down into absolute R-band magnitude sub-samples and, similar to the results of Einasto (1991) and Madgwick et al. (2003), that red galaxies exhibit a steeper slope.

Meneux et al. (2006) split the VIMOS-VLT Deep Survey (VVDS; Le Fèvre et al. 2005) galaxies into red and blue spectral types by the method of Zucca et al. (2006) and computed their clustering out to $z \simeq 1.2$. They found that the deprojected clustering scale length of the red spectral type galaxies ($r_0 \sim 5h^{-1}$Mpc) was greater than that of the blue type galaxies ($r_0 \sim 3h^{-1}$Mpc) at all redshifts investigated. Carlberg et al. (1997) find similar results for their sample at $z \simeq 0.6$. More recently Coil et al. (2008) used the DEEP2 galaxy redshift survey to address the same question at $z = 1$. They also found that red-sequence galaxies are more strongly clustered ($r_0 \sim 5$ and $3h^{-1}$Mpc for red and blue samples), but additionally showed that those galaxies with intermediate colours had an intermediate clustering strength. They computed the relative bias between their sub-samples, finding a smoothly increasing bias with rest-frame $U - B$ colour up to the red sequence.

Recently, two further deep surveys have been used to study the clustering of galaxies in the range $0.2 < z < 1.2$: the Canada-France-Hawaii Telescope Legacy

Survey (CFHTLS, Ilbert et al. 2006) and the zCOSMOS field (Lilly et al. 2007). McCracken et al. (2008) used the CFHTLS, splitting the galaxy sample by template-fit spectral type and B-band luminosity. Based on photometric redshifts, they find that the difference in clustering strength between galaxies fit by early type and late type templates is roughly constant across all redshifts studied. The luminosity dependence at a given redshift, however, is minimal. Intriguingly, the early type population at $z \sim 0.5$ shows hints of an inverse dependence of luminosity on clustering strength. The authors attribute this behaviour to fainter red galaxies being preferentially located in cluster environments. Meneux et al. (2009) meanwhile, use the spectroscopic zCOSMOS data, finding a weak dependence on luminosity which remains when stellar mass is used instead.

Each of these studies has been based upon an optical colour selection, which limits these studies to $z \leq 1$. Beyond this redshift optical studies become increasingly biased against passive and dusty galaxies and so are only able to probe the clustering of unobscured star-forming galaxies. The relative clustering strengths for complete samples of passive and star-forming galaxies are therefore poorly understood beyond $z \sim 1$. Infrared survey data are not biased in this way and in addition the K-band luminosity is a reasonable proxy for the stellar mass of a galaxy. Infrared-selected samples are therefore vital in the study of high redshift galaxy clustering.

Utilising infrared data, the high redshift progenitors of nearby high-mass, passive systems have been suggested as being passively-selected BzK galaxies (Daddi et al. 2004; Kong et al. 2006; Hartley et al. 2008), ultra-compact z=2 red galaxies (e.g., Zirm et al. 2007; Toft et al. 2007; Cimatti et al. 2008), sub-mm galaxies (Hughes et al. 1998; Swinbank et al. 2006), 24μm sources selected at $z > 1$ (e.g., Magliocchetti et al. 2008), distant red galaxies (DRGs) (Franx et al. 2003; Foucaud et al. 2007), extremely red objects (EROs) (Roche et al. 2002; McCarthy et al. 2004) and Ultra-luminous infrared galaxies (ULIRGS) (e.g. Farrah et al. 2006). Each of these populations have been found to cluster very strongly, with r_0 values in the range $7 < r_0 < 14h^{-1}$Mpc (Roche et al. 2002; Blain et al. 2004; Farrah et al. 2006; Foucaud et al. 2007; Magliocchetti et al. 2008).

In the previous chapter, which was also published as Hartley et al. (2008) (henceforth H08), the colour selection criteria of Daddi et al. (2004) was used to investigate the clustering properties of passive and star-forming galaxies at $1.4 < z < 2.5$. It was shown that, when limiting the samples to $K_{AB} < 23$, the passive galaxies were significantly more clustered than star-forming galaxies. This result was found in angular clustering and also when deprojected to find the real-space clustering strength; with clustering scale lengths of $\sim 15h^{-1}$Mpc for the passive sample (pBzK) and $\sim 7h^{-1}$Mpc for the star-forming sample (sBzK).

Although all of these different galaxy populations overlap to some extent (Lane et al. 2007), their biases differ and they are sensitive to different epochs. It is therefore difficult to compare these samples with low-redshift samples or to study global galaxy evolution. The analysis in H08 in particular left some open questions. It was established that passive galaxies are more clustered even at such high redshifts, but the passive population is exclusively bright and the star-forming galaxies exhibited a strong limiting-magnitude dependence on clustering strength. It is therefore not clear whether it is luminosity or passivity that is more significantly correlated with clustering strength. Even at low-redshift debate remains over such a distinction (Norberg et al. 2002), complicated by the dominance of passive systems at the bright end of the luminosity function.

Williams et al. (2009) used a two-colour rest-frame selection to select a quiescent cloud and star-forming track over the range $1 < z < 2$. Using this novel selection, they compute the clustering scale lengths of their two samples, finding very similar lengths to those of the BzK selections in H08. They also report evidence for a mild evolution in clustering with luminosity for the star-forming galaxies by splitting their samples into two luminosity bins. Coil et al. (2006) also find, at $z = 1$, only a modest dependence for clustering strength to increase with B-band luminosity, covering $0.7L^* < L < 1.4L^*$.

A further remaining question is *when* is the passive population built up? Though it was found in H08 that this build-up occurs over $1.4 < z < 2.5$, and Williams et al. (2009) find similar evidence for their sample at $1 < z < 2$, these ranges in redshift are very broad. Can we identify a narrow epoch at which the clustering strengths of passive and star-forming galaxies of a given luminosity are equal? At such an epoch passive and star-forming galaxies would occupy the same mass of dark matter halos and differences in their internal processes could be clearer. In this chapter I attempt to disentangle the effects on clustering strength due to passivity and luminosity and to further refine the redshift range over which the passive population of galaxies is first established.

3.2 Sample preparation

The analysis presented in this work is primarily based on the third data release (DR3) of the UKIDSS UDS. The principal improvement from the DR1 to the DR3 is the addition of H-band data to match the K-band and J-band data already released as the DR1. Additional data in the K-band also provides an incremental improvement in depth. The 5σ, Vega depths within $2''$ apertures for the DR3 in J, H and K-bands are 22.8, 22.1 and 21.8 respectively. These depths make it the deepest single-field near infrared survey of its area to date. For details of the stacking procedure, mosaicing, catalogue extraction and depth estimation the reader is referred to Almaini et al. (in prep.) and Foucaud et al. (2007).

In addition to the deep optical data in the B, V, R, i' and z'-bands described in chapter 1 section 1.4, this work utilises *Spitzer* data reaching 5σ depths of 24.2 and 24.0 (AB) at 3.6μm and 4.5μm respectively from a recent *Spitzer* Large Program (SpUDS, PI:Dunlop) and U-band data taken with CFHT Megacam. Due to the much larger PSF of the Spitzer data, matching them to the near-infrared data is a difficult deblending problem which was performed as part of the photometric redshift determination by Dr Cirasuolo in Edinburgh. As before, border regions, areas around bright stars and obvious cross-talk artifacts were masked and any sources within such regions were discarded. The co-incident area with the UDS taking masking into account is 0.63 deg^2.

The world-public DR3 catalogue extracted from the K-band image is used as the basis for selection, upon which a cut at $K_{AB} = 23$ within $2''$ is imposed. This cut was made to minimise photometric errors and the number of spurious sources as well as ensure a high level of completeness and robust photometric redshifts (see below). This limit is fainter than the corresponding value for M_K^* at $z \simeq 1.5$ (Cirasuolo et al. 2010 and fig. 3.3). Corresponding $2''$ diameter magnitudes for each object were extracted directly from the optical, J and H-band images at the position of the source after a detailed astrometric re-alignment to the UDS K-band image.

From this combined catalogue stars were removed in the following way: bright

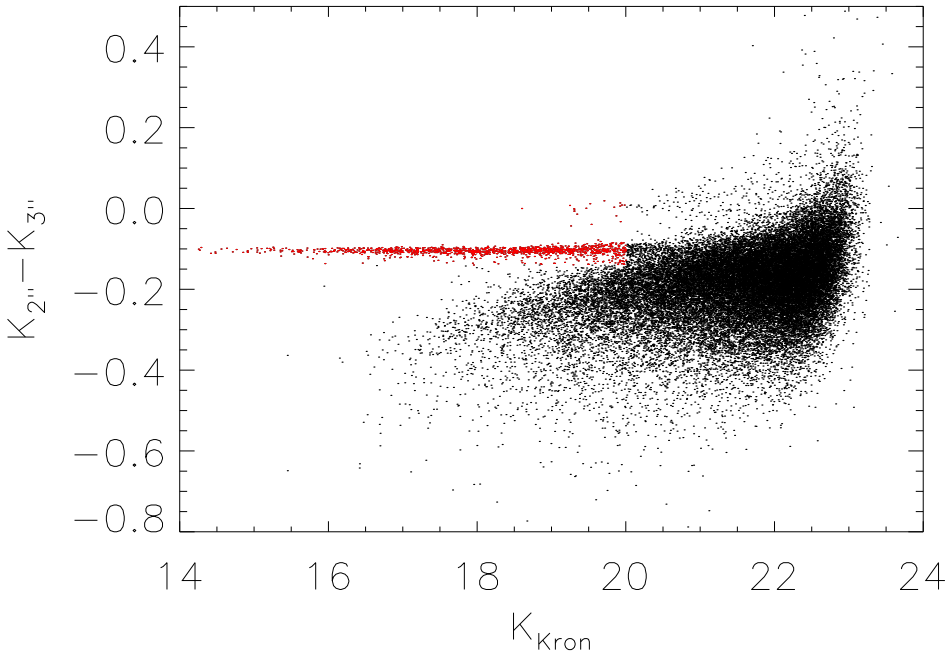

Figure 3.1. Difference in K-band $2''$ and $3''$ diameter magnitudes against Kron magnitudes (AB) for the sources in the UDS DR3 catalogue. Bright stars form a tight sequence (red points).

($K_{AB} < 20$) stars were removed by excluding those with $K_{3''} - K_{2''} > -0.14$, where the stars are clearly separated from the galaxies (Figure 3.1). The remaining, fainter stars form a stellar locus in the $(B - z') - (z' - K)$ plane (as noted by Daddi et al. 2004) and can easily be removed by the criterion $(z' - K) < 0.3(B - z') - 0.5$ (Cirasuolo et al. 2010). Heavily saturated stars and the surrounding contaminated regions were carefully masked out during the analysis. These cuts left 33923 galaxies from which the selections were made.

3.2.1 Redshifts

The photometric redshifts used during this chapter, in common with the rest of this thesis, were not computed by myself. A general description of the photometric redshift technique is given in Chapter 1, Section 1.3.3. However, I shall provide some more specific details relevant to this work in this section.

The magnitudes obtained from the images were used to determine photometric redshifts (photo-zs) and stellar ages by a χ^2 minimisation over a large suite of model templates. The template set comprised only those of synthetic SEDs for which the star-formation histories are known. These histories are parameterised by an age since the onset of star-formation and the e-folding time, τ, characterising the decay of star-formation rate with time:

$$SFR_o = SFR_{\text{initial}} \times e^{-age/\tau}; \tag{3.1}$$

where SFR_o is the SFR at the time of observation.

The templates were constructed using Bruzual & Charlot (2003) with a Salpeter initial mass function (Salpeter 1955), and including a treatment of dust content (see Cirasuolo et al. 2007, 2010 for a full account). These templates cover nine values of the exponential star-formation rate decay, τ (0.1, 0.3, 0.5, 1, 2, 3, 5 and 10 Gyrs, and an instantaneous burst), and allow stellar ages up to the age of the universe at any given redshift. The dust reddening is allowed to vary between $0 < A_V < 2$, following the Calzetti law (Calzetti et al. 2000), and line of sight neutral Hydrogen absorption, calculated from Madau (1995), is also taken into account.

Any source with an unacceptable fit was removed from the catalogue ($\chi^2 > 15$, 4.0% of the galaxy sample) as they will likely be mis-assigned when the sample is broken up by photometric redshift and luminosity, and the derived stellar ages are unlikely to be reliable for such objects. Upon investigation the majority of these discarded sources were found to be cross-talk artifacts (26%), QSOs (36%) or the minor members of pairs or mergers (23%). The majority of the remaining discarded objects were of very low surface brightness. Many of these fainter objects would not have made it into the volume-limited sub-samples. The fraction of otherwise useful objects rejected by this $\chi^2 > 15$ criterion is therefore $< 0.6\%$. Rest-frame magnitudes and colours were computed by integrating the appropriate filter over the best-fitting template of each object. This work requires the use of absolute rest-frame K-band magnitudes and rest-frame $U - B$ colours. It should be noted that at $z > 1$ the rest-frame K-band moves beyond the $4.5\mu m$ band of Spitzer and so is dependent on a good template fit at shorter wavelengths.

The star-formation history is required in order to select the passive and star-forming samples used later in this chapter. While the accuracy of photometric redshifts are more strongly dependent on the number of photometric points, there is nevertheless a concern that by using only synthetic templates systematic biases may be introduced. The agreement between the photometric redshifts used in this work and the spectroscopic redshifts are as good as those from a full template set (see Figure 1.8, section 1.3.3). However, a global $\Delta z/(1 + z)$ statistic may hide systematic differences between photometric redshifts produced from different template sets. Figure 3.2 compares the photometric redshifts and derived properties obtained for the restricted template set used in this chapter and a full template set, analogous with the previous chapter.

The most fundamental quantity, the photometric redshifts (upper left panel), are largely consistent between the two template sets at $z < 2$. The upper right panel shows how the mean absolute difference in photometric redshift between the two template sets varies with photometric redshift. The typical global uncertainty in redshift is $\Delta z \sim 0.12$, though the value for any given galaxy will depend on redshift, magnitude and SED. The differences in photometric redshift are smaller than the typical error at $z < 1.5$ and of the same order at $1.5 < z < 2$. Above $z = 2$ the discrepancy increases and measurements at these redshifts should therefore be treated with caution. It should be noted, however, that even at $z = 3$ the mean difference in photometric redshift is less than the bin widths used in the analysis (0.5 in redshift). The derived quantities: absolute rest-frame K-band magnitude and $U - B$ colour are shown in the lower two panels. In both cases the reduced template set appears to have reproduced these quantities fairly well, with very few extreme outliers. However, there does seem to be some systematic effect in the $U-B$ colours, particularly for colours bluer than $U - B = 0.4$. However, as these are not

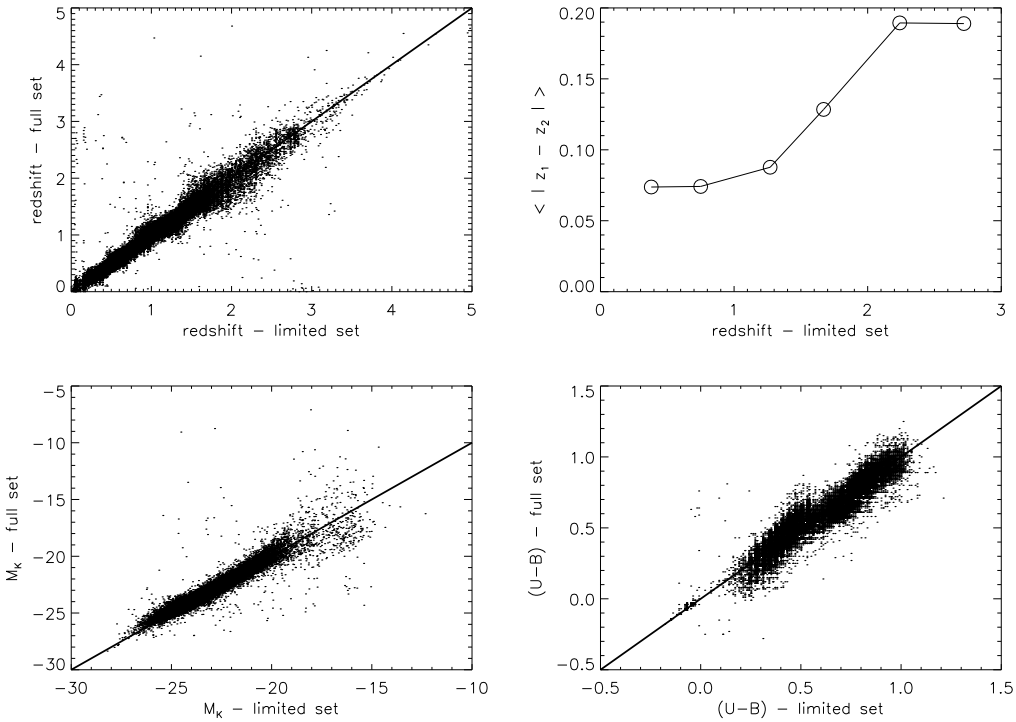

Figure 3.2. Comparison of photometric redshift-dependent quantities derived from two different template sets: a full set and a set restricted to synthetic templates (see text). **Upper left:** photometric redshifts. **Upper right:** mean difference in photometric redshift against photometric redshift of the synthetic template set. **Lower left:** K-band luminosity. **Lower right:** $U - B$ colour.

close to the red/blue selection boundary (section 3.4.1) the impact of this effect will be minimal.

3.3 The global evolution of clustering

One of the most fundamental measures of the evolution of galaxies is how their clustering strength evolves as a function of time and how it depends upon their stellar mass content. Before approaching the distinction between star-forming and passive galaxies it is useful to study how the clustering strength of all galaxies brighter than some limit evolves with time.

In order to compare measurements at different redshifts, the sub-samples must be constructed such that Malmquist bias is avoided; i.e. the sub-samples must be volume limited, rather than magnitude limited. The sub-samples used are shown in Figure 3.3 (dot-dashed lines) together with the data and the evolving M_K^* (Cirasuolo et al. 2010, dashed line). The majority of sub-samples are of galaxies with luminosities similar to or slightly fainter than M_K^*. The measurements are therefore representative of the galaxies which dominate the stellar mass of the universe at the epochs under study.

Deprojected correlation lengths were computed for each of the sub-samples shown in Figure 3.3, provided that subsample contained ≥ 150 galaxies. The determina-

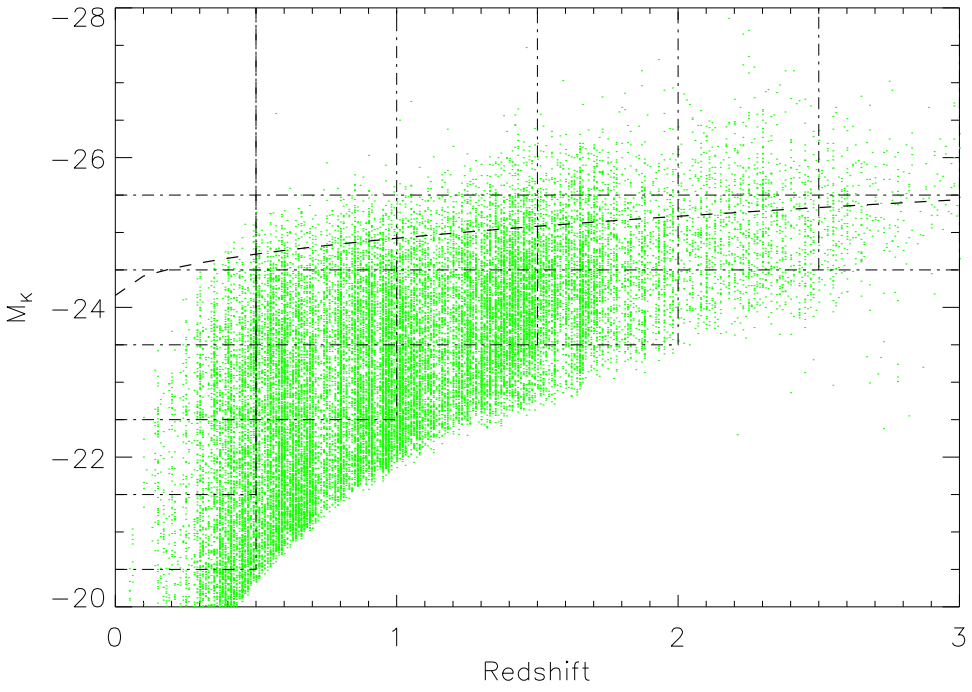

Figure 3.3. The distribution of UDS galaxies used in this work in the $z - -M_K$ plane. The dot-dashed lines show the sub-sample regions for which clustering lengths are computed, provided they contain > 150 galaxies. Also plotted as the dashed-line curve is the fit for an evolving M_K^* as given in Cirasuolo et al. (2010).

tion of the correlation lengths largely followed the procedure outlined in Chapter 1, Section 1.3.1 and Chapter 2, Section 2.3.2.

Angular correlation functions were computed using the estimator of Landy & Szalay (1993). The correlation functions were then fit by χ^2 minimisation with a power law of fixed slope, $\delta = -0.8$. The fit was made over mid to large scales to avoid biases arising from multiple halo occupation. The small-scale limit corresponds to $0.3h^{-1}$Mpc and the large scale cut off is at 0.2 degrees where survey area limits our measurements. This method is the same as that presented in Chapter 2. Unlike Chapter 2 however, cosmic sampling variance is not taken into account, using simple bootstrap error estimates instead. The rationale for this decision is discussed in Section 3.8.

A second point at variance with Chapter 2 is the redshift distribution that was used during deprojection. While the sBzK and pBzK samples were assumed to have a highly peaked redshift distribution which was being broadened by the redshift errors, such an assumption is inappropriate for the samples considered here. This change in the prior for the redshift distribution makes deconvolving out the redshift errors extremely problematic. Rather than introduce further unknown uncertainties into the correlation length measurement, the raw n(z) was used during deprojection for these samples. This issue is discussed further in section 3.8.1.

Using the method outlined above clustering scale lengths, r_0, were compute for each of the volume limited sub-samples shown in Figure 3.3, provided that each

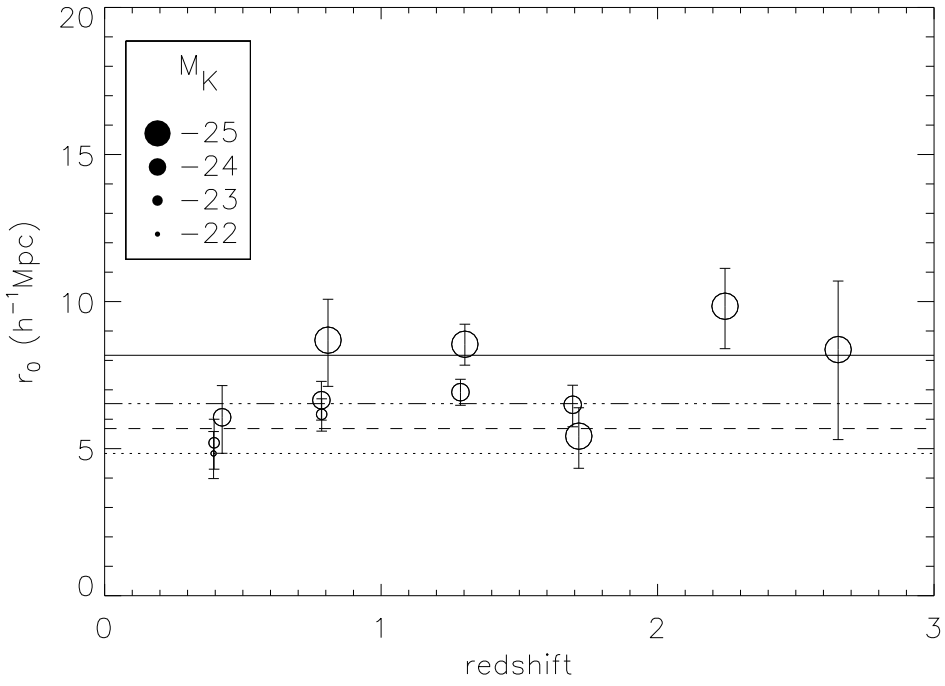

Figure 3.4. The global evolution of clustering in the UDS field. Galaxies were split into the volume-limited sub-samples shown in Figure 3.3 and the real-space correlation lengths were computed (see text). The evolution of clustering strength with redshift is at most mild for all luminosity ranges and can be reasonably fit by lines of constant clustering (shown top to bottom for decreasing luminosity samples).

subsample contained at least 150 galaxies. These measurements are shown in Figure 3.4. The first thing of note in these measurements is the lack of strong evolution of r_0 with redshift at a given M_K. Furthermore, there is very little luminosity segregation, consistent with the findings of Coil et al. (2008), although the most luminous galaxies are slightly more strongly clustered at $z \sim 1$.

3.4 The clustering of red and blue galaxies

3.4.1 Selections

In much of the literature concerning the evolution of galaxy clustering, separation of populations by colour, rather than star-formation properties, is preferred. Rest-frame colours are simpler to compute and therefore allow simpler comparison between different works.

In this work the separation of red and blue is performed making use of the best fit star-formation histories of the sample galaxies. At low redshift the red sequence is found to be very tight, consisting mainly of objects with a strong 4000Å break. The uncertainties in the redshifts and templates blur out the red sequence in these data to some degree. However, those galaxies which are best fit by a template with a strong break are likely to have more accurate redshifts and $U - B$ colours.

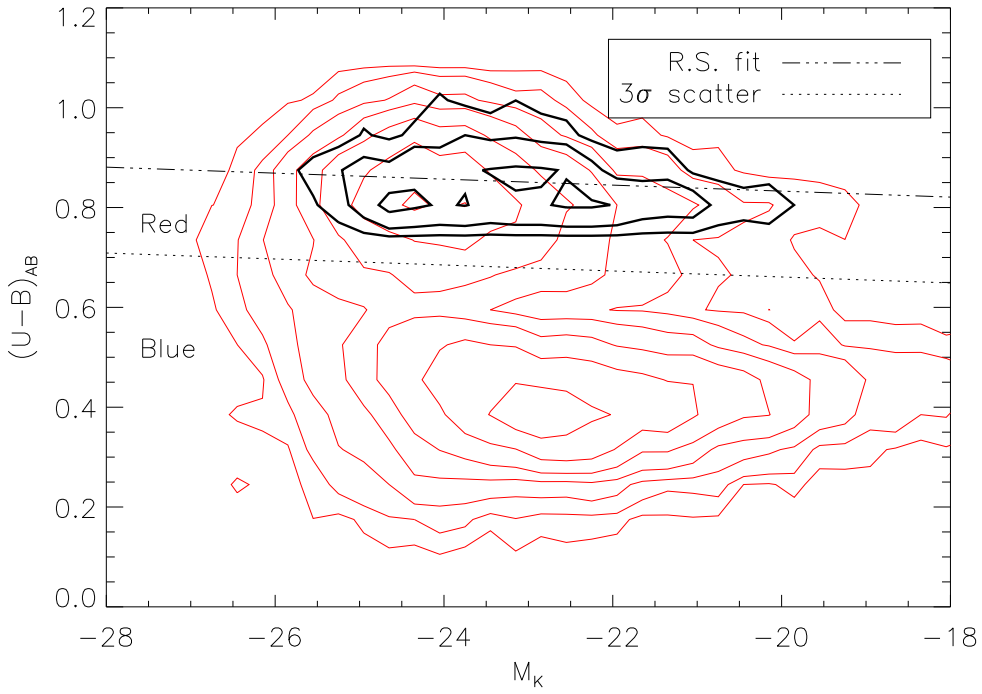

Figure 3.5. Colour-luminosity density plot of the full galaxy sample (light, red contours) and of those used to define the red sequence (heavy, black contours). The red sequence was defined by fitting a line of the form $U - B = a \times M_K + b$ to the galaxies which were best fit by a burst template with age > 1 Gyr. This fit is shown by the dot-dashed line, while the dotted line is the 3σ lower envelope. Galaxies redder than the dotted line are defined as 'red' for the purposes of sample selections.

The red sequence is initially defined using the most conservative and clean sample of passive galaxies drawn from the catalogue. This sample is constructed to be those galaxies which are simultaneously old (age $>$ 1Gyr) and are fit with a burst template, as these values imply a strong 4000Å break. A χ^2 minimisation to fit an equation of the form $U - B_{AB} = a \times M_K + b$ to these galaxies is then performed. The 'red' sample is taken to include all galaxies within 3σ scatter of this fit (or redder). In Figure 3.5 contours showing where galaxies lie in the $(U - B)_{AB}$ vs. absolute K-band magnitude plane are plotted. Overlaid in black are the contours for the burst galaxies described above, the associated best fit (dashed-dotted line) and red/blue selection criterion (dotted line).

This Red/Blue division is suitable for galaxies close to the median redshift of these old, burst-SED galaxies ($z = 0.7$). However, the red sequence is known to evolve with time, with redder colours at lower redshifts (e.g. Brammer et al. 2009). Accurately accounting for this evolution with photometric redshifts is problematic. The typical errors on the $U - B$ colours are 0.15 magnitudes, which is of the same order as the expected colour evolution. For simplicity, the red sequence is not treated in such detail and instead a fixed colour selection boundary is used. It should be noted that the 'red' sample includes part of the green valley, and so it is unlikely many intrinsically red galaxies at high redshift are being missed.

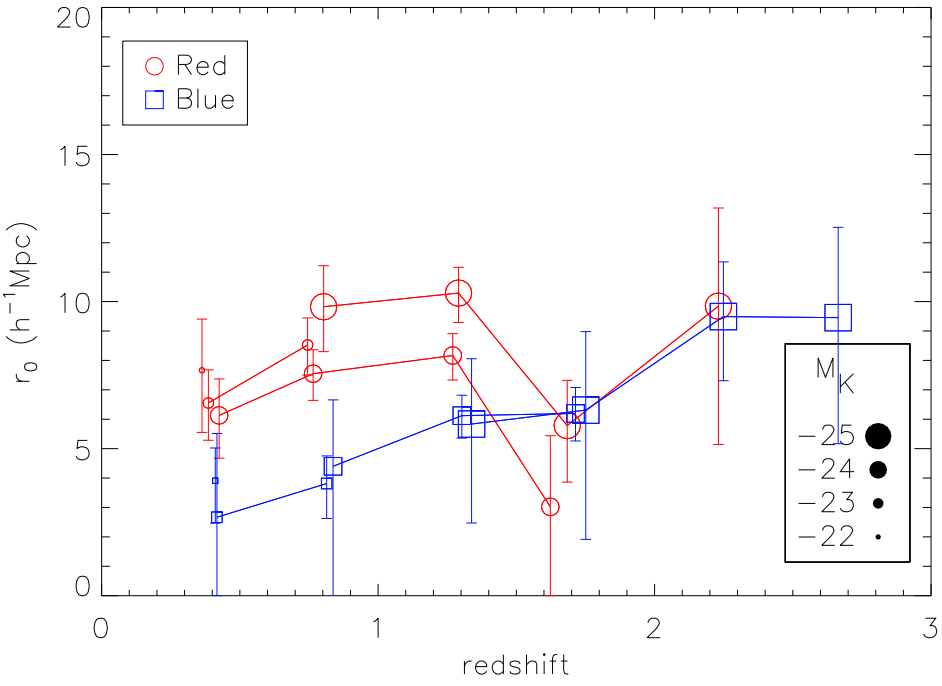

Figure 3.6. The sub-samples in Figure 3.3 were further separated into red and blue in rest-frame $U - B$ colour and the clustering computed once again.

3.4.2 Clustering results

In Figure 3.6 r_0 values are again computed for the subsamples in Figure 3.3, but this time they are for the separate red and blue galaxy samples, defined in Figure 3.5. The process for these red and blue sub-samples is otherwise identical to that used for the full sample. For $z < 1$, red galaxies are significantly more clustered than blue galaxies. This difference is particularly pronounced for the brightest subsamples.

Above $z = 1.5$ the clustering strengths of the blue and red subsamples appear to converge, indicating that this may be the principlal epoch at which the red sequence is populated. The red population is inhomogeneous, however, containing both passive and dusty star-forming objects. The star-formation histories from the template fits are therefore used to separate the passive and star-forming galaxies in the following section.

3.5 Passive and star-forming galaxies

3.5.1 Sample selection

In this section the definition of the passive galaxy sample is defined. At low redshift, passive galaxies form a well-defined red sequence in a diagram of rest-frame $U - B$ colour versus absolute magnitude. The red sequence is not exclusively composed of passive systems, however, and typically contains a 30% contamination from dusty star-forming galaxies (e.g. Wolf et al. 2005). At high redshifts ($z > 1$) contamination by dusty star-forming objects becomes increasingly problematic (Daddi et al. 2004).

In addition, a simple $U-B$ cut does not make full use of all the available photometry. For these reasons, the passive sample is defined with two criteria: (a) the galaxy must be 'red' in rest-frame $U-B$ colour and (b) the best-fitting galaxy template must imply a very low level of residual star formation. These criteria are further detailed below.

In order to remove the galaxies that are red due to dust rather than age the information about the star-formation histories obtained from the templates is used. To be selected as a passive galaxy it is required to be red, as defined above, older than $1Gyr$ and to have a best-fit SFR (from the photo-z templates) below some cut-off value (as given by equation 3.1, where $\tau = 0$ for a burst model). By varying this SFR cut-off criterion it is possible to additionally study how the inclusion of galaxies with greater recent star formation rates influences the clustering strengths of the sub-samples.

For a primary passive sample, a highly conservative limit to residual star-formation rate is required: $SFR_o \leq 0.1\%$ of $SFR_{\rm initial}$ (in addition to being red in rest-frame $U-B$ and with a stellar age $>$ 1Gyr). This strict limit was chosen to minimise the contamination by dusty star-forming objects whilst retaining sufficient objects with which to compute meaningful clustering lengths. There are well-known degeneracies between age, metallicity and dust content. These cannot be completely overcome, even with the multiple bands of deep photometry covering the UDS field. However, these parameters imply a strong break in the galaxy SED and hence only the most extreme dusty contaminants will be included. The number of objects satisfying these selection criteria is 3991 (33% of the 'red' sample). The vast majority of passive galaxies selected in this way also lie within the 'pERO' colour selection boundaries (see section 3.8.1).

As a baseline against which to compare the passive sample, a star-forming sample with a single criterion is chosen: $SFR_o > 10\%$ of $SFR_{\rm initial}$. This star-forming sample consists of 22856 objects. Three further passive samples were then constructed by relaxing the SFR criterion used in the primary passive sample to include galaxies with $SFR_o \leq 0.5\%$, $SFR_o \leq 1\%$ and $SFR_o \leq 10\%$ of $SFR_{\rm initial}$. In each of these three additional samples the requirements for red $U-B$ colour and stellar age are kept the same as the primary sample. The number of objects in each of these other passive samples are 6111, 7677 and 10032 respectively. These samples will occasionally be referred to by the SFR limit imposed during selection.

3.5.2 Results

Deprojected clustering lengths of the passive and star-forming samples are plotted in Figure 3.7 with the SFR limits described in section 3.5.1 shown in each panel. The age $>$ 1Gyr criterion means that there are very few galaxies beyond $z = 2$, as the Universe is only $\sim$ 3Gyrs old at that point. For the most conservative passive sample (SFR $\leq 0.1\%$) the clustering strengths of the passively-selected galaxies are greater than those of the equivalent star-forming galaxies for almost all sub-samples up to $z = 1.5$. This result, shown in the upper left panel of Figure 3.7, shows that the clustering length is more strongly dependent on passivity than on luminosity, with very little luminosity segregation in the star-forming sample. This difference continues to at least the median redshift of BzK-selected galaxies (chapter 2).

Above $z \sim 1.5$, however, the respective clustering strengths of passive and star-forming galaxies appear to converge. Further survey data will be required to confirm this finding, which is discussed in section 3.8. As the SFR requirement is relaxed

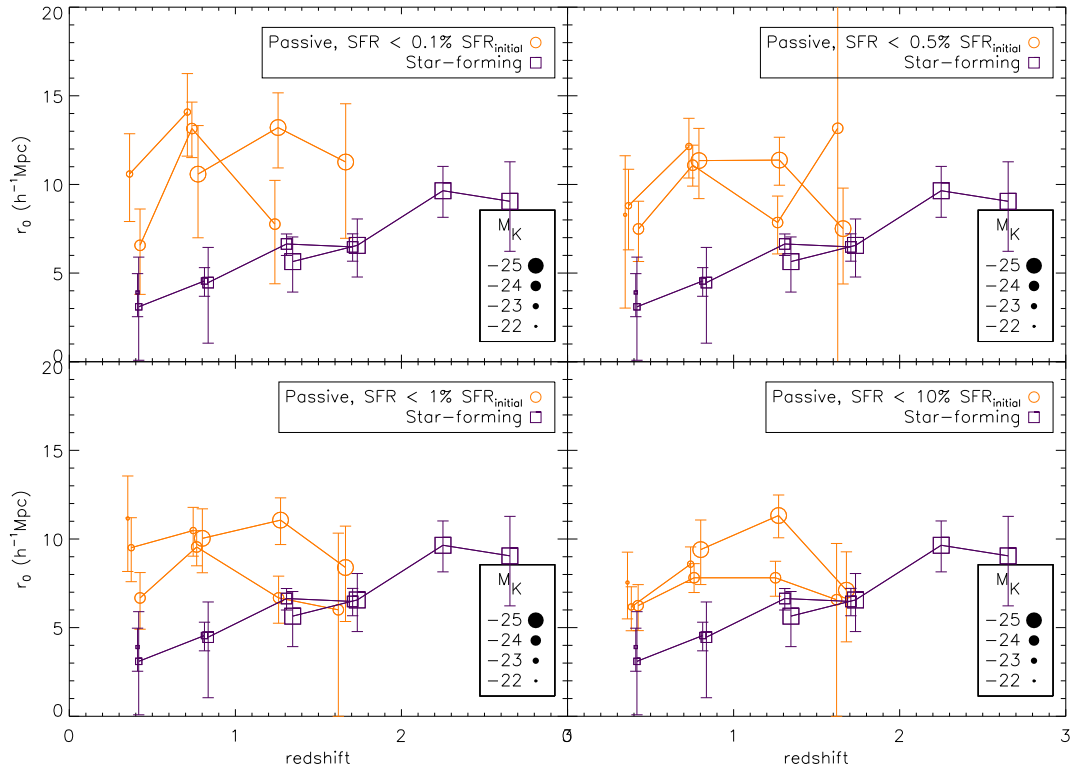

Figure 3.7. Evolution of the clustering strengths with redshift for the four passive samples with the single comparison star-forming sample. The orange and purple symbols represent the passive and star-forming sub-samples respectively, with the size representing the absolute K-band magnitude. With the most conservative passive definition (requiring red rest-frame $U-B$ colour, stellar age >1 Gyr and $SFR_o \leq 0.1\%\ SFR_{\rm initial}$), the passive galaxy sub-samples are significantly more strongly clustered than the star-forming sub-samples for $z<1.5$. Above $z=2$ the number of passive galaxies becomes too small to make robust clustering measurements. As the passive galaxy selection criterion of SFR (given in each panel) is relaxed, the clustering strengths of star-forming and passive sub-samples become more similar. Each point is plotted at the mean redshift of the sub-sample it represents.

to allow galaxies with higher residual star-formation rates into the passive selection, the strength of clustering of the passive sub-samples decreases, which confirms that truly passive galaxies are more strongly clustered among the 'red' population.

For illustration, in Figure 3.8 angular clustering measurements for passive and star-forming galaxies over all sub-samples in the two redshift ranges, $1<z\leq 1.5$ and $1.5<z\leq 2$ are shown. The passive samples use the conservative definition ($SFR_o \leq 0.1\%\ SFR_{\rm initial}$) described in section 3.5.1. The vertical dot-dashed line shows the lower limit of the fitting range used. The solid lines are the best power-law fits and, in agreement with the earlier measurements for BzK-selected galaxies (Chapter 2 and H08), the passive galaxies are found to cluster more strongly.

3.6 Samples in the $(B-z')-(z'-K)$ plane

In this section I re-visit the 2-colour selection of Daddi et al. (2004). The aim here is to derive a series of simple boundaries which can be used to approximate the

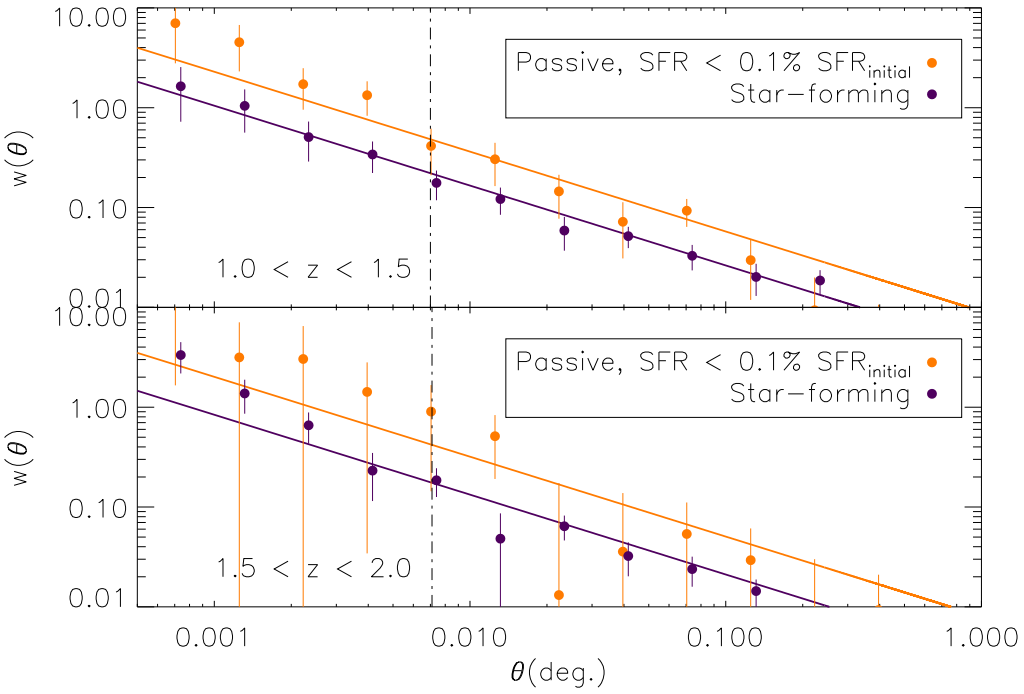

Figure 3.8. Angular clustering measurements for conservative passive galaxies ($SFR_o \leq 0.1\%\ SFR_{\rm initial}$) and star-forming galaxies within the redshift ranges $1 < z \leq 1.5$ (upper) and $1.5 < z \leq 2$ (lower) with the star-forming points off set for clarity. Subsamples in luminosity have been combined for illustrative purposes. Solid lines show the best power-law fits, using a constant slope of -0.8, at large-scales ($0.3\ h^{-1}$Mpc (dot $-$ dashedline) $-\ 0.2$ deg.). In both cases the passive samples are more strongly clustered, consistent with the measurements of pBzK and sBzK-selected galaxies (chapter 2). Errors are determined from a bootstrap analysis.

passive samples used earlier in this work. The purpose of doing so is to facilitate the reproduction of the results presented in the previous section in complementary fields, rather than create a robust photometric definition for a passive sample. The $(B-z')_{AB}-(z'-K)_{AB}$ plane was chosen because the pBzK selection of Daddi et al. (2004) is well understood, and also because of the track identified in Lane et al. (2007). This track has colours consistent with the SED of an E/S0 type galaxy, evolved through redshift, and is well separated from the rest of the distribution. It suggests that it may be possible to separate passive and star-forming galaxies over all redshifts with relative ease.

The use of such simple photometric sample definitions may be preferable to full star-formation history fitting when comparing different data sets. In minimising the complexity of selection, the intrinsic differences in the samples should be more apparent and hence a better estimate for cosmic sample variance can be obtained. The colour boundaries outlined below are defined using the Subaru and UKIRT filters. Conversion to other filter sets can be made with reference to Hewett et al. (2006) and Furusawa et al. (2008).

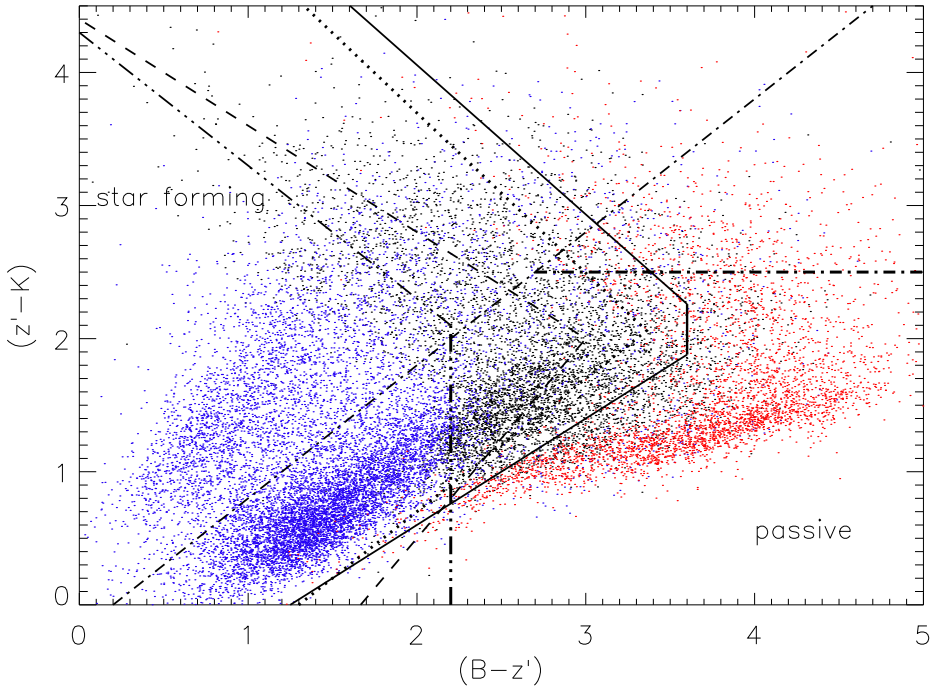

Figure 3.9. $(B-z')_{AB}-(z'-K)_{AB}$ colour-colour plot for actively star-forming sample (blue points), the 'conservative' passive galaxies (red) and those which fall between these criteria (black). Also shown are the commonly used BzK selection criteria of Daddi et al. (2004) (dashed-dotted lines) and the borders defined in the text. The solid line marks out the boundary which approximates the conservative ($SFR_o \leq 0.1\%\ SFR_{initial}$) selection, the dotted line is for galaxies with $SFR_o \leq 0.5\%$, the dashed line is for $SFR_o \leq 1\%$ and the dashed-triple dotted line is for $SFR_o \leq 10\%$. Only half of the star-forming points are plotted for clarity.

3.6.1 Selection

The plane was split into cells, 0.2×0.2 in colour, and within each cell the fraction of galaxies that are in the passive sample was computed. Boundaries were chosen to provide a simple selection technique which includes those cells with $> 50\%$ passive galaxies. These borders are shown in Figure 3.9 together with the conservative passive sample and star-forming samples used earlier in this chapter.

The boundaries for the $SFR_o \leq 0.1\%\ SFR_{\rm initial}$ sample are:

$$\begin{aligned} & (z'-K)_{AB} < 0.8 \times (B-z')_{AB} - 1 \\ \text{or} \quad & (B-z')_{AB} > 3.6 \\ \text{or} \quad & (z'-K)_{AB} > -1.4 \times (B-z')_{AB} + 7.3; \end{aligned} \tag{3.2}$$

for the $SFR_o \leq 0.5\%\ SFR_{\rm initial}$ sample:

$$\begin{aligned} & (z'-K)_{AB} < (B-z')_{AB} - 1.3 \\ \text{or} \quad & (z'-K)_{AB} > -1.2 \times (B-z')_{AB} + 6.1; \end{aligned} \tag{3.3}$$

for the $SFR_o \leq 1\%\ SFR_{\rm initial}$ sample:

$$\begin{aligned} & (z'-K)_{AB} < 1.5 \times (B-z')_{AB} - 2.5 \\ \text{or} \quad & (z'-K)_{AB} > -0.8 \times (B-z')_{AB} + 4.4; \end{aligned} \tag{3.4}$$

and for the $SFR_o \leq 10\%\ SFR_{\rm initial}$ sample:

$$\begin{aligned} & (B-z')_{AB} > 2.2 \\ \text{or} \quad & (z'-K)_{AB} > -1.0 \times (B-z')_{AB} + 4.3; \end{aligned} \tag{3.5}$$

These 2-colour-defined passive samples include all galaxies in the relevant region regardless of their best-fit age, $U-B$ colour or SFR. The number of objects in each sample are 5438, 7287, 8986 and 12242 respectively, and the number of objects in the star-forming sample is 20645.

3.6.2 Clustering of the $B-z$, $z-K$ colour-selected samples

The clustering properties of the these galaxy samples were treated in the same way as the earlier samples. In each case the properties of galaxies selected by the photometrically-defined borders, shown in Figure 3.9, reflect those of the intended population: the majority of sub-samples having very similar clustering strengths. Each of the conclusions drawn previously are equally applicable to these photometrically-selected samples.

The border-defined passive samples include slightly larger numbers of galaxies than the more physically motivated selections presented earlier in this work. The additional galaxy numbers enable measurements beyond $z = 2$ for some of these samples. Though highly uncertain, these measurements indicate equivalence of clustering strength at $z > 2$.

Defining a sample in this way enables not only simple comparison between different data sets, but also avoids the need for star-formation histories from the template fits. In Figure 3.11, the measurements for the passive and star-forming samples are again computed, but this time the redshifts and K-band luminosities were computed from a full template set, including empirical galaxy templates. These measurements are in excellent agreement with those using the restricted template set.

3.7 Halo models from DM theory

As stated previously, the correlation length of a galaxy sample can be directly related to the typical hosting halo mass. The formalism given in Mo & White (2002) is used here to find the clustering strengths of dark matter halos of given masses as a function of redshift. These models gather together previous work (Press & Schechter 1974; Bond et al. 1991; Mo & White 1996; Jing 1998; Sheth et al. 2001) and use the ellipsoidal collapse model which has been calibrated against dark-matter only N-body simulations. Their aim is to provide the abundances and clustering of dark matter halos as functions of redshift and mass across a wide dynamic range. A much more detailed description of the formalism is provided in Chapter 5.

In Figure 3.12 these models for halos of mass $10^{10}{\rm M}_\odot$to $10^{14}{\rm M}_\odot$ are plotted together with the results from the conservative passive sample (see section 3.5.1) and star-forming sample. At $z < 2$, star-forming galaxies generally occupy halos of mass $10^{12}{\rm M}_\odot$ or smaller, while passive galaxies occupy halos of mass $10^{13}{\rm M}_\odot$ to

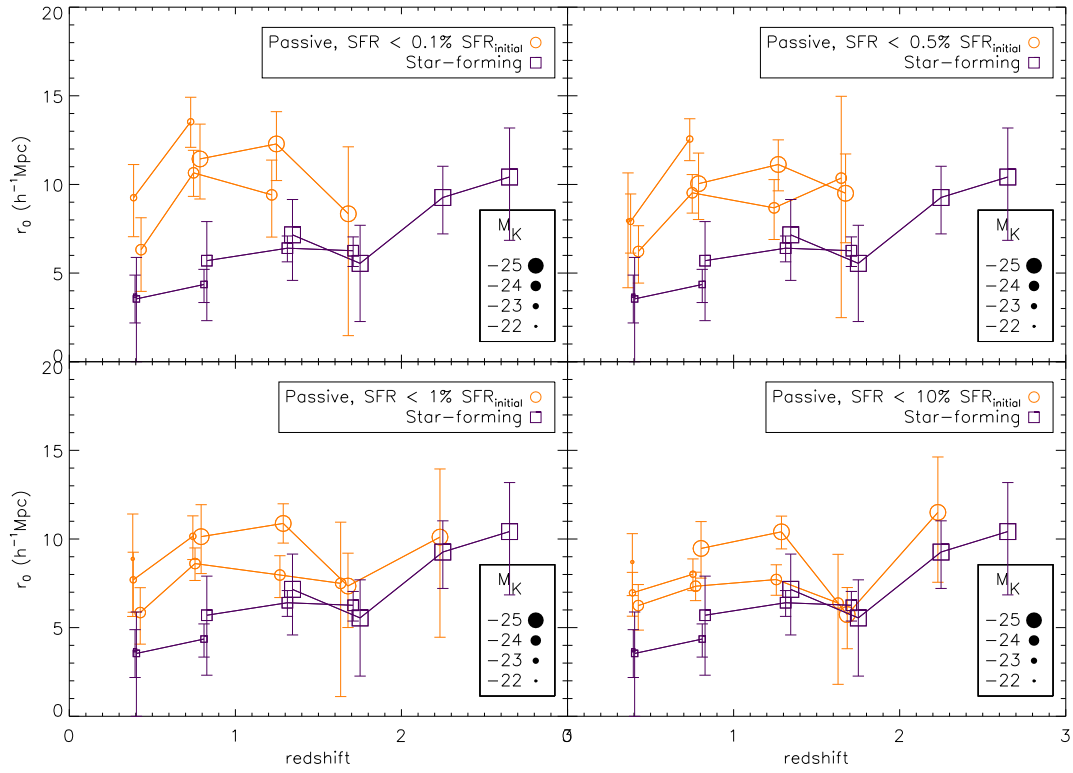

Figure 3.10. Clustering strengths of the 2-colour-defined galaxies. The symbols have the same meanings as in Figure 3.7. The values for r_0 are largely consistent with those in Figure 3.7.

$5 \times 10^{13} M_{\odot}$. The general trend for star-forming galaxies is for galaxies of a given luminosity to occupy halos of greater mass at higher redshift. This is a manifestation of downsizing (Cowie et al. 1996; Foucaud et al. 2010, submitted). This downsizing in host halo mass for star-forming galaxies appears to extend beyond z=2. At these very early times bright star-forming galaxies are hosted by massive dark matter halos ($M \sim 5 \times 10^{12} M_{\odot}$). Presumably these are among the progenitors of low redshift group-dominant galaxies.

The passive galaxies occupy halos up to an order of magnitude more massive than the star-forming galaxies. In contrast to the star-forming galaxies, the brightest passive galaxies appear to show constant clustering strength over the range $0.5 < z < 1.5$. If confirmed by future survey data, this result suggests that bright passive galaxies are hosted by similar mass halos across this redshift range.

3.8 Discussion

In this chapter it has been shown that passive galaxies, defined by fitting templates over 11 bands from the UV to *Spitzer*'s IRAC bands, are more strongly clustered than star-forming galaxies at $z < 1.5$. Clustering strength is intimately linked with the mass of dark matter halo that can host a galaxy of the given selection. Passive galaxies therefore on average occupy more massive dark matter halos than their star-forming counterparts. These results indicate that at $z < 1.5$ the passivity of a galaxy sample is a strong indicator of the typical mass of dark matter halo that

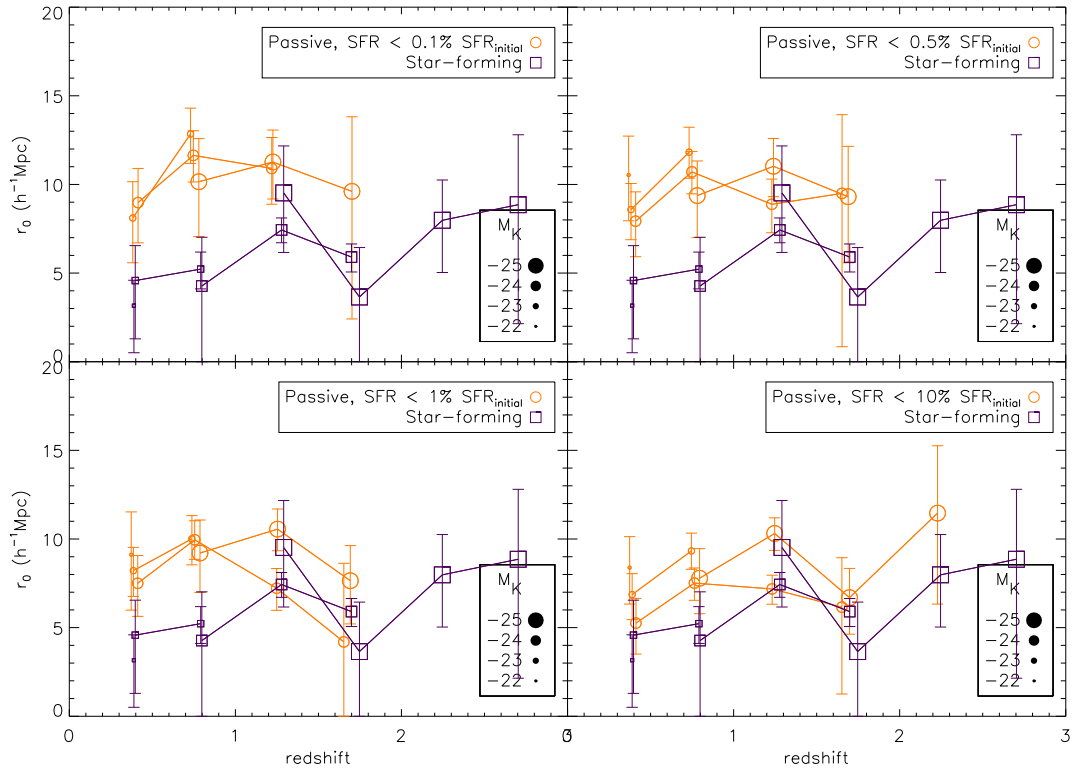

Figure 3.11. Same as figure 3.10 but with photometric redshifts and K-band luminosities derived from a full template set, rather than limited to synthetic templates. The results of the two sets are very similar.

host them.

The K-band luminosity is a reasonable proxy for the stellar mass of a galaxy, though galaxies may dim by approximately one magnitude after they stop forming stars (e.g Lilly & Longair 1984; Cowie et al. 1996). Even taking this dimming into account the conservative passive sub-samples ($SFR \leq 0.1\%\ SFR_{\rm initial}$) are typically more strongly clustered than the star-forming samples for a given stellar mass. Furthermore, there is an absence of a strong luminosity dependence on the clustering strength of the star-forming samples. Of the two effects that we set out to disentangle, passivity and luminosity, passivity is therefore the more significant indicator of clustering strength at $z < 1.5$. However, it should be noted that due to limited survey area it is not possible to probe the most extreme high luminosity galaxies ($M_K < -26$) with the UDS data set.

Above $z = 1.5$ star-forming and passive galaxy clustering are not able to be distinguished significantly. There is significant redshift evolution in the star-forming galaxy clustering, while that of the passive galaxies remains flat. However, we should not expect an inversion of the clustering strength at higher redshifts as passive galaxies must have at some earlier epoch been star-forming galaxies and their clustering strength cannot decrease with time. The correlation lengths of passive and star-forming galaxies therefore apparently converge at this epoch. If this behaviour is confirmed it would suggest that the epoch $z \sim 2$ is the epoch in which the passive and star-forming samples are first becoming distinct. Hence, it is likely to be the major epoch at which the red sequence is being populated. This finding makes the

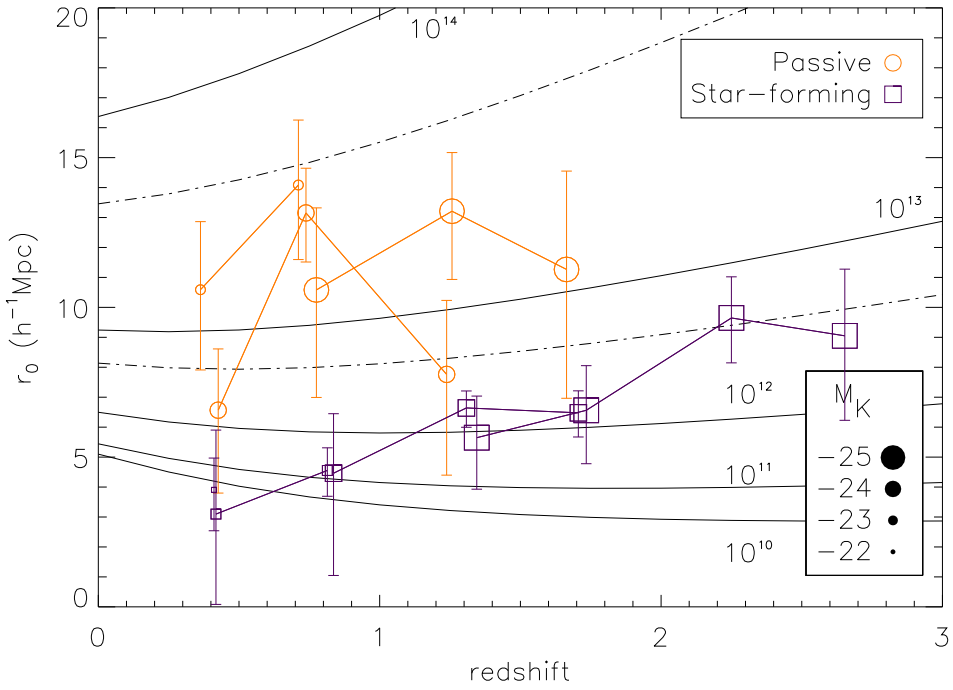

Figure 3.12. The clustering of the conservative passive sample and star-forming sample (symbols have the same meanings as in Figure 3.7) are compared with the clustering predictions of dark matter halos from Mo & White (2002). Lines of constant halo mass (in $M_\odot$) are shown, with additional dashed-dotted lines for $5 \times 10^{13} M_\odot$(upper) and $5 \times 10^{12} M_\odot$(lower). Downsizing in the star-forming population is evident with the same luminosity galaxies found in less massive halos towards z=0. The same cannot be said of the passive galaxies, which typically occupy halos of mass $M \geq 10^{13} M_\odot$.

further study of the $1.5 < z < 2.5$ range critical and I intend to return to and improve upon this work as the UDS data push deeper and spectroscopic samples become available.

In the hierarchical formation of structure the first halos to collapse are those which eventually merge to form the most massive halos at low redshifts. This result then points towards a time sequence: passive galaxies formed earlier in those first halos while those of similar stellar mass, but still forming stars, developed in lower mass halos that collapsed later. Galaxy and halo evolution is accelerated at the earliest epochs with respect to the present day. The galaxies formed at these strongly clustered, high density peaks of the matter distribution are therefore likely to become fully evolved, passive galaxies more rapidly than the general population. Discounting mergers, the stellar mass of a galaxy is limited by the available gas reservoir so a passive galaxy could have formed significantly earlier than a star-forming galaxy, but end up with very similar stellar masses at the epoch of observation. Their respective halos, however, will have built-up mass since they first collapsed and so those of the earlier formers will be more massive. In this way the difference in clustering strength is a natural result of hierarchical mass assembly in halos and downsizing in galaxies (c.f. Foucaud et al. 2010, submitted).

In addition to the relative differences in clustering strength at fixed redshift, the

measurements also show a potential difference in the redshift evolutions of the passive and star-forming samples. The star-forming samples follow the behaviour that would be expected as a result of the downsizing scenario for galaxy evolution. Sub-samples of a given K-band luminosity (stellar mass) taken from this population exhibit a decline in clustering strength towards $z = 0$. The most luminous *passive* galaxies show tentative evidence for constant clustering strength across $0.5 < z < 1.5$. A constant value of r_0 of such magnitude indicates that the hosting dark matter halos are no less massive at lower redshift.

Direct comparison with the results of previous work is extremely difficult as the UDS is currently unique and this study is the first of its type utilising K-band selection. Qualitatively, these results agree very well with the findings of Coil et al. (2008), McCracken et al. (2008) and Meneux et al. (2006): each of these studies finds a longer clustering lengths for red or passive galaxies over blue or star-forming galaxies below $z = 1$. Furthermore, they are in agreement with previous works that any luminosity dependence on clustering in the sample is minimal in comparison with that of passivity over the range of galaxy properties that we have investigated.

McCracken et al. (2008) find that at intermediate redshifts ($0.5 < z < 1$), galaxies fit by an early-type template possibly have a inverse luminosity dependence, with less luminous galaxies having slightly longer correlation lengths. This behaviour is well known at low redshift, where passive dwarf galaxies exhibit strong clustering and are found to be associated with clusters of galaxies (Conselice et al. 2003; Zehavi 2005). The measurements presented here for red galaxies at $z < 1$ are not inconsistent with these findings.

3.8.1 Discussion of uncertainties

Throughout this work there are uncertainties that are extremely difficult to quantify. Though dust attenuation is taken into account during template fitting, there is a well-established degeneracy between the stellar age and dust content. Even with the wide range of photometry available, it is not possible to fully overcome this degeneracy. In addition, the typical redshift uncertainties are $\sim 0.05(1 + z)$ and typical uncertainties in $U - B$ colour are 0.15. The ERO definitions of Pozzetti & Mannucci (2000) can be used to examine whether the passive selection method is robust. Each of the passive objects with best-fit redshifts within the range $1 < z \leq 2$ meets the $i - K > 4$ (Vega) ERO criterion, with 1423 of the 1457 objects lying in the passive ERO (pERO) region, defined by i,J,K colours. In addition, when plotted in the $(B - z)_{AB} - (z - K)_{AB}$ plane, the conservative passive sample populates the areas in which we would expect to find passive galaxies (Section 2.1.1, Figure 3.9 and Lane et al. 2007). The SED-fit star formation properties that are used in this work therefore seem to be relatively robust, in comparison with previous broad band techniques.

Even with robust colours and selection criteria, the photometric redshift technique can introduce uncertainties into clustering measurements by causing a galaxy to be assigned to the wrong sub-sample in redshift. Passive galaxies have a strong break implied by our selection process and hence the photometric redshift determination is more reliable for these galaxies than for star-forming galaxies. There is an ongoing ESO large programme, the UDSz, to obtain ~ 4000 redshifts of galaxies at $z > 1$ in the UDS field. Early indications suggest that the photometric redshifts for the passive galaxies are indeed highly reliable, with error estimates of $\delta_z \sim 0.02(1+z)$ (though at the time of writing the majority of these are at $z < 1.75$).

Spectroscopic redshifts are particularly difficult to obtain at $z > 1.5$ and hence photometric redshifts are poorly trained at such redshift. It is precisely beyond this point that the red and blue samples converge, due in part to a decrease in the red galaxy clustering. On the other hand, the blue and star-forming samples' clustering strengths increase over this range. Sub-sample misassignment, particularly in redshift, will tend to dilute the clustering signal. For the observed convergence to be driven by errors in the photometric redshifts would therefore require the clustering signal from the red sample to be washed out, while that from the blue sample is not. This in turn would require the redshifts of the blue and star-forming galaxies to be much more accurate than those of the red galaxies. Such a case is highly unlikely, but a substantial sample of redshifts are ideally required to confirm this.

Of greater concern is the possibility that due to photometric redshift errors, star-forming galaxies are more likely to be assigned to the wrong sub-samples. The errors may then drive the difference in clustering strength that are observed between the passive and star-forming samples. A contaminating galaxy will be uncorrelated with the intended sample and so will dilute the angular clustering signal. If such dilution is significant but not accounted for in the redshift distribution, then it might be expected that the r_0 measurements for the star-forming galaxies are smaller than other comparable studies.

In chapter 2 the clustering strength of sBzK-selected galaxies for a sample of median redshift ~ 1.5 cut at $K = 23$ was found to be $\sim 6.75\ h^{-1}$Mpc. This value is similar to the star-forming sub-samples of similar redshifts presented here. Although selections in other bands are not directly comparable fig 3.4 shows that clustering strength is only weakly dependent on K-band luminosity. Any moderately bright star-forming sample can therefore be used as a comparison. Coil et al. (2008) found that star-forming galaxies at $z = 1$ have modest clustering strengths of $\sim 4h^{-1}$ Mpc, with similar values found by Adelberger et al. (2004) (BM/BX-selected galaxies, $z = 1$) and McCracken et al. (2008) ($z = 0.6$). The $z = 1$, UV-selected galaxies in Heinis et al. (2007) have a slightly larger correlation length ($4.1h^{-1}\text{Mpc} < r_0 < 5.5h^{-1}\text{Mpc}$), similar to those reported here. At similar redshifts the correlation lengths are slightly longer ($3h^{-1}\text{Mpc} < r_0 < 5h^{-1}\text{Mpc}$), but consistent, with most other studies.

A second method that can be used to test this potential issue was also used in Magliocchetti & Maddox (1999). They noted that there is a systematic error introduced into the r_0 values when it is assumed that the redshift distribution is accurately known. They take account of this error by assuming the errors in the redshifts of their sub-samples are Gaussian random and apply a co-efficient, $[(12\sigma^2/\Delta z^2)+1]^{1/2\gamma}$, to broaden the redshift distribution. In this work modelling the individual n(z) bins as top-hat functions would be inaccurate and so the r_0 measurements are not adjusted in this way. However, this method can still provide an estimate of whether the results are driven by redshift errors. Each galaxy has an associated error in redshift, derived from the χ^2 distribution, which is assumed here can be represented by a simple Gaussian. In this way each galaxy in a sub-sample has a probability of being at a given redshift and these individual probabilities are then binned and summed for all galaxies within a sub-sample. The resulting distribution is then used in place of the original n(z) during deprojection. The r_0 values in this case increase over those using the raw n(z), as expected from a broader redshift distribution. However, the difference in clustering between the star-forming and passive sub-samples remains, though difference in implied halo masses is slightly reduced.

This method is very much a simplification and the errors associated with the photometric redshifts are likely to be much more complex. In particular the errors on the star-forming galaxies' redshifts may deviate significantly from a Gaussian. Photometric redshift codes are known to favour certain redshifts, introducing artificial clumping in redshift space. The spectroscopic redshifts obtained from the UDSz will be vital in understanding possible biases arising from using photometric redshifts. The conclusion from the tests outlined above, however, is that errors in the redshift determination are unlikely to be responsible for the results that have been found.

A further, and potentially important source of error, is that of cosmic sampling variance. In contrast to the previous chapter, the effects of sampling variance are not taken into account in this work. The method employed in Chapter 2 consisted of splitting the field into 4 and computing a measurement for each sub-field. This method provides a estimate of the sample variance for a field only one quarter of the size of the original. The ideal method for estimating the sample variance is to repeat these measurements for comparable, independent fields. To facilitate such work in the future the samples were re-defined, using simple photometric selection in Section 3.6. However, it should also be noted that the clustering of sub-samples at different redshifts are largely uncorrelated, yet consistent differences between passive and star-forming galaxies are found for subsamples with $z < 1.5$. It is unlikely that such consistent disparity in the clustering results is due to cosmic sampling variance alone.

Finally, the difference in clustering that has been found between the star-forming and passive galaxies may have been exagerated by multiple halo occupation. If passive galaxies have a higher probability of multiple occupation by a single halo, then the lower limit of the fitting range ($0.3h^{-1}$Mpc) may be insufficient to entirely avoid occupation biases. In figure 3.13 the correlation length measurements are repeated with different fitting ranges: 0.5Mpch^{-1} - 0.5 deg. for the upper plot and $1.0h^{-1}$Mpc - 0.5 deg. for the lower. Though the measurements become less certain due to the poorer data used in fitting, it is clear that the results for the star-forming galaxies are extremely robust. The measurements for the passive galaxies are less certain due to the reduced statistics, however, there does seem to be an overall reduction in the r_0 vlaues. Despite this reduction, these alternative values are within the errors on the original measurements and are entirely consistent with all of the conclusions drawn throught this chapter.

3.9 Conclusions and unanswered questions

Using the deepest contiguous near infrared survey currently available over an area of ~ 0.7 deg^2 I have investigated the clustering strengths of galaxies over the range $0 < z < 3$. Best fitting templates, including a dust contribution, have been measured for each galaxy. From these templates the rest frame $U-B$ colour, absolute K-band magnitude, stellar age and e-folding time of star-formation, τ, are derived. These quantities are then used to define 'red' and 'blue' galaxies, a conservative passive sample and active star-forming sample.

Each of these samples are then further sub-divided by redshift and K-band luminosity. Angular correlation function measurements are made for each of these sub-samples and power laws are fit over the scales $\theta_{lim} < \theta < 0.2$ degrees, where θ_{lim} corresponds to $0.3h^{-1}$Mpc. This limited fitting range is used so as to avoid the bias

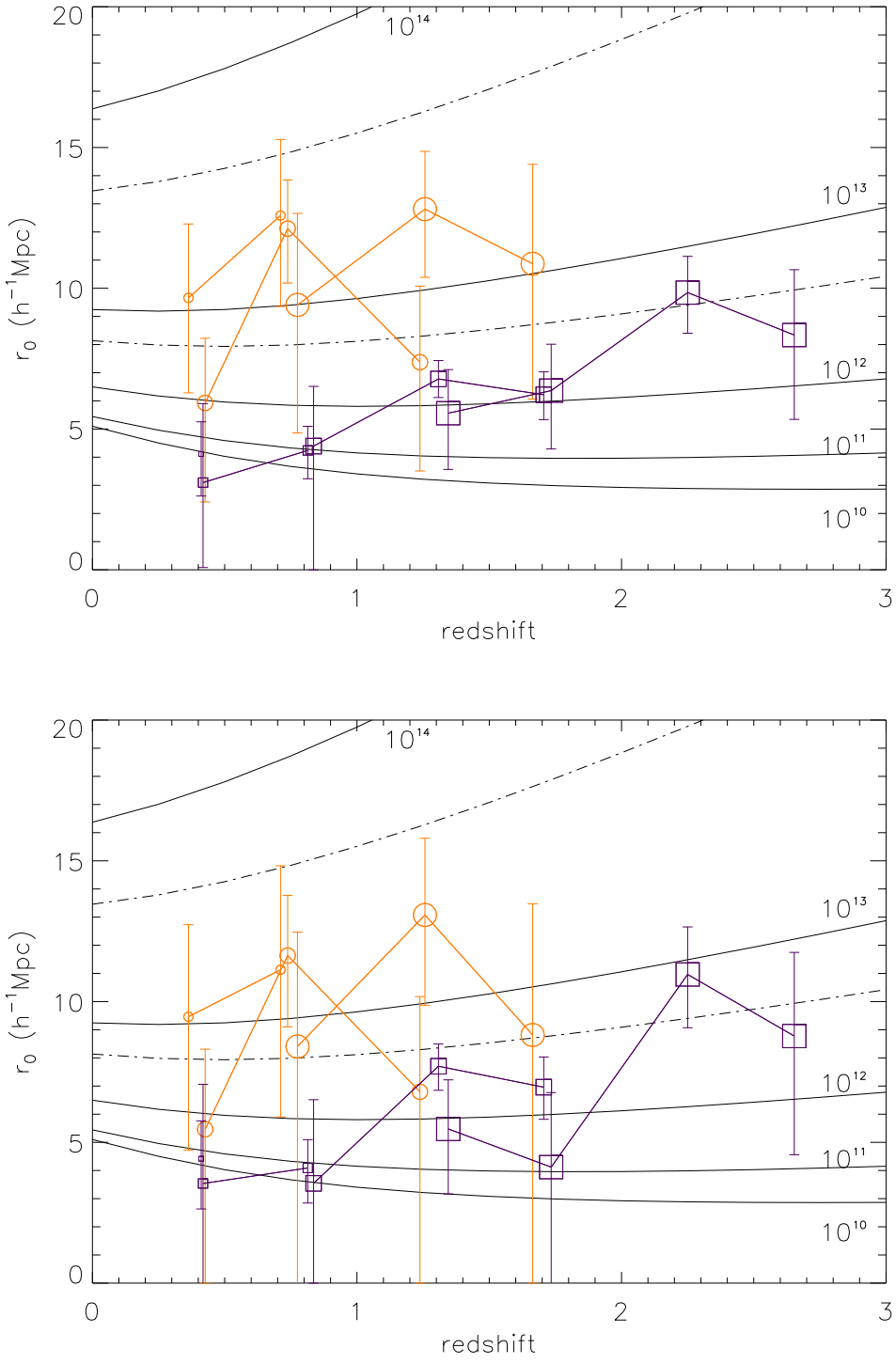

Figure 3.13. Clustering lengths for the conservative passive sample and star-forming sample have been re-computed using two different fitting ranges to investigate possible biases due to multiple halo occupation. In the upper plot the fitting range used was 0.5Mpch^{-1} - 0.5 deg.; while in the lower plot it was 1.0Mpch^{-1} - 0.5 deg..

caused by multiple halo occupation (on small scales) and to avoid the scales where the correlation function and integral constraint correction become similar. Using photometric redshift distributions, the angular correlation functions are deprojected to find real space correlation lengths.

The results show that the conservatively-defined passive sample is more strongly clustered than the star-forming sample for all luminosities at $z < 1.5$. Above this redshift the clustering strengths appear to converge, though data of greater depth or area are required to improve the statistics before it is possible to make a firm conclusion. Furthermore, by relaxing the strict passive galaxy selection I have shown that passive galaxies are the most strongly clustered of the 'red' galaxies.

Clustering strength is intimately related to the typical mass of dark matter halo hosting a sample, and K-band luminosity is a reasonable proxy for the stellar mass of a galaxy. Passive galaxies of a given stellar mass are therefore hosted by more massive halos at $z < 1.5$. At higher redshift the convergence of clustering strengths suggests that the red sequence is in the early stages of being populated.

I have also studied how the clustering of passive and star-forming sub-samples compares with those of dark matter halos of varying mass. The clustering strengths of star-forming galaxies decline towards $z = 0$ for a given luminosity (stellar mass), and hence are typically found in less massive dark matter halos. This finding is consistent with the wealth of evidence supporting downsizing.

The UDS imaging project is ongoing, and expected to gain at least 1 magnitude in J,H,K depth by 2011, and potentially substantially deeper. This will enable the extension of this work to higher redshift and lower luminosities. In addition, the ongoing spectroscopic programme (UDSz) will provide several thousand spectra over the redshift range probed by this analysis. This will enable a variety of more detailed studies, in addition to substantially improving the reliability of photometric redshifts. The addition of similar deep fields from upcoming surveys (e.g. UltraVISTA) will also allow robust estimates of the sample variance. I therefore anticipate major progress in our understanding of the galaxy populations at the crucial epoch when the red sequence first becomes established.

In the meantime the results from this chapter provide vital data for the further development of theoretical models of galaxy formation, such as those used to construct semi-analytic galaxy formation models. A reflection of current models in the light of these data will be presented in Chapter 5. In addition, the results presented here allow us to form a hypothesis which may be testable with another, related technique. This hypothesis is that passive and star-forming galaxies are equally clustered at high redshift ($z > 2$), but become distinct at lower redshifts. In the next chapter I shall use a marked correlation function analysis of these same galaxies and attempt to test this hypothesis.

Chapter 4

Marked-correlation functions

In this chapter I shall test some of the conclusion of the previous chapter through the use of a related technique, the marked correlation function. This work constitutes a preliminary investigation and will require a number of tests beyond those presented here before it is ready for publication.

4.1 Introduction to marked-correlation functions

Weighted, or 'marked', point processes were developed in statistics studies some time ago (Stoyan 1984; Stoyan & Stoyan 1994), but have had seen little use in astronomy until recently. Alimi et al. (1988); Boerner et al. (1989) and Valls-Gabaud et al. (1989) were the first authors to apply weighted statistics to the study of galaxy properties, with each choosing luminosity as their mark of interest. However, it wasn't until Beisbart & Kerscher (2000) attempted to disentangle the dependencies of luminosity and morphology on galaxy clustering that marked correlation functions started to gain more widespread use. The abundance of high quality survey data (e.g. York 2000), detailed models of the evolution of galaxies (Somerville et al. 2001; Springel et al. 2005; Bower et al. 2006; Croton et al. 2006; De Lucia & Blaizot 2007; Font et al. 2008) and the halo model theoretical framework (Neyman & Scott 1952; Cooray & Sheth 2002) have all contributed to the development and successes of marked correlation statistics. Throughout much of this chapter I shall follow the notation and conventions laid out by Skibba et al. (2009), modified for use with angular correlation functions.

The auto-correlation function is routinely used to investigate how the clustering of galaxies depends on a quantity of interest (e.g. luminosity, colour). Typically this is performed by splitting the sample into subsamples and computing the correlation function of each subsample. Doing so can easily result in small samples (exemplified by the previous chapter) and hence large statistical errors. Furthermore, there is a loss of information in assigning objects to sub-samples in this way. In a sample where the clustering strength is only enhanced for the objects with the most extreme values for a mark, the dependence on the quantity may be missed. In a marked correlation function splitting into subsamples is often unnecessary and, if a continuous mark is used, there is no loss of information in the quantity of interest.

Transforming the auto-correlation function to a weighted correlation function requires a single, simple step: each pair is counted by the product of its normalised marks rather than as a single pair. This leads to,

$$W(\theta) = \frac{\Sigma W_i W_j}{\Sigma R_i R_j} - 1 = \frac{WW(\theta)}{RR(\theta)} - 1 \tag{4.1}$$

where W_i and W_j are values for the quantity of interest of a galaxy pair separated by angle θ and normalised by the mean of all sample galaxies. WW, analogous to DD in equation 1.3 (Chapter 1, Section 1.3.1), is the resulting weighted pair count. If the W_i are equal for all galaxies, then the weights are each equal to 1 and equation 4.1 reduces to the natural estimator for the angular auto-correlation function. Clearly, the strength of clustering on a particular scale will reflect how likely a pair of galaxies with high values for a mark are to be found at such a separation. However, the galaxies themselves are clustered and so to allow more straightforward interpretation of the measurements the marked correlation function is defined as

$$M(\theta) = \frac{W(\theta) + 1}{w(\theta) + 1}. \tag{4.2}$$

M(θ) is therefore the clustering of the marks themselves, rather than the galaxies. Using the natural estimator for $w(\theta)$ equation 4.2 becomes

$$M(\theta) = \frac{WW}{DD}, \tag{4.3}$$

with the random-random pair counts cancelling out, provided the weighted and unweighted selection functions are the same. This property of the marked correlation function makes it very fast to estimate. As the auto-correlation function, $w(\theta)$, and weighted correlation function, $W(\theta)$, are affected in the same way by the survey geometry and edge effects, the marked correlation function is more robust than the equivalent auto-correlation function. However, in practice a more robust estimator, such as that suggested by Landy & Szalay (1993), is usually preferred, with the weighted correlation function given by,

$$W(\theta) = \frac{WW - 2WR + RR}{RR}. \tag{4.4}$$

Similar to the auto-correlation function (c.f. Chapter 1, section 1.3.1) WW, WR and RR are normalised by their sums over all scales. However, in this case the random galaxies receive a random mark taken from the distribution of data marks in addition to a random position. The marked correlation function is then computed by $(W(\theta)+1)/(w(\theta)+1)$ as before. It should be noted here, however, that the integral constraint used to correct auto-correlation function measurements is neglected. A correction of this type requires that we know the amplitude and power-law slope of the correlation function and we do not wish to assume either. However, the effects of the finite field size cancel out to some extent as it influences both the weighted and unweighted correlation functions. For a field the size of the UDS this slight bias is very small. An example is given in the appendix.

Using the above form of the marked correlation function, Skibba et al. (2009) investigated the environmental dependence of colour and morphology from the galaxy zoo project (Lintott et al. 2008). In agreement with previous work using auto-correlation functions, they found that red populations and early type galaxies were found in denser environments and more highly biased with respect to the overall galaxy population. However, they then further examined colour-marked correlation

functions at fixed morphology and morphology-marked correlation functions at fixed colour. They found that morphology was relatively insensitive to environment, with much of the morphology-density dependence, known since Dressler (1980), the result of a colour-density correlation.

This striking result was the latest of several similar results to have come out over the last few years. Skibba & Sheth (2009) found both luminosity and colour dependent clustering in the SDSS galaxy sample, while the technique has also successfully been applied to the halos of dark matter simulations. Wechsler et al. (2006) confirmed the result of Sheth & Tormen (2004); Gao et al. (2005) by using a formation time mark in a study of dark matter halo clustering. Using the Millennium simulation, Gao et al. (2005) had earlier found that the oldest 20% of halos within a given mass range were significantly more clustered than the youngest 20% of halos of the same masses. Using a marked correlation function, Wechsler et al. (2006) generalised the result for a continuous distribution of formation time. They found a clear dependence of clustering on both of the measures of halo formation that they used (concentration and normalised formation time). This dependence was found to be stronger for lower mass halos. Harker et al. (2006) also used marked correlation functions to study dark matter halo formation. However, their analysis focused on the environmental biases of formation time, finding that halos in denser environments were formed earlier.

4.2 Choosing marks

When computing a marked correlation function, we are essentially free to choose whichever mark we please. However, the quantities of interest, for example colour or stellar age, may have very different ranges of values. How can we determine which of these is the more significant mark? The significance (or otherwise) of a mark may be determined in two ways. The first is by bootstrap resampling, similar to the method used to find error estimates in the auto-correlation function. The re-sampled galaxies have both a position and a mark and so should reproduce the variance of the marks and their clustering reasonably robustly. The second method is to scramble the marks by assigning them to each galaxy randomly and repeating many times. In this case we obtain an estimate of the variance a marked correlation function may have when there is no dependence on the mark. Both methods are commonly used in the literature but bootstrap resampling is preferred in this work. If a mark is not correlated with the environment at a given scale, then a value of unity is recovered. So, a degree of significance can be estimated based on the deviance from unity and the bootstrap errors.

However, a significance level may not be sufficient, and a more direct comparison between different marks may be desired. To what extent is it reasonable to rescale a mark such that two different marks occupy the same range of values? This question has already been approached by Skibba et al. (2009). They found that the results of the marked correlation function are relatively insensitive to rescaling, either by multiplication or by taking logs, so long as the marks are positive definite. A test of this assertion using the pBzK and sBzK samples from chapter 2 is presented in the appendix. The results of that test were found to be broadly in agreement with the statement from Skibba et al. (2009). The marks used in the remainder of this chapter are therefore scaled to lie within the range $0 < W < 1$, to facilitate comparison between different marks. The dependence on the shape of the distribution is more

complex, however, and further testing is required to quantify any effects.

4.3 Results

In this section I shall test the conclusions of the previous chapter using marked correlation functions of the same galaxy sample. Specifically, the relevant conclusions were:

- ⊙ Red galaxies cluster more strongly than blue galaxies regardless of K-band luminosity at $z < 1.5$.
- ⊙ Of the red galaxies, passive galaxies are the more strongly clustered population, with approximately constant clustering strength over the range $0 < z < 1.5$.
- ⊙ The combination of downsizing in the star-forming galaxies and the constant clustering strength as a function of redshift (possibly even declining above $z = 1.5$) of passive galaxies suggests that the clustering of the two populations will become indistinguishable by $z \sim 2.5$.

These conclusions can be used to make predictions regarding the behaviour of marked correlation functions:

- ⊙ A mark of $U - B$ colour will exhibit stronger mid-to-large-scale environmental dependence than a mark based on K-band luminosity (scaled to the same range in marks, $0 < W < 1$).
- ⊙ This first prediction will cease to be the case beyond $z \simeq 1.5$.
- ⊙ A passivity-marked correlation function will exhibit a stronger large-scale environmental dependence than a $U - B$ MCF.
- ⊙ A mark reflecting 'passivity' will exhibit an environmental dependence even if only red galaxies are used as the galaxy sample.
- ⊙ The passivity-marked correlation function of the combined red and blue sample will be consistent with $M(\theta) = 1$ at mid-to-large scales by $z \simeq 2.5$.

In each of the following two sub-sections, two of these predictions shall be examined, beginning with the comparison of $U - B$ colour and K-band luminosity.

4.3.1 Colour or luminosity?

Following the formalism of Skibba et al. (2009) described in the first section of this chapter, marked correlation functions were computed using firstly $U - B$ colour as the mark (the distribution of which is shown in Chapter 3, Figure 3.5), and then K-band luminosity. As in the previous chapter, the requirement of volume-limited samples was maintained, but no further luminosity binning was used. The luminosities were mapped to marks by firstly taking the modulus and then scaling them, such that the volume-limiting luminosity had a mark $M = 0$, and a luminosity of $M_K = -28$ corresponded to a mark of $M = 1$. Only three objects in the sample have $M_K < -28$, and two of these lie at very high photometric redshift ($z \sim 5.5$). In fact these redshifts are likely to be incorrect and therefore cannot be reliably

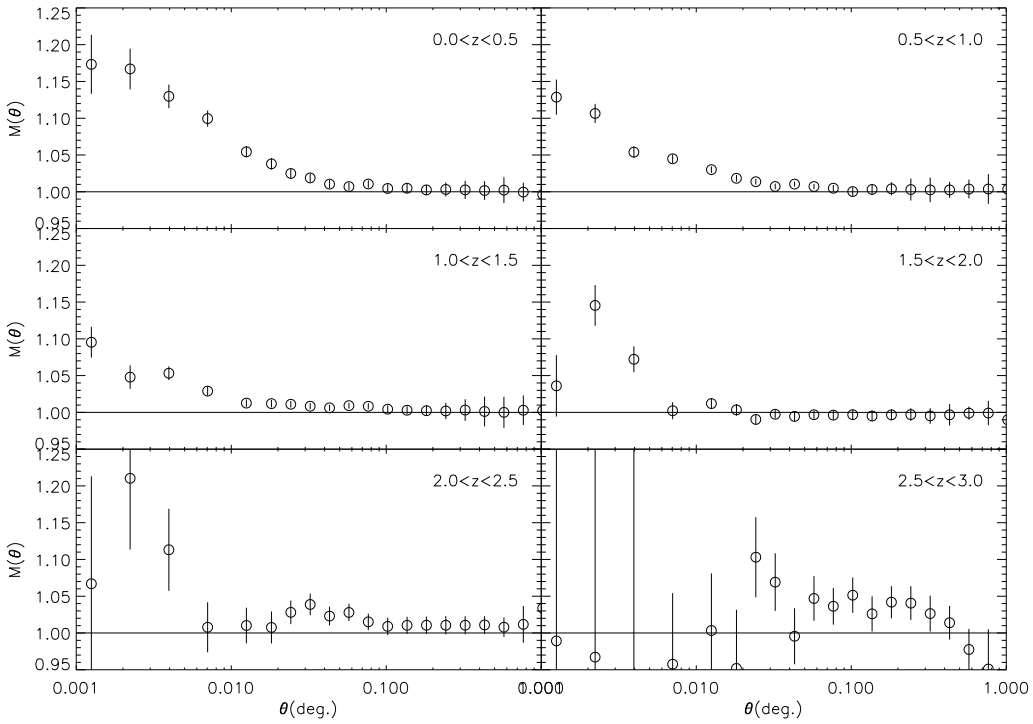

Figure 4.1. Rest-frame $U - B$-marked correlation functions of different redshift sub-samples. Each sub-sample is volume limited (see chapter 3), but there is no further separation by K-band luminosity. A value of $M(\theta) > 1$ indicates that the mark property ($U - B$ colour in this case) is correlated with the environment at that scale.

assigned to any sample. Mapping luminosities to marks in this manner had the effect that the mark range diminishes as redshift increases. It is therefore expected that, regardless of any intrinsic redshift evolution in the clustering properties of the galaxy sample, the luminosity-marked correlation function will become suppressed at higher redshifts. However, the same is also partially true for the $U - B$-marked correlation function. The number of red galaxies diminishes at higher redshifts and there is also mild evolution in the colour of the red sequence towards bluer colours.

Splitting the sample by K-band luminosity as well as redshift is no longer required nor desired. The sub-samples therefore have greater numbers in this work, and the redshift bin widths can be slightly reduced (though they must remain significantly wider than the photometric redshift error estimates). In order to check consistency with the results of the previous chapter, however, the $U - B$-marked correlation function is initially computed using $\Delta z = 0.5$. Figure 4.1 shows these marked correlation functions for samples in the range $0 < z < 3$.

The lowest redshift panel of Figure 4.1 is similar to the results found by Skibba et al. (2009) from the SDSS. There is a clear environmental dependence on the mark at small and mid scales. The measurements are significantly elevated above unity out to 0.1 degrees, which corresponds to $\sim$ 2Mpc (physical) at $z = 0.5$ in the assumed cosmology. While the halo occupation effects of massive clusters could plausibly extend that far, it should be noted that the SXDS/UDS field was chosen to avoid obvious low-redshift structure of this sort. The measurements in the following

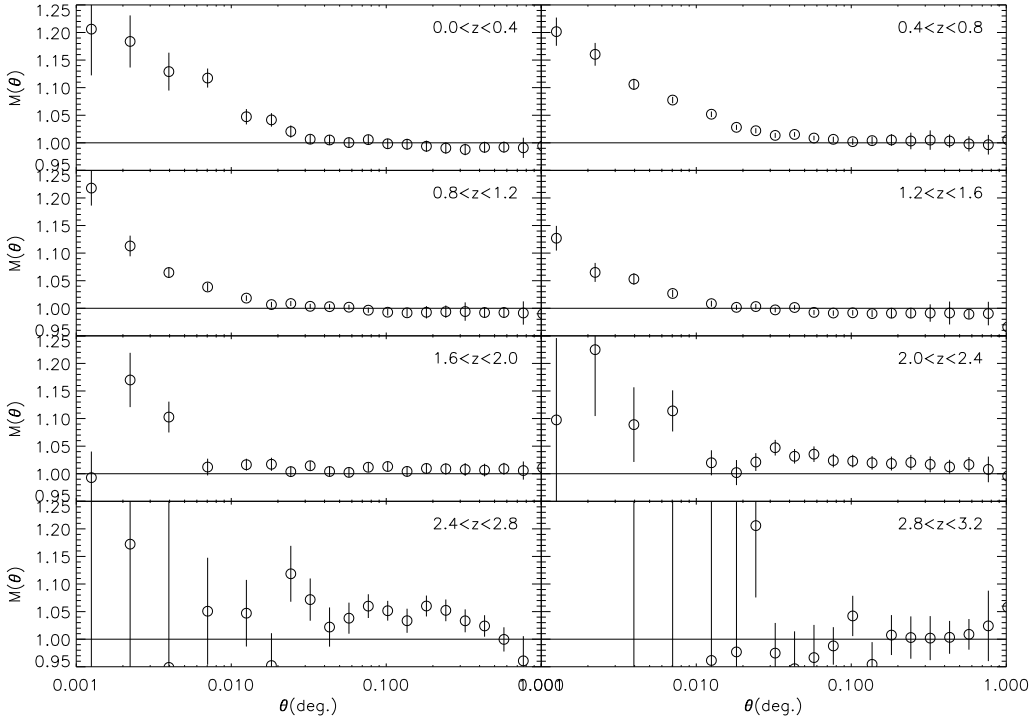

Figure 4.2. Same as Figure 4.1 but with a finer binning in redshift. A finer binning is preferable as structure gets increasingly washed out as bin widths are increased.

two panels, out to $z = 1.5$, show very similar behaviour, though the small scale dependence appears to be smaller. It is unfortunately not possible to compare two panels of different redshifts directly as the volumes they are probing are not equal. To do so will require development of this technique within the halo model (see chapter 5), or a large spectroscopic data set, such as the ESO large programme, UDSz, which will become available shortly. However, by $z = 2$ the only points significantly above unity are those at scales corresponding to the single halo regime. Taken at face value, this would suggest that the bias of galaxies is independent of rest-frame $U - B$ colour at $z > 1.5$, but that redder galaxies are more likely to be found in the same dark matter halos. The behaviour described here corresponds very well with the results of the previous chapter (c.f. Figure 3.6). Beyond $z = 2$ the measurements become increasingly uncertain, though there does appear to be an enhancement at small scales out to $z = 2.5$.

As noted previously, by not splitting samples by K-band luminosity, it is possible to reduce the redshift bin widths whilst retaining a reasonable number of objects. The measurements displayed in Figure 4.2 have been computed in bins of $\Delta z = 0.4$. The conclusions which can be drawn from these measurements are broadly in line with those using a coarser binning in redshift. Where $M(\theta)$ deviates from unity, it is typically slightly higher than the closest corresponding point from the first set of results. This is simply the effect of reducing the bin widths. As a greater radial distance range is used, the uncorrelated galaxies at different redshifts will wash out the clustering signal. This issue was also of relevance in chapter 2, where the broadening of the BzK-selected galaxy redshift distributions had to be accounted for

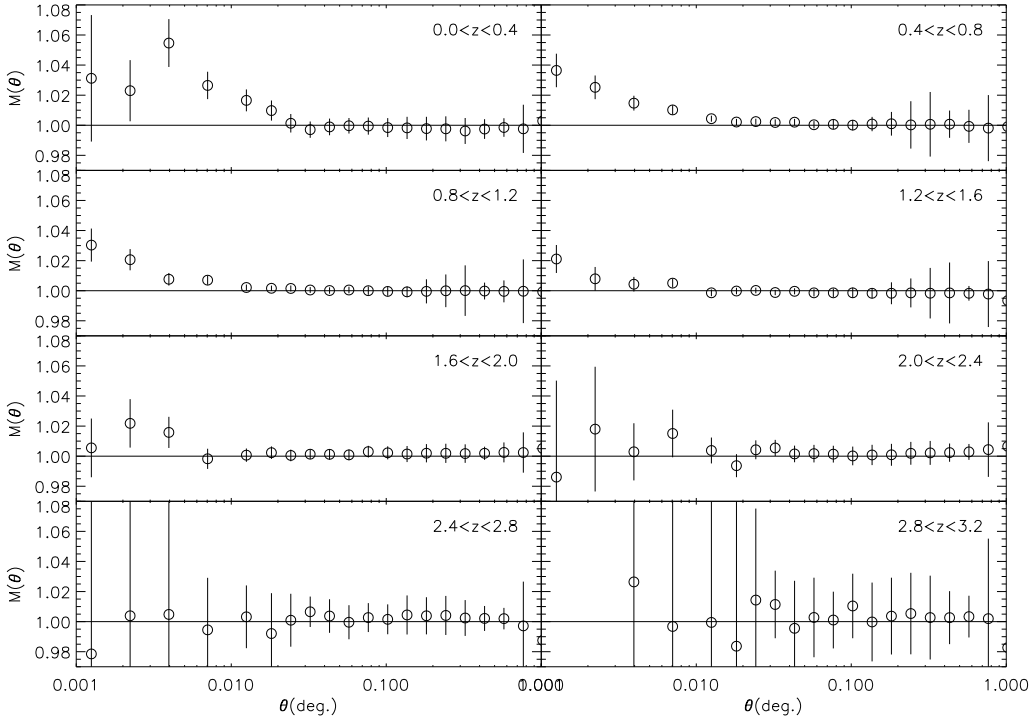

Figure 4.3. Rest-frame K-band luminosity-marked correlation functions. The typical amplitudes of $M(\theta)$ are substantially smaller than those of the $U-B$-marked correlation functions.

when deprojecting $w(\theta)$. Due to this washing out effect, it is desirable to minimise the redshift bin width as far as is reasonable. A width of $\Delta z = 0.4$ is used in the remainder of this chapter.

The next set of measurements were made using a mark derived from the rest-frame K-band luminosity as described above. These results are shown in Figure 4.3. Though they share a similar form with the $U-B$-marked correlation functions at lower redshifts, any luminosity dependence on clustering is too small to reliably measure with the current data set above $z = 1.6$. Furthermore, the amplitudes of the K-band luminosity-marked correlation functions are much smaller than those using $U-B$ colour marks in the same redshift range. The first prediction made from the results of chapter 3, that $M(\theta)$ will be greater for $U-B$ marks than for K-band luminosity marks at mid to large scales (scales corresponding to $> 1\mathrm{h}^{-1}\mathrm{Mpc}$), is true in the first two redshift bins ($z < 0.8$). However, at $0.8 < z < 2.0$ both sets of data are barely above unity at mid and large scales. At higher redshifts ($z > 2$), $U-B$ colour appears to be correlated with large-scale environment, however the measurements become increasingly uncertain at such distances.

The situation is substantially different at small scales. Colour has a strong dependence on small-scale environment out to $z \sim 2.4$ while the environmental dependence on luminosity has all but disappeared by $z \sim 1.6$. This absence of a correlation is partly due to the lack of discriminating range in luminosity. However, for reasonably bright galaxies ($M_K < -23$) at high redshift ($1.6 < z < 2.4$) galaxies redder in $U-B$ are more likely to be found in regions of higher galaxy density.

4.3.2 Passivity as a mark

While $U-B$ colour and K-band luminosity are straightforward marks to use, it is not clear what an appropriate mark could be to reflect the 'passivity' of a galaxy. Two intuitive choices that would initially seem reasonable are the time since the last significant episode of star-formation and the specific star-formation rate (SSFR). Unfortunately neither of these choices are useful with the data in hand. The star-formation history parameters that have been computed for each galaxy are the best fit to the entire composite stellar population. Discriminating multiple episodes of star-formation is simply beyond reasonable for the data available. The passive samples in chapter 3 were defined in part by requiring the current SFR, implied by the best fit stellar age and τ parameter, to be below some fraction of the initial SFR. This quantity is very closely related to the SSFR and a reasonable estimate of the current SSFR could be made from the data, but is not a sufficient criterion. The reason why it is not sufficient is that many of the sample galaxies are best fit by a burst or small value for τ and a stellar age of only a few hundred Myrs. The implied SSFR for these galaxies at the time of observation is therefore zero, but they are not the desired population. To separate such galaxies from the red and passive sample that is desired, additional colour and stellar age criteria were used. For the conservative passive sample, however, galaxies which satisfied both the SFR and stellar age criteria also satisfied the colour criterion. Rather than the use the SSFR then, a mark was constructed from the best-fit star formation histories to reflect the level of star formation over the last 1Gyr.

The star-formation rate in units of the initial star-formation rate, SFR_i, is given by

$$SFR = SFR_i \; e^{-t/\tau}. \tag{4.5}$$

The mass of stars in units of SFR_i formed over the last Gyr is therefore

$$M_{stars} = SFR_i \int_{t_0-1}^{t_0} e^{-t/\tau} dt, \tag{4.6}$$

where t_0 is the time at observation. However, as SFR_i is unknown, a passivity mark, M_p, is instead defined as M_{stars}/SFR_i. Performing the integration, this becomes

$$M_p = \tau \left(e^{-(t_0-1)/\tau} - e^{-t_0/\tau} \right). \tag{4.7}$$

Had we integrated over the age of the galaxy we would recover the SSFR. If the star-formation history of the galaxy is shorter than 1Gyr then the mark may take values of $M_p > 1$, which are undesired. For galaxies with age $<$ 1Gyr then, the integral is taken over their entire age and scaled up to 1Gyr:

$$M_p = \frac{\tau}{age} \left(1 - e^{-t_0/\tau} \right) \qquad (\text{age} < 1\text{Gyr}). \tag{4.8}$$

In addition, to avoid infinities, galaxies fit by burst templates are given $\tau = 0.05$ (half the value of the lowest τ value otherwise allowed by the templates). These two equations give a well behaved mark distribution between zero and one. Finally, for ease of comparison with the other marks presented here, the marks are inverted: $M = 1 - M_p$, so that values of $M(\theta)$ are expected to be greater than unity. This mark distribution is shown in Figure 4.4.

The limited number of values for the stellar age and SFR decline, τ, which a

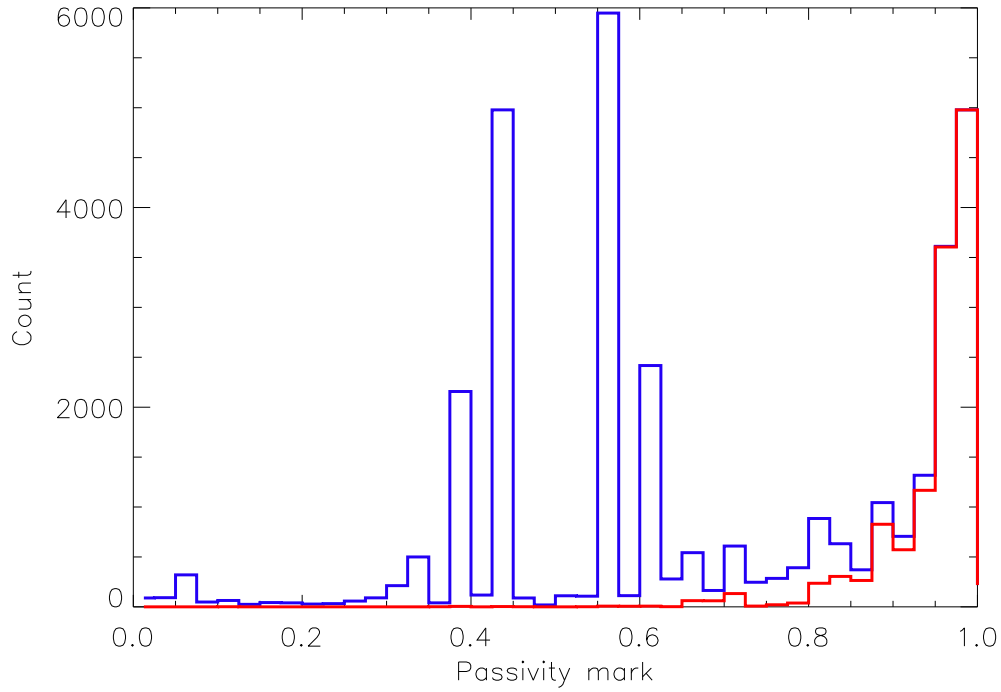

Figure 4.4. Distribution of the the 'passivity' mark defined in the text for the full sample (blue) and the red sample defined in chapter 3 (red). The combinations of discrete values for age and τ cause a small number of values to be strongly favoured.

galaxy can take, causes the mark distribution to be a series of separated discrete values rather than a continuous range. Furthermore, a relatively small number of these values are strongly favoured. The resulting distribution is not ideal and makes direct comparison with the $U-B$ and K-band luminosity marks more difficult. In particular, there are very few objects with marks $M < 0.3$. The discriminating range of marks is therefore slightly reduced, which means that the $M(\theta)$ values will be slightly suppressed relative to the $U-B$-marked correlation function. As previously stated the luminosity-marked correlation function suffers a similar problem, due to the rapid decline in galaxy numbers brighter than L^*.

Nevertheless, the passivity-marked correlation functions would still be expected to display environmental dependence due to the significant difference in clustering strengths of passive and star-forming galaxies (chapter 3). These MCFs are shown in Figure 4.5. Their shapes are very similar to those found using a $U-B$ mark, but at $z < 0.8$ the small scale amplitudes are smaller. At $z > 0.8$ there is no significant difference between the two, with the possible exceptions of the rather uncertain measurements at $2.0 < z < 2.4$ and $2.4 < z < 2.8$. At these redshifts the contamination of the red population by dusty star-forming galaxies is expected to be more severe, and so a greater difference between passivity and $U-B$ colour-marked correlation functions may be anticipated. With regard to the third prediction above, however, the measurements are inconclusive. The final prediction is also in doubt, but larger samples are ideally required to confirm or deny it.

The distributions of the $U-B$ colours and passivity mark are substantially dif-

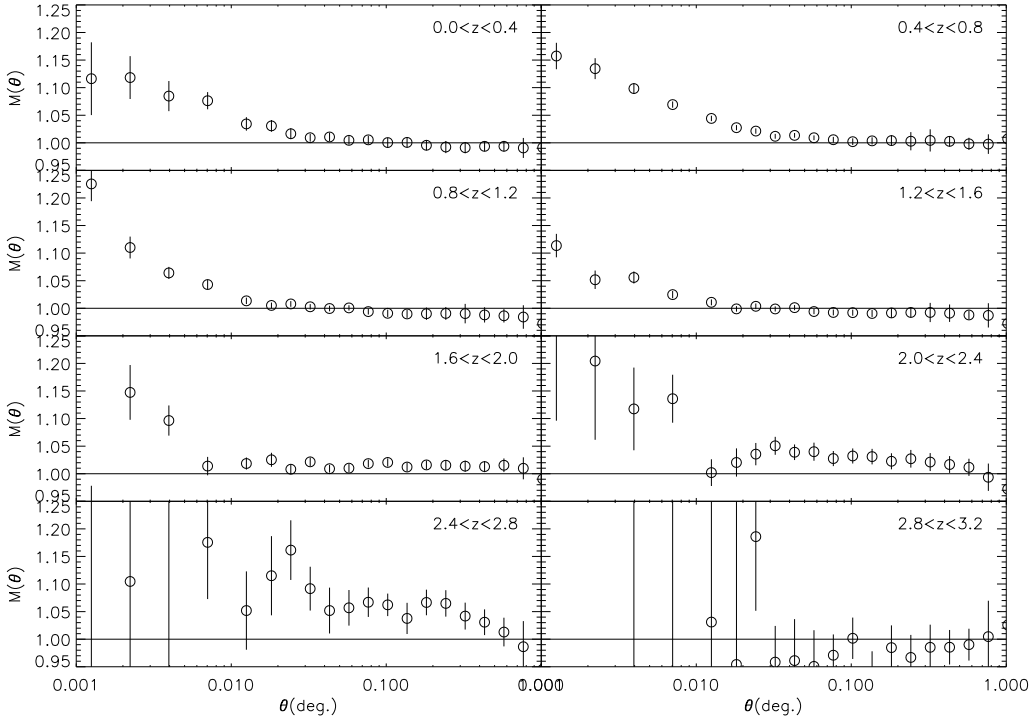

Figure 4.5. Marked correlation functions using the 'passivity' mark defined in the text. The amplitudes of $M(\theta)$ are slightly smaller than those using a $U-B$ colour mark for $z < 0.8$, especially at small scales. Above $z = 1.6$ the large-scale $M(\theta)$ measurements are marginally higher than those of the $U-B$ MCF.

ferent. The low-redshift, small-scale effect could therefore be due to the differences in mark distributions, but a great deal of testing on mock galaxy samples would be required to quantify their influence on $M(\theta)$. Such testing is beyond this preliminary investigation. However, if passivity really is more strongly correlated with environment than colour is, it should be noticeable in a passivity-marked correlation function of red galaxies. This approach is taken below.

Passivity-marked correlation functions were computed from the red sample, which was defined in chapter 3. These are shown in Figure 4.6. Contrary to the fourth prediction at the start of this section, there is apparently very little correlation with environment. There are some marginally significant points at small or mid-scales at $z < 1.6$. However, the clear influence of the SFR criterion on the values of r_0 that was found in chapter 3 is not reflected in these measurements. What possible explanation could there be for this behaviour?

The first possibility is that the measurements are a representation of the real physical state. If we are to take these results at face value then they imply that colour is a more fundamental indicator of the environment that a galaxy is in than its star-formation history. Aside from the fact that this interpretation is counter to the conclusions of the previous chapter, it is not particularly physically intuitive. While both passive galaxies and dusty star-forming galaxies are often associated with dense environments (the latter possibly caused by a recent interaction), why should the colour of a dusty galaxy correlate with environment in the same way as

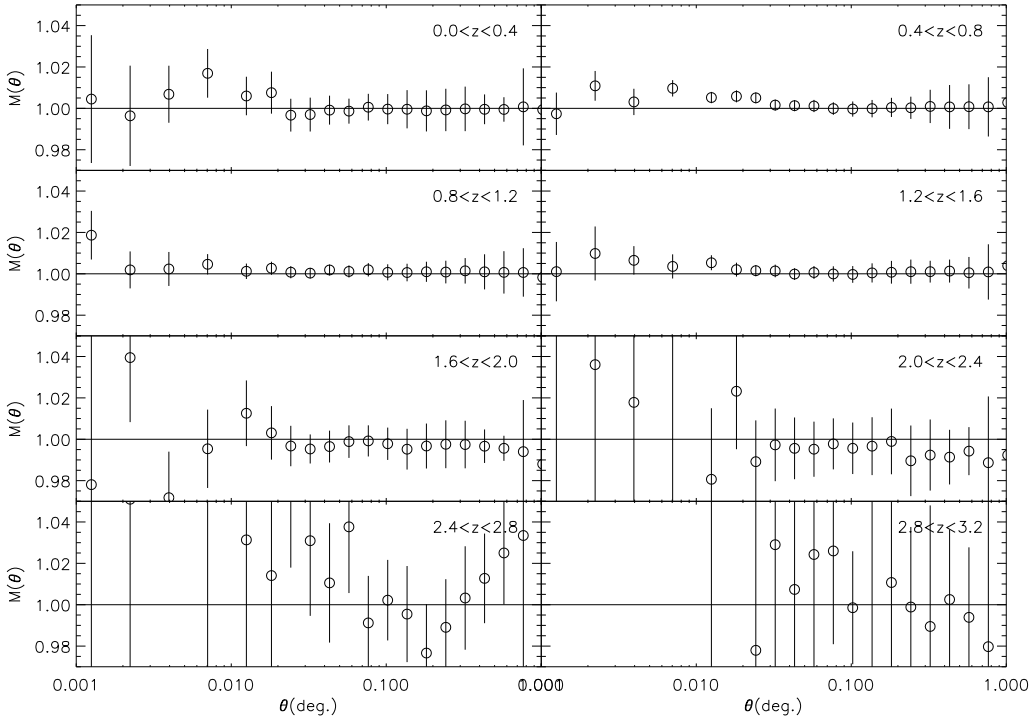

Figure 4.6. Marked correlation functions using the 'passivity' mark defined in the text within the red sample.

a passive galaxy?

Perhaps instead it is a fault of the photometric redshift fits and derived properties. There are considerable uncertainties in the photometric redshifts, and even greater uncertainties in the star-formation histories that are implied by the template fits. It is possible that these fits are unable to sufficiently distinguish the SFHs of red galaxies. While such uncertainties are a concern, the measurements of the multiple passive samples in chapter 3 suggest that, statistically at least, the SFHs are reasonable. Finally, it may simply be that the passivity mark constructed here is too insensitive to pick out the subtle environmental effects at work. I would suggest that this is the most likely explanation, or at least the dominant effects at work. In the following, an attempt to refine the passivity mark for use with the red sample is described.

As can be seen in Figure 4.4, there are very few galaxies in the red sample with a passivity mark below $M = 0.5$. It is possible that the range of marks allowed is inappropriate to the task of differentiating red passive from red star-forming environment correlations. By restricting the mark range to $0.5 < M < 1.0$ and rescaling such that these values now occupy the range $0 < M < 1$, perhaps more subtle effects can be highlighted. There are eleven objects which are in the red sample but have passivity marks $M < 0.5$. As these marks imply a negative mark upon being rescaled, these objects were discarded. Six of these objects lay in the $2.4 < z < 2.8$ bin, while the others were evenly distributed among the lower redshift samples. It is therefore not expected that these discarded objects should severely bias the new measurements below $z = 2.4$. The MCFs computed using these

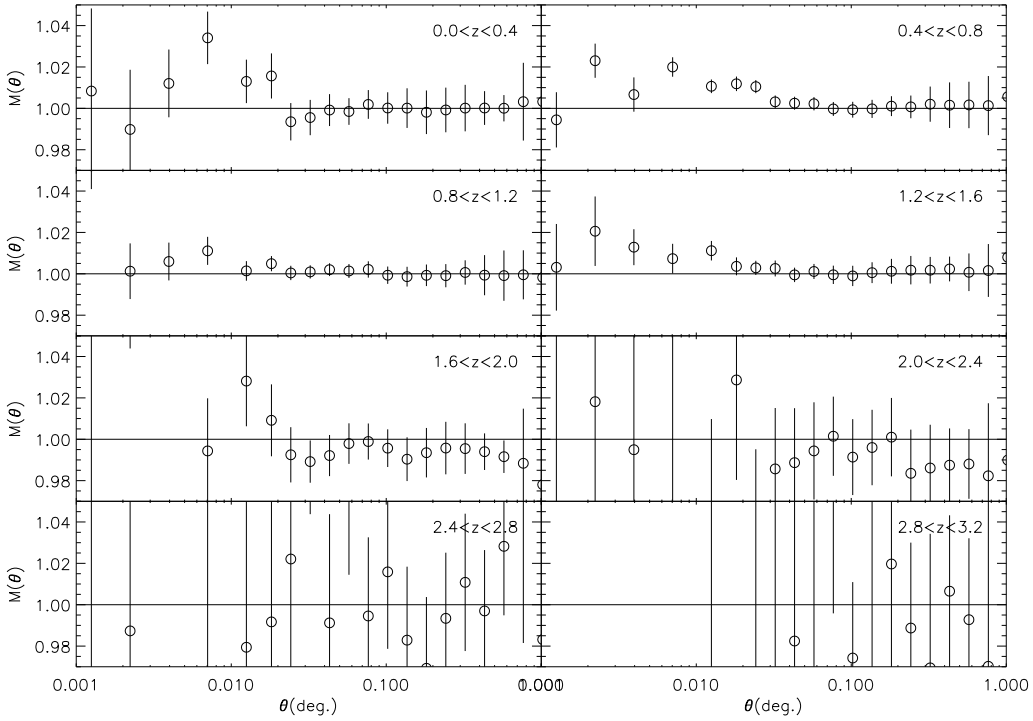

Figure 4.7. Marked correlation functions using the rescaled passivity mark (see text). Though there is some enhancement of the measurements, particularly at lower redshift, the environmental dependence of passive within the red population remains largely inconclusive.

rescaled marks are shown in Figure 4.7. Rescaling the marks has indeed brought out some of the more subtle features. It appears from these measurements that at $z < 1.6$ there is an environmental dependence on passivity. But any dependence is very small and larger data sets would be required for a firm conclusion.

4.4 Conclusions

In this chapter I have tested a number of predictions which were based on the results of chapter 3. These were

- ⊙ A mark of $U - B$ colour will exhibit stronger mid-to-large-scale environmental dependence than a mark based on K-band luminosity.
- ⊙ This first prediction will cease to be the case beyond $z \simeq 1.5$.
- ⊙ A passivity-marked correlation function will exhibit a stronger large-scale environmental dependence than a $U - B$ MCF.
- ⊙ A mark reflecting 'passivity' will exhibit an environmental dependence even if only red galaxies are used as the galaxy sample.
- ⊙ The passivity-marked correlation function of the combined red and blue sample will be consistent with $M(\theta) = 1$ at mid-to-large scales by $z \simeq 2.5$.

The first of these was found to be true for $z < 0.8$, in agreement with the body of literature finding a stronger dependence on colour than luminosity in auto-correlation functions of galaxy samples (e.g. Coil et al. 2008). However, at higher redshifts ($z > 0.8$) it becomes difficult to reliably discern large-scale enhancements in either the colour or luminosity-marked correlation functions. Despite this, the second prediction is also somewhat uncertain. There appears to be a correlation between large-scale environment and $U - B$ colour at $z > 2.0$, while there is none with rest-frame K-band luminosity. Deeper data and more reliable estimates of the rest-frame $U - B$ colour will become available in the future, and they may help to confirm or deny one or both of these predictions.

The final three predictions were all concerned with the behaviour of the passivity-marked correlation functions. It is difficult to define a suitable mark for a galaxy's passivity, and this preliminary investigation has shown that there is room for substantial improvement in this area. The third and final predictions required the computation of the passivity MCF from the full red and blue combined sample. Similar to the first two, neither of these predictions can be reliably confirmed or ruled out. The large differences between the mark distributions of colour and passivity make direct comparison extremely difficult. However, in the higher redshift sub-samples ($1.6 < z < 2.8$), the passivity mark does seem to have a slightly greater large-scale environmental correlation. It is at these redshifts that we might expect there to be more contamination of the red sample by dusty objects.

Unfortunately, the remaining prediction that was to be tested, the fourth, is also inconclusive. Simply restricting the galaxy sample to those with red $U - B$ colours and computing the passivity MCF produces results which suggest there is no correlation between passivity and environment, at any scale or redshift. In other words, the measurements indicate that all of the environmental dependence on the passivity MCF can be explained by correlations between environment and colour. After rescaling the passivity marks, however, there does appear to be a residual correlation with environment, but it is marginal at most.

There is clearly a lot more work required in quantifying the biases introduced by the shape of the mark distribution. There is also much to be done in discerning the optimum mark to use for the 'passivity' of a galaxy. This latter task will undoubtedly be assisted by the ~ 4000 spectra of UDS galaxies which are currently being processed.

Chapter 5

Modelling

In this chapter I present some of the background theoretical framework upon which our understanding of galaxy evolution is built and then use it to interpret the results of a comparison between a mock observation from the literature and the observations of earlier chapters. Much of this framework constitutes what is usually referred to as the 'halo model' (Press & Schechter 1974; Cooray & Sheth 2002). This framework includes the formalism presented in Mo & White (2002), which was used in this thesis to compute the halo masses in chapters 2 and 3, and is also used to compute the clustering of halos later in this chapter. This section concludes with the formulation of the halo occupation distribution. This formalism is a new direction in which I am taking my work and is used to interpret the results in the latter half of the chapter.

Following the halo model I introduce a branch of theoretical techniques, semi-analytic models of galaxy formation (Cole et al. 1994), and detail one particular implementation (De Lucia & Blaizot 2007; Kitzbichler & White 2007). The publicly available mock observations from this model are then used in a comparison with the clustering behaviour of the passive and star-forming galaxies presented in chapters 2 and 3. This comparison highlights two issues with the model which are known, and further possible discrepancies regarding the evolution of clustering.

5.1 The Halo Model

In this section I describe the 'halo' model, including the formalism which was used to produce the dark matter masses which are used throughout this thesis. Press & Schechter (1974) suggested that we could treat the dark matter distribution of the Universe as a collection of discrete, gravitationally bound 'halos'. These halos are often further simplified by assumptions such as spherical symmetry, but regardless of the details, halo models underly much of the theoretical framework of galaxy formation. The original formalism of Press & Schechter (1974) was expanded upon in Bond et al. (1991); Mo & White (1996); Jing (1998); Sheth & Tormen (1999) and Sheth et al. (2001), and following confirmation by numerical techniques (e.g. Jenkins et al. 2001), the halo paradigm became firmly established.

The theory of the hot big bang and inflationary expansion (e.g. Peebles 1980; Guth & Jagannathan 1998) had also become widely accepted following observations of the cosmic microwave background (CMB, Penzias & Wilson 1965; Bennett et al. 1996). In particular, the temperature fluctuations that were observed by early CMB experiments were identified as regions of over or under-density and hence were responsible for the large-sale structure that we observe in the present-day Universe

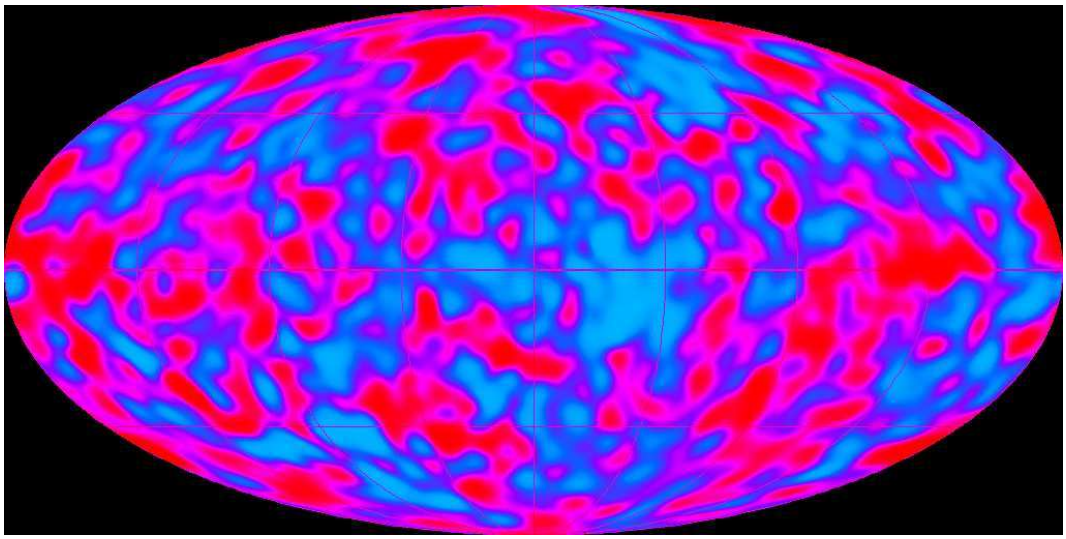

Figure 5.1. The cosmic microwave background, as seen by the COBE satellite. The colours correspond to variations in the temperature of the black body spectrum observed, and are a result of very small density fluctuations at the time of recombination. These fluctuations have amplitudes of only $\sim 10^{-5}$. Credit: NASA.

(Peebles 1982). The Cosmic Microwave Background Explorer (COBE Bennett et al. 1996) was a space-based mission to observe the whole sky, the results of which are shown in Figure 5.1. These measurements became a benchmark in the study of the CMB and facilitated new studies into the development of cosmic structure.

The effects due to gravity on a perturbed matter distribution are reasonably straightforward to compute, at least in the linear regime. The growth of the fluctuations observed by COBE could therefore be treated analytically in the linear regime (Peebles 1980), while numerical methods are employed when non-linear effects become important. This is the approach taken by Peacock & Dodds (1996) and though they did not express their work in the language of the halo model (something that was performed later by Smith et al. 2003), their work forms an important background for what will be discussed. Their argument and fit to the non-linear dark matter is outlined in the following subsection.

5.1.1 Linear and non-linear evolution of the dark matter distribution

The distribution of temperature density fluctuations observed in the CMB is approximately scale invariant and is well described by a simple power-law power spectrum:

$$P(k) = Ak^n \tag{5.1}$$

with n close to unity and A determined by the rms density fluctuation within a sphere of radius 8 h^{-1}Mpc at $z = 0$ (termed σ_8). The power spectrum can be expressed as a dimensionless function, $\Delta^2(k)$, following Peebles (1980):

$$\Delta^2(k) = \frac{d\sigma^2}{d\ ln\ k} = \frac{V}{(2\pi)^3} 4\pi k^3 P(k). \tag{5.2}$$

Here, the normalisation of the power spectrum is represented by the quantity V, a

normalisation volume. The primordial power spectrum, measured from the CMB, and the present-day linear power spectrum are linked by a transfer equation:

$$P_z(k) = Ak^n.T^2(k) \tag{5.3}$$

with

$$T(k) = \frac{\ln(1+2.34q)}{2.34q}\left[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4\right]^{-1/4} \tag{5.4}$$

and $q = k/[(\Omega_{CDM}+\Omega_B)\ h^{-2}\mathrm{Mpc}]$ (Bardeen et al. 1986). The redshift-dependent quantities Ω_{CDM} and Ω_B together make up the matter component of the universe which evolves as

$$\Omega_m(z) = \frac{\Omega_{m,0}(1+z)^3}{E^2(z)} \tag{5.5}$$

where $E(z) = \left[\Omega_{\Lambda,0}+(1-\Omega_0)(1+z)^2+\Omega_{m,0}(1+z)^3\right]^{1/2}$. The power spectrum is related to a measurable quantity, the 2-point correlation function, by the Fourier transform:

$$\xi(r) = \int \Delta^2 \frac{dk}{k}\frac{sin(kr)}{kr}. \tag{5.6}$$

In the linear regime, the 2-point correlations depend on the scale factor as $\bar{\xi}_L \propto [a(t)g(\Omega)]^2$, while on the smallest scales (largest wavenumbers), where the dark matter has decoupled from the Hubble flow, they depend on the background density, $\bar{\xi}_{NL} \propto a(t)^3$ (the symbol $\bar{\xi}$ represents the integrated correlation function). The factor $g(\Omega)$ is the generalisation of the growth factor for cosmologies with non-zero vacuum energy, given by Carroll et al. (1992) as

$$g(\Omega) = \frac{5}{2}\Omega_m\left[\Omega_m^{4/7}-\Omega_v+(1+\Omega_m/2)(1+\Omega_v/70)\right]^{-1}. \tag{5.7}$$

The aim of Peacock & Dodds (1996) (and of Hamilton et al. 1991 before them) was to characterise the dependence on scales between these two extremes.

Hamilton et al. (1991) suggested an empirical scaling between linear scales and transitional scales of the integrated correlation function. They argued that a region collapsing under gravity could be considered as being made up by infinitesimal shells. If these shells are not allowed to cross, then it could be possible to formulate a mapping (fit by numerical models) between shells on non-linear scales and those from linear predictions. This idea was taken up by Peacock & Dodds (1994) and modified to use the power spectrum in place of the integrated correlation function used by Hamilton et al. (1991). Subsequently, in Peacock & Dodds (1996), they improved upon their method, increasing the accuracy with which it fit the results of numerical models.

In the method of Peacock & Dodds (1996), the dimensionless power spectrum in the non-linear regime is written (following Peebles 1980):

$$\Delta^2_{NL}(k_{NL}) = f_{NL}\left[\Delta^2_L(k_L)\right], \tag{5.8}$$

with $k_L = \left[1+\Delta^2_{NL}(k_{NL})\right]^{-1/3} k_{NL}$. In the virialised regime this equation must

reduce to

$$f_{NL}(x) \propto x^{3/2}[g(\Omega)]^{-3}. \tag{5.9}$$

The fitting function that they chose has the form

$$f_{NL}(x) = x\left[\frac{1 + B\beta x + [Ax]^{\alpha\beta}}{1 + ([Ax]^{\alpha} g^3(\Omega)/[Vx^{1/2}])^{\beta}}\right]^{1/\beta}, \tag{5.10}$$

with best-fit parameters: $A = 0.482(1 + n/3)^{-0.947}$, $B = 0.226(1 + n/3)^{-01.778}$, $\alpha = 3.310(1 + n/3)^{-0.244}$, $\beta = 0.862(1 + n/3)^{-0.287}$ and $V = 11.55(1 + n/3)^{-0.423}$. In this case n is the effective power-law slope: $d\ ln\ P/d\ ln\ k$ where $k = k_L/2$. This solution for the non-linear dark matter clustering is an important prerequisite for the halo occupation distribution presented later.

5.1.2 Halo abundances and distributions

In the years that followed, a large body of work concerning the dark matter distribution and the halo model built up, and in Mo & White (2002) many of these results were gathered into a single framework. Given an initial power spectrum and cosmology, this framework predicts the abundances and distributions (the biases with respect to the dark matter distribution) of dark matter halos of any given mass at any given redshift. It remains in frequent use and is the origin of the halo masses quoted in previous chapters of this thesis. I shall outline this model here.

The foundation is that of Press & Schechter (1974): the abundance of halos as a function of mass and redshift per comoving volume is given by

$$n(M, z)dM = \left(\frac{2}{\pi}\right)^{1/2} \frac{\bar{\rho}_0}{M}\frac{d\nu}{dM}\exp\left(-\frac{\nu^2}{2}\right) dM. \tag{5.11}$$

This halo mass function was updated by Sheth & Tormen (1999) to better fit the results of numerical simulations:

$$n(M, z)dM = A\left(1 + \frac{1}{\nu'^{2q}}\right)\left(\frac{2}{\pi}\right)^{1/2} \frac{\bar{\rho}_0}{M}\frac{d\nu'}{dM}\exp\left(-\frac{\nu'^2}{2}\right) dM. \tag{5.12}$$

This modification to Press & Schechter (1974) was justified by the possibility of halos collapsing ellipsoidally rather than spherically (Sheth et al. 2001; Sheth & Tormen 2002). The quantities in the equation for the mass function are detailed in the following. The fitted quantities A and q are $\simeq 0.322$ and 0.3 respectively, while the replacement of ν with ν' is simply $\nu' = \sqrt{a}\nu$ where $a = 0.707$.

It is the quantity ν, then, which determines the dependences on both mass and time (including cosmology).

$$\nu = \delta_c/(D(z)\sigma(M)) \tag{5.13}$$

where $\delta_c = 1.69$ is a constant, $D(z)$ is the growth factor from Carroll, Press & Turner (1992) ($D(z) = g(z)/[g(0)(1 + z)]$) with Ω_m given by equation 5.5 and Ω_Λ given by

$$\Omega_\Lambda(z) = \frac{\Omega_{\Lambda,0}}{E^2(z)}. \tag{5.14}$$

Finally, $\sigma(M)$ is the rms density fluctuation within the Lagrangian radius of a halo, mass M, at $z = 0$:

$$\sigma^2(R) = \frac{1}{2\pi^2} \int_0^\infty k^3 P(k) \bar{W}^2(kR) \frac{dk}{k}, \tag{5.15}$$

with $R(M) = \left(\frac{3M}{4\pi\bar{\rho}_0}\right)^{1/3}$. This argument provides the mass function of dark matter halos at any given epoch, but not their distributions. Mo & White (1996) used the extension to Press & Schechter (1974) of Bond et al. (1991) and developed an analytic model for halo clustering. Using N-body models they showed that, at the scales for which linear clustering is appropriate, a simple multiplicative factor is all that is needed to recover the cross-correlation of halos of a given mass with dark matter from the dark matter auto-correlation. This bias factor is

$$b(M, z) = 1 + \frac{\nu^2(M, z) - 1}{\delta_c}. \tag{5.16}$$

The auto-correlation is the quantity which is typically measured,

$$\xi_{hh}(r, z) = b^2(M, z) \xi_{mm}(r, z). \tag{5.17}$$

So, under the assumption that each halo hosts only one galaxy of the population of interest, the large-scale galaxy auto-correlation function can be related directly to the mass of the halos which host them. This assumption is typically not valid, however, and will introduce a bias in the recovered mass. The magnitude of this bias was studied in Zheng et al. (2007), and is an important consideration in section 5.3.

Jing (1998) showed that the bias formula of Mo & White (1996) was inaccurate, and so Sheth et al. (2001) found a new formula, compatible with the ellipsoidal collapse model:

$$b = 1 + \frac{1}{\delta_c} \left[\nu'^2 + s\nu'^{2(1-c)} - \frac{\nu'^{2c}/a^{1/2}}{\nu'^{2c} + s(1-c)(1-c/2)} \right] \tag{5.18}$$

with a=0.707, s=0.5 and c=0.6. This bias can be related to the bias of the typical $z = 0$ descendent halos by

$$b = 1 + \frac{D(z)}{\delta_c} \left[\nu'^2 + s\nu'^{2(1-c)} - \frac{\nu'^{2c}/a^{1/2}}{\nu'^{2c} + s(1-c)(1-c/2)} \right]. \tag{5.19}$$

We now have the bias of dark matter halos of given mass at a given redshift with respect to the matter distribution. From the Peacock & Dodds (1996) argument above, we also have the clustering of the matter distribution across both linear and non-linear scales for any given redshift. In principle, then, we can associate a galaxy population, for which we compute $\xi(r)$, with the mass of the dark matter halo which hosts them. In reality it is not quite so simple. A galaxy population is not found exclusively in halos of a narrow mass range. In particular, galaxies found within halos of a given mass will also be found in more massive halos, often with there being more than one galaxy per halo. The first aspect is partly solved by using an abundance averaged bias of all halos more massive than the halo mass of interest. More generally however, this issue has in part motivated the development of halo occupation models (see the following subsection).

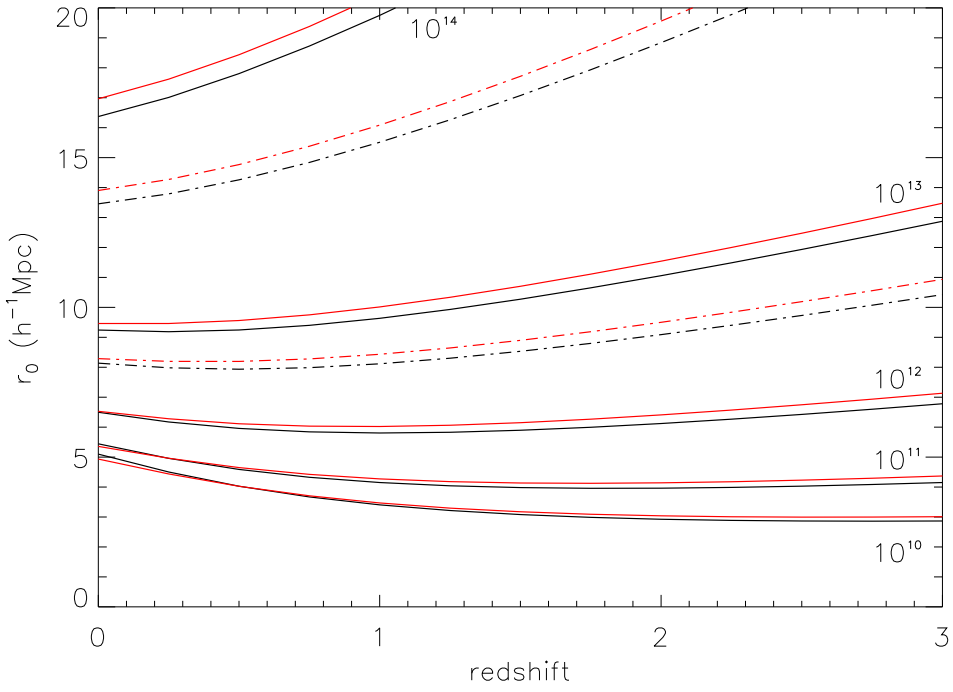

Figure 5.2. Clustering strengths of dark matter halos as a function of redshift computed from the formalism of Mo & White (2002) for the two cosmologies which are used in this thesis. In black are the clustering of halos in the cosmology used throughout most of this thesis; in red are the halo clustering strengths in the cosmology used for the Millennium simulation. The lines are of constant mass (M_{200}) as indicated (in solar masses). The dot-dashed lines are for masses of 5×10^{13} (upper) and 5×10^{13}M$_\odot$(lower).

In the previous chapters of this thesis, the redshifts of the target galaxies were insufficiently well known to compute $\xi(r)$ directly (this is true even in most redshift surveys, where the projected correlation function is often preferred). Instead, the angular correlation function was computed and then deprojected to find the clustering scale length, r_0. Finding this scale length from the model of Mo & White (2002) is relatively straightforward. Using the definition for Δ_8 from Peebles (1980),

$$\Delta_8 = (C_\gamma (r_0/8\ h^{-1})^\gamma)^{1/2}, \tag{5.20}$$

where $\Delta_8(M,z) = \bar{b}(M,z)\sigma_8 D(z)$, $C_\gamma = 72/((3-\gamma)(4-\gamma)(6-\gamma)2^\gamma)$ and $\gamma = 1.8$, the canonical slope of the correlation function. The bar above b indicates that it is the abundance averaged bias that is to be used here. Figure 5.2 shows r_0 values as a function of redshift for different halo masses, which have been computed here in two different cosmologies. The black lines are for the cosmology used throughout this thesis, as plotted in Figure 3.12, and from which the masses in chapter 2 are determined. In red those for the cosmology used in the Millennium simulation are shown, and are used later in this chapter. The two sets of data are very similar, mainly due to the fact that the value for σ_8 is the same in each case.

5.1.3 Halo occupation distribution

The most significant development in analytic halo model theory since Mo & White (2002) has been halo occupation distribution (HOD) modelling (Peacock & Smith 2000). As stated in the previous subsection, relating galaxies to halo masses through the linear bias requires the implicit assumption that there is only one galaxy of the sample under examination within each halo. This assumption is generally not true. HOD models seek to make predictions of observable quantities, such as galaxy clustering, by populating dark matter halos with galaxies in a plausible manner. The use of HOD models in the study of galaxy clustering is particularly fruitful as the contribution to the two-point correlation function by galaxy pairs within a single halo can be separated from pairs of galaxies in separate halos. Of the current HOD formalisms that of Magliocchetti & Porciani (2003); Porciani et al. (2004) and Magliocchetti et al. (2007) is favoured here for its simplicity, though it should be noted that a distinction between central and satellite galaxies of a halo (as used in Zheng et al. 2007) may be important (Berlind et al. 2003; Zheng et al. 2005).

Provided the halo mass function and galaxy occupation function do not evolve significantly over the redshift range of interest the mean co-moving number density of galaxies in a sample can be determined as

$$\bar{n} = \int n(M)N(M)dM, \tag{5.21}$$

where $n(M)$ is the halo mass function and $N(M)$ the occupation number as a function of halo mass. We are free to parametrise occupation function in any way we please, however a simple and sensible choice is that of Porciani et al. (2004):

$$\begin{aligned} N(M) = &\quad N_0(M/M_{min})^{\alpha} \quad \text{if} \quad M \geq M_{min} \\ = &\quad 0 \qquad\qquad\quad \text{if} \quad M < M_{min}. \end{aligned} \tag{5.22}$$

The quantities α, N_0 and M_{min} are determined by minimising the χ^2 between the theoretical correlation function resulting from the following formalism and the measured correlation function. Before beginning the computation of the theoretical correlation function there is one more quantity, related to the occupation function, that is required. The second factoral moment of the occupation number, $\sigma(M) =< N(N-1) >$, is

$$\sigma(M) = \beta(M)^2 N(M) \tag{5.23}$$

with

$$\begin{aligned} \beta(M) = 0 &\qquad \text{when } N(M) = 0, \\ \beta(M) = \frac{\log_{10}(M/M_{min})}{\log_{10}(M_0/M_{min})} &\qquad \text{when } N(M) < 1, \\ \beta(M) = 1 &\qquad \text{when } N(M) =\geq 1. \end{aligned} \tag{5.24}$$

With this functional form for the occupation distribution and the halo and dark matter models introduced earlier it is possible to compute a theoretical correlation function. Such a computation is usually done in two separate regimes, the 1-halo and 2-halo regimes. Observationally, the correlation function is determined by counting

pairs. If we happen to know which dark matter halos hosted which galaxies we could separate the counts of galaxy pairs within the same halo (the 1-halo regime) from those of pairs which are hosted by different halos (the 2-halo regime). The combined correlation function is the sum of the correlation functions in these two regimes:

$$\xi(r) = \xi_{1h}(r) + \xi_{2h}(r). \tag{5.25}$$

These two terms are computed as follows:

$$\xi_{1h}(r) = \int \frac{n(M)\sigma(M)}{n_{gal}^2} dM \int \rho(x|M)\rho(x+r|M)d^3x, \tag{5.26}$$

$$\xi_{2h}(r) = \int \frac{n(M_1)N(M_1)}{n_{gal}} dM_1 \int \frac{n(M_2)N(M_2)}{n_{gal}} dM_2$$
$$\times \int \rho(x_1|M_1)\rho(x_2|M_2)\xi(r_{12}|M_1, M_2)d^3r_1 d^3r_2. \tag{5.27}$$

Each of these terms requires a convolution of the halo density profile, $\rho(r)$, and are more easily computed in k-space. The two terms then become:

$$n_{gal}^2 \Delta_{1h}(k) = \int n(M)\sigma(M)|u_M(k)|^2 dM, \tag{5.28}$$

$$n_{gal}^2 \Delta_{2h}(k) = \int n(M_1)N(M_1)u_{M_1}(k)dM_1$$
$$\times \int n(M_2)N(M_2)u_{M_2}(k)\Delta(k, M_1, M_2)dM_2. \tag{5.29}$$

When expressed in this way, the halo density profile is replaced by its Fourier transform:

$$u_m(k) = \frac{k^3}{2\pi^2} \int_0^{r_{cut}} \rho(r) \frac{\sin(kr)}{kr} 4\pi r^2 dr \tag{5.30}$$

where the density profile for a halo is normalised by the mass within its virial radius and $\Delta(k, M_1, M_2)$ is given by

$$\Delta(k, M_1, M_2) = \Delta_{dm}(k)b(M_1)b(M_2) \tag{5.31}$$

if $r > R_{vir,1} + R_{vir,2}$ and zero otherwise. The bias as a function of halo mass, $b(M)$, is found from the model of Mo & White (2002), introduced earlier, and Δ_{dm} is the dimensionless power spectrum of Peacock & Dodds (1996). Finally, the 1-halo and 2-halo correlation functions are just the reverse Fourier transforms of Δ_{1h} and Δ_{2h}:

$$\xi(r) = \int \Delta(k) \frac{\sin(kr)}{kr} \frac{dk}{k}. \tag{5.32}$$

Unfortunately, observational data typically do not allow direct computation of the 3-d correlation function and so there is one further step that must be made: projection to the angular correlation function. Projection of the spatial correlation function to the angular correlation function was formulated many years ago (Limber 1954; Peebles 1980) and the inversion of this technique was used in earlier chapters.

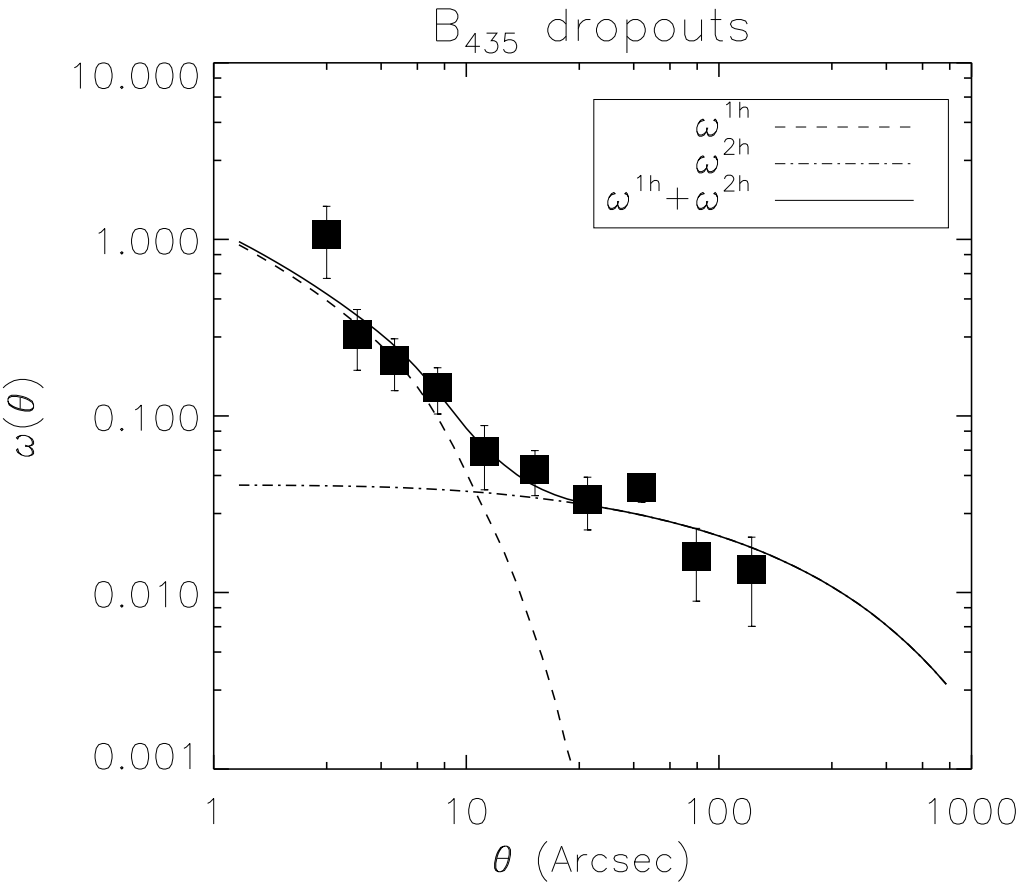

Figure 5.3. An example of the halo occupation distribution technique for an angular correlation function from Lee et al. (2006). The 2-halo term dominates only at large scales, with a plateau on the smallest scales. The 1-halo term dominates on small scales but the region of cross-over between the two is very narrow. The data points are their angular correlation function computed for B-dropout galaxies.

$$w(\theta) = 2\frac{\int_0^\infty \int_0^\infty N(z)^2 \xi(r,z) dz\ du}{\left[\int_0^\infty N(z) dz\right]^2}. \tag{5.33}$$

Thus we have a synthetic $w(\theta)$ from the halo model which can be fit to observations. An example, taken from Lee et al. (2006), is shown in figure 5.3 with the 1-halo and 2-halo regimes labelled. Particular points to note are how narrow the angular range is over which the two contributions are comparable, and also the small scale plateau of the 2-halo term. Despite the power of the HOD technique, achieving a fit to an observational data set is highly time consuming and expensive and a full implementation with the earlier results of this thesis shall have to be left as future work. However, the formalism itself can be of use in interpreting results as shall be discussed later in this chapter. The approach taken in the remainder of this chapter is to use a publicly-available mock galaxy catalogue (Lemson & Virgo Consortium 2006) built on the merger trees extracted from a high-resolution but large-volume dark matter N-body simulation. By comparing predictions from

such a model with the observational results and using the halo model formalism, it is possible to determine reasons for discrepancies between the two, and perhaps suggest areas where the model can be improved.

5.2 Semi-analytic models of galaxy formation

Semi-analytic (S-A) models are a branch of theoretical galaxy formation models. Their name represents the approach taken, to use partly numerical methods and partly analytic solutions. Though in fact this description is often a gross over-simplification. In the hierarchical model of galaxy formation, galaxies form in the potential wells of halos of dark matter. The first ingredient in semi-analytic models is therefore the dark matter structure as a function of time.

The dark matter structure and merger histories can be taken from a Press & Schechter approach (Press & Schechter 1974), similar to the methodology outlined above, or from the results of a dark matter N-body simulation. In the case of a simulation, the halo mass function is distorted slightly at the high mass end due to the limited volume of the simulation. However, there are advantages too, namely the ability to include effects due to the larger-scale environment, should they be required. In recent years the development of semi-analytic models has been led by groups that also have interests in numerical models, with the result that most current S-A models are built on the skeleton of an N-body simulation.

Many of the quantities that we measure by observations of galaxies: colour, brightness, metalicity content etc., are the result of either the existence of stars or their evolutionary byproducts. The most fundamental product of a semi-analytic model must therefore be the rate of formation of stars as a function of epoch. This in turn requires knowledge of the amount of gas in a galaxy and the gas cooling rate. Processes which do not affect these quantities are 'bells and whistles' and are of secondary importance. Several empirical and theoretical models for the cooling of gas exist (Raymond & Smith 1977; Sutherland & Dopita 1993; Mewe et al. 1995). One of these codes is used together with an efficiency of converting cold gas to stars to describe the instantaneous SFR. The SFR as a function of time and the initial mass function (IMF) of stars formed from a given gas mass are the principle ingredients required to determine the observable properties of the modelled galaxies (though the effects of dust are often important too).

Other important ingredients such as feedback from supernovae and active galactic nuclei (AGN) can influence the cooling rate or cold gas fraction. In cases such as these the details are unknown and so they are simplified to energy or momentum injections with free parameters which can be tuned to fit observations. An example of a procedure used in building a semi-analytic model is shown in Figure 5.4, taken from Baugh (2006).

5.2.1 S-A models using the Millennium simulation

The models used in this chapter are those of De Lucia & Blaizot (2007) and the light cones constructed by Kitzbichler & White (2007). These models are based on the Millennium simulation: a dark matter only N-body simulation of 10^{10} particles in a co-moving cubic volume with side $500\mathrm{h}^{-1}\mathrm{Mpc}$ (Springel et al. 2005). In the cosmology assumed in the simulation, this number implies a particle mass of $8.6 \times 10^{8}\mathrm{h}^{-1}\mathrm{M}_{\odot}$. The Plummer equivalent softening length (Plummer 1911) is

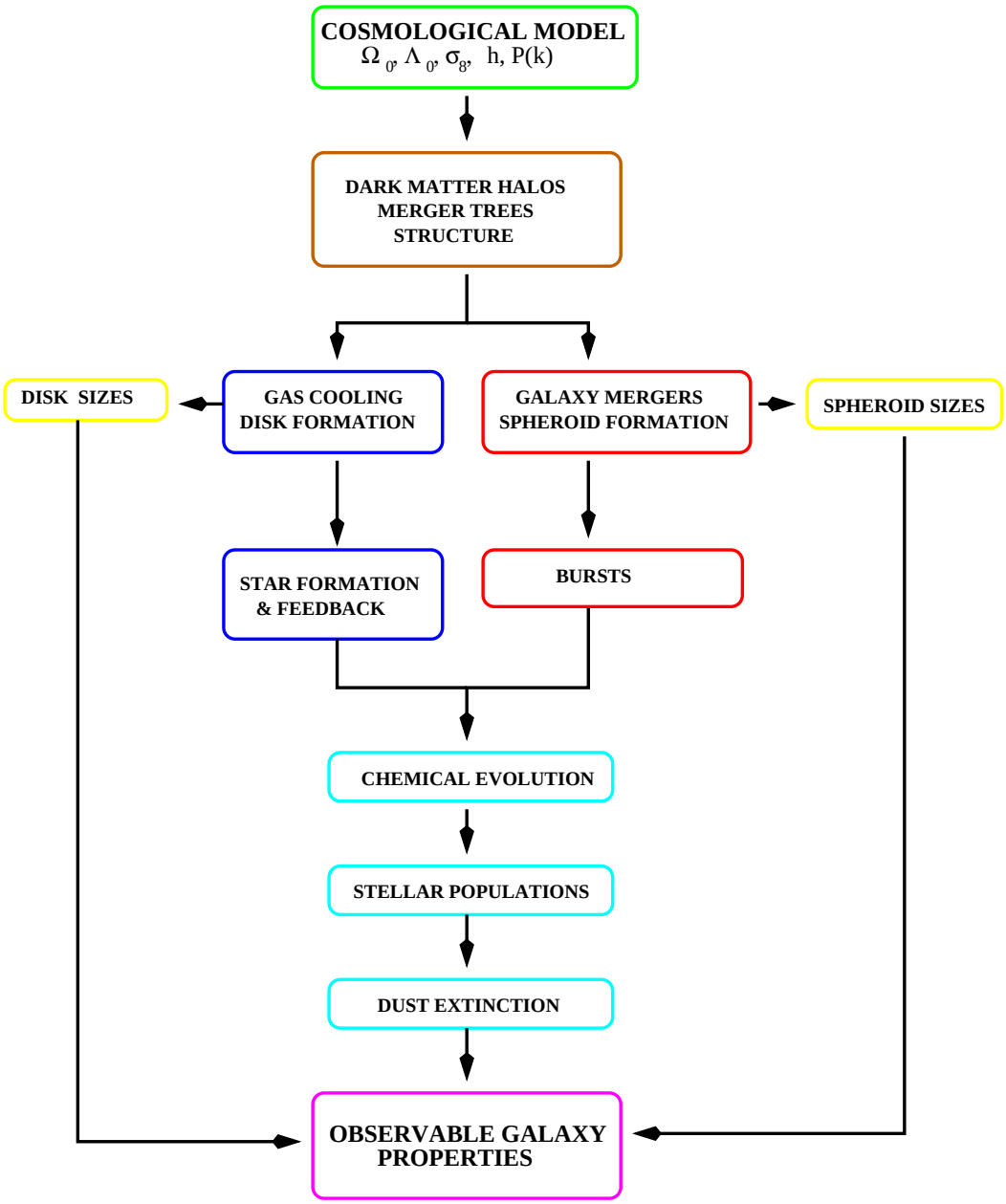

Figure 5.4. A simplified flow chart of the steps taken to build a semi-analytic model of galaxy formation, taken from Baugh (2006). The dark matter merger trees are often taken from N-body simulations and match the structure of the observed Universe extremely well. It is in the gas cooling, merging and feedback processes that the galaxy observables are principally determined.

$5h^{-1}$kpc. The cosmology is a flat ΛCDM with $\Omega_m = 0.25$ and $\sigma_8 = 0.9$. Halos are extracted by a friends-of-friends (fof) algorithm with linking length $b = 0.2$ and then gravitationally bound sub-halos are identified from each fof group. Any sub-halo of 20 particles or more is saved and used as a site for a galaxy in the semi-analytic model. Sub-halos are tracked so that satellite galaxies and their merger histories can be studied from the redshift they were first identified to $z = 0$.

The synthetic galaxies of De Lucia & Blaizot (2007) are an incremental improvement upon those of Croton et al. (2006). The two models therefore share many of the same implementations for physical processes. The most significant of these are briefly described here for completeness. The 'ingredients' of a model can be broadly separated into two sets: those that are required to grow galaxies and those that are required to produce observables. The first of these sets includes parameters for the handling of mergers, feedback processes and star-formation efficiency. The principle requirement of these processes is to produce the cold gas mass required for the galaxies star-formation histories to match observation (Madau 1995). The prescription follows from that of White & Frenk (1991), as each halo collapses it is assumed that the gas fraction of this halo is that of the universe (~ 0.17 for the Millennium simulation cosmology). This gas is then instantaneously shock heated to the virial temperature of the halo. The cooling rate of the gas is then computed. At early times and in low-mass halos the gas cools very rapidly (the rapid cooling regime), while in high-mass halos at late times cooling is very slow (the static hot halo regime). In either case, gas only cools onto the central galaxy of a halo or sub-halo and not onto satellite galaxies. Star formation is triggered by a threshold in surface density and the star-formation efficiency and recycled fraction are just two of the many parameters which can be adjusted to match observations.

However, in a simple model such as this too many stellar-massive galaxies are produced and too many low-mass halos are bright. Both ends of the luminosity function must be suppressed in order to match observations. There are three main processes at work: re-ionisation, star-formation feedback and feedback from AGN. By heating the gas during re-ionisation, low mass halos are prevented from forming stars at very early times. The cooling rate is proportional to the square of the gas density (White & Rees 1978) and so a greater proportion of gas can escape the gravitational potential of low-mass halos.

Feedback from star-formation and AGN are a little more complex and are also influenced by mergers. Feedback from star-formation is in the form of supernovae, implemented by an instantaneous transfer of energy into the gas. The amount of energy injected is another of the tunable parameters, but is kept within plausible limits. An alternative to changing the amount of energy per supernova is by changing the star-formation initial mass function (IMF). A standard Chabrier (2003) IMF is used in De Lucia & Blaizot (2007). This feedback and the surface-density threshold produce episodic star-formation.

During mergers, however, up to 40% of the available gas can be turned into stars at once. When a galaxy enters the halo (including sub-halos) of another galaxy a timer begins, and after twice the dynamical time (Binney & Tremaine 1987) it is merged with the central galaxy of that halo. When a merger occurs a collisional star-burst follows (Somerville et al. 2001) and the central black holes also merge. In addition, the central black hole of the new descendant galaxy is fed by gas and shines in a quasar phase (Kauffmann & Haehnelt 2000). Unlike Kauffmann & Haehnelt (2000) however, a quasar phase can be triggered by either a major or minor merger

in the model of De Lucia & Blaizot (2007). However, during this phase of an AGN there is no feedback to the gas component. Jets from quasars are highly collimated and it is not at all clear how it is possible to transfer energy from them to the gas within a galaxy or beyond. Furthermore, the models do not need quasar feedback to reproduce low-redshift galaxy properties as star-formation feedback is sufficient for all but the most massive galaxies. However, it could equally be considered that quasar feedback is included as part of the star-formation feedback during a merger.

Feedback from AGN is instead limited to the 'radio mode' by De Lucia & Blaizot (2007). At low redshift, powerful radio galaxies are almost exclusively massive elliptical galaxies in or near the centres of massive halos (Dunlop et al. 2003). In models which do not use radio AGN feedback such galaxies are frequently too stellar massive and too blue (Benson et al. 2003), indicating that too much gas is cooling onto them at late times. In the radio mode scenario the radio lobes create bubbles in the hot static halo which further heat the group or cluster gas, preventing further cooling. Motivated by the desire to produce brightest cluster galaxies which exhibit downsizing behaviour and the low-redshift luminosity function, radio mode feedback is a plausible answer to a previously long-standing problem (Bower et al. 2006).

The second task, to produce observables, requires stellar evolution models and a treatment of the dust content of galaxies. Stellar evolution in the De Lucia & Blaizot (2007) and Kitzbichler & White (2007) models uses the Bruzual & Charlot (2003) stellar population code with the Padova 1994 stellar evolution tracks (see references within Bruzual & Charlot 2003) and Chabrier IMF (Chabrier 2003). The dust is treated slightly differently between the two models. De Lucia & Blaizot (2007) use Devriendt et al. (1999) for the diffuse inter-stellar medium and Charlot & Fall (2000) for the absorption of the molecular stellar birthing clouds. The metalicity dependence on the dust extinction follows Hatton et al. (2003) and each galaxy is given a random inclination. The aims of De Lucia & Blaizot (2007) were to match the low-redshift universe and reproduce BCGs in the context of downsizing. Kitzbichler & White (2007) wished to study galaxies at high redshift where the metalicities and dust content is much more variable. For this reason they constructed a much more detailed treatment of dust, incorporating a redshift scaling of the dust-to-gas ratio.

The mock observations of Kitzbichler & White (2007) were constructed from light cones. A light cone is created by positioning an observer at the centre of the simulation volume at redshift $z = 0$ and tracing all the light-like world lines which intersect this observer. A field of view is chosen, 1.4×1.4 deg. in the case of Kitzbichler & White (2007), and the remainder of the light cone is discarded. For the volume of the Millennium simulation, this mock observation extends to $z = 0.17$. In order to probe the high redshift universe the volume must therefore be repeated. This is done by firstly slanting the line of sight across the volume, enabling more of the volume to be used before repetition of structure is seen. Repetition of structure is, however, inevitable in very deep mock observations, but provided they are located at different angular positions the effects are minimal (Kitzbichler & White 2007). Equiped with these mock observations and the theoretical framework described earlier in this chapter, I re-select the samples used in chapters 2 and 3 and compare the model predictions to the data.

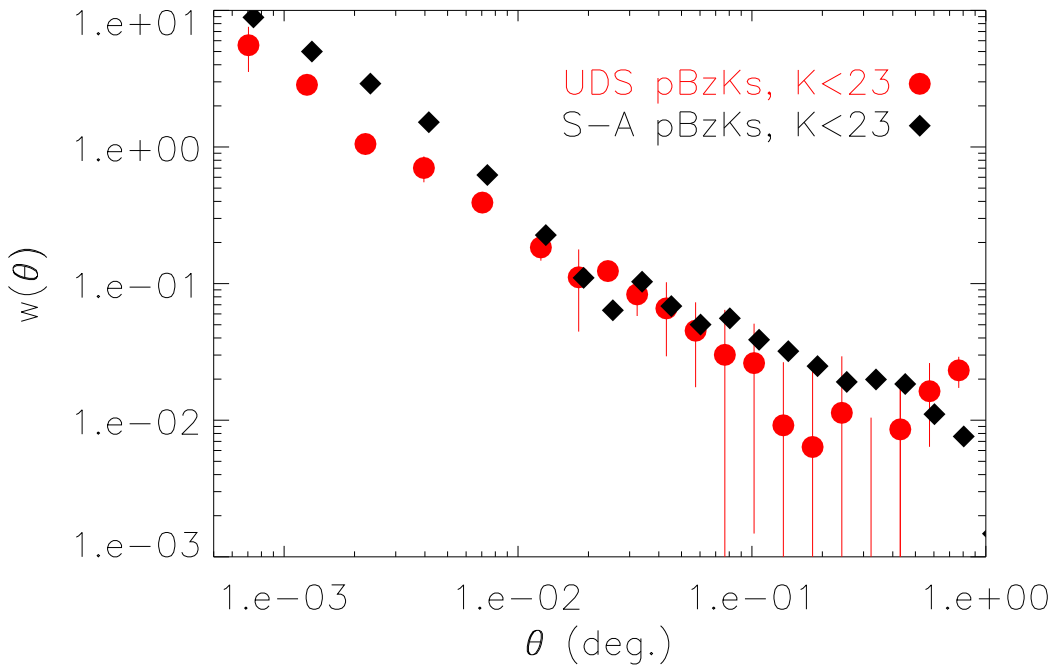

Figure 5.5. Correlation function of pBzKs taken from the semi-analytic model of Kitzbichler & White (2007) (black diamonds) compared with that from the UDS (red circles). On large scales the two match extremely well. However, on small scales, where the effects of halo occupation are most clearly seen, the model produces a higher correlation function.

5.3 Comparison of the results of BzK-selected galaxies

Attempting to directly reproduce observational results, such as the angular clustering of BzK-selected galaxies, from mock galaxies is difficult. For instance, $B-z$ and $z-K$ colours are defined within $2''$ apertures in observed data sets, while the equivalent colours from S-A models are perhaps more closely related to total magnitudes. Ideally, the mock galaxies would be used to produce a mock image with the appropriate level of noise added, rather than a catalogue. However, despite such quantities as the bulge to disk ratio being modelled, such a step would likely introduce further uncertainties and perhaps make interpretation more difficult rather than simpler. In either case, such mock images are not available for the models that are used here, and it is well beyond the scope of this work to produce them. Colours computed from the magnitudes provided in the public database (Lemson & Virgo Consortium 2006) are therefore assumed to be appropriate in this work. As most of the objects of interest are at high redshift, and therefore have small angular extent, this assumption should be reasonable.

The B and K filters used to produce magnitudes in the S-A model are both Johnson filters, while the z-band filter is that of the SDSS. Using the same method as described in chapter 2, the $B-z$ and $z-K$ colours were corrected to the system used by Daddi et al. (2004). In fact, there was relatively little difference between colours computed from the model filter sets and those of the UDS/SXDS data set.

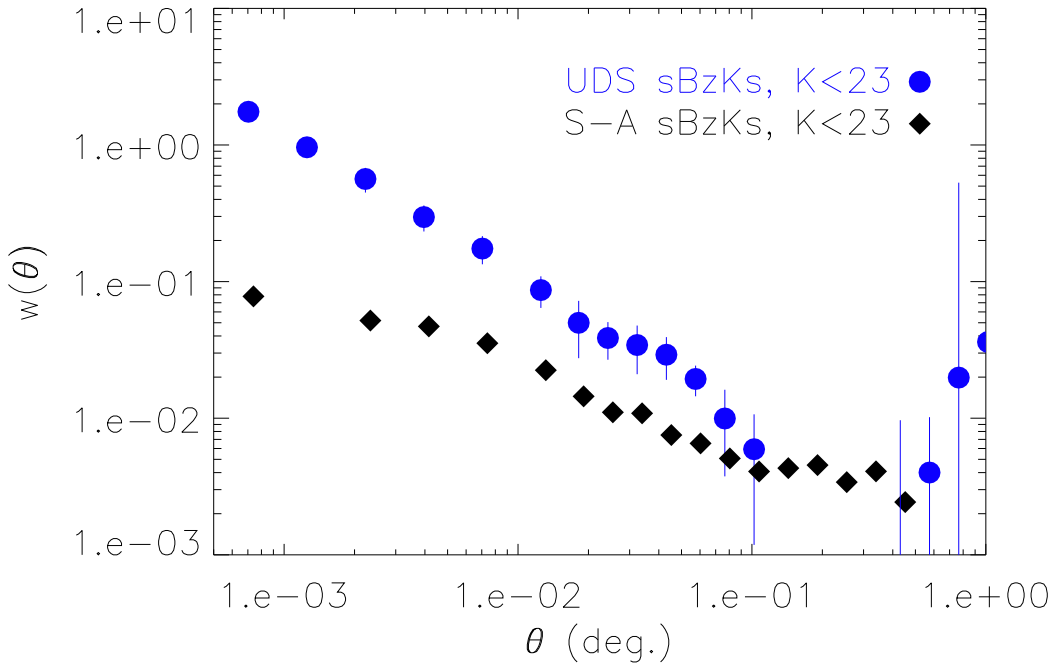

Figure 5.6. Correlation function of star-forming BzK-selected galaxies from Kitzbichler & White (2007) (black diamonds) and the UDS (blue circles). The model does not reproduce the sBzK clustering at any scale. This difficulty is most likely linked to the small-scale clustering over-prediction in the pBzK measurement (see text).

Using the standard selection definition of Daddi et al. (2004), pBzK and sBzK samples were extracted from one of the mock galaxy catalogues of Kitzbichler & White (2007), with a magnitude cut at $K_{AB} = 23$. Angular correlation functions were then computed as detailed in chapters 1 and 2 and these are plotted along with the observational results in Figures 5.5 and 5.6. The errorbars on the model measurements have been suppressed for clarity, but they are substantially smaller than those of the observations as the mock field size is almost four times that of the UDS.

The measurements of passive BzK-selected galaxies are similar. The model appears to have done a good job of reproducing the large-scale clustering strength. As the large-scale clustering is related to the host dark matter halo mass, this agreement is highly encouraging. However, on small scales the model points are substantially higher than those of the observed galaxies. As this discrepancy is found on small scales the cause must be related to the halo occupation distribution of pBzKs in the model.

There is a known problem with the model galaxies used in this work, but also across semi-analytic models more generally (Croton et al. 2006; Kitzbichler & White 2007; Font et al. 2008). S-A models tend to produce too many satellite galaxies, and those that are produced are frequently too red in colour. As detailed in the model description above, when a galaxy falls into a more massive halo and becomes a satellite, its gas supply is immediately removed, as gas is only allowed to cool onto

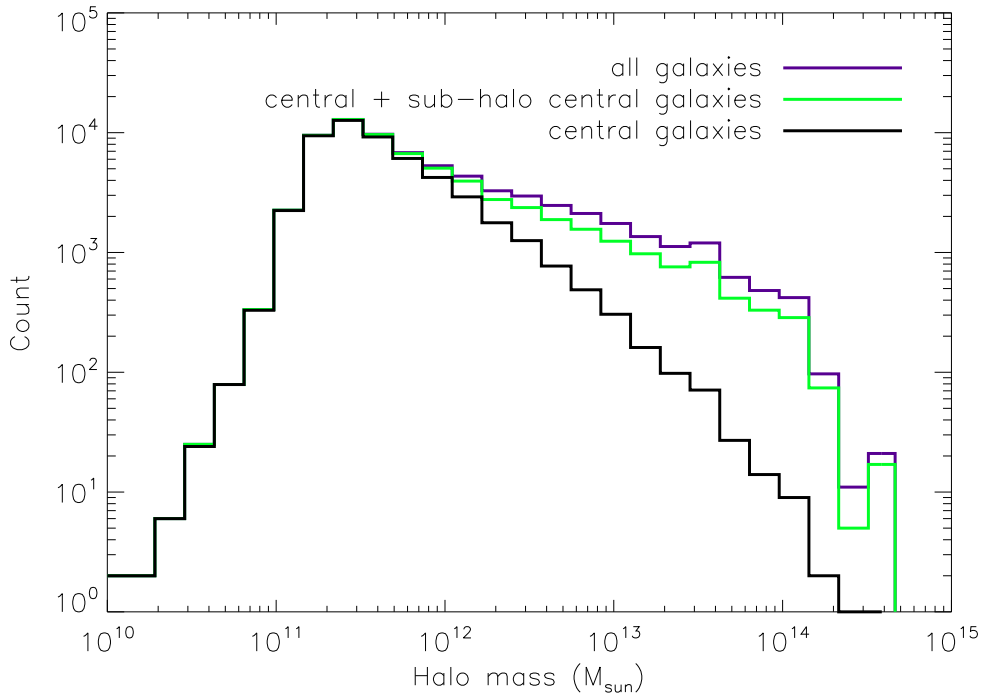

Figure 5.7. The total number of galaxies hosted by dark matter halos of different masses within $1 < z < 1.5$ and with $K < 23$ from the mock catalogues. The galaxies are separated by their type. i.e. whether they are centrals, the main galaxy of a sub-halo, or in a sub halo of a sub halo. The black line gives the counts for central galaxies only; green is for central galaxies plus the main sub-halo galaxies, while the purple line shows the counts for all galaxies in the sub-sample.

the central galaxy of a parent halo. Within a fairly short time ($t \leq$ 1Gyr) these galaxies develop a red SED. An attempt to correct this problem was made by Font et al. (2008), but with only limited success. Though they were successful in reducing the number of red satellites by extending the time that they retain a gas supply, many of these galaxies were instead located in the green valley, reducing the clear colour bimodality. The colour of satellite galaxies remains an outstanding problem in galaxy formation models. This known problem is the most likely explanation for the discrepancy between model and observations in this case.

The measurements of model and observed sBzK-selected galaxies, shown in Figure 5.6, appear at first glance to be greatly at odds with each other. Given the tendency for satellite galaxies to be red, and hence selected as pBzKs rather than sBzKs, a difference at small scales should be anticipated. However, not only are the small scales different, the large-scales also show a large discrepancy. Much of this difference in the two sets of measurements can be explained by the same issue that the pBzK sample highlighted. The removal of blue satellites modifies the halo occupation distribution of the model sBzK sample. The form of $w(\theta)$ is very similar to that of the example 2-halo term shown in Figure 5.3. The 1-halo term must therefore be negligible, and hence the conversion of satellites from star-forming to passive must be highly efficient. However, $N(M)$ also appears in equation 5.27, the 2-halo term. The 2-halo term is effectively the correlation function of the dark

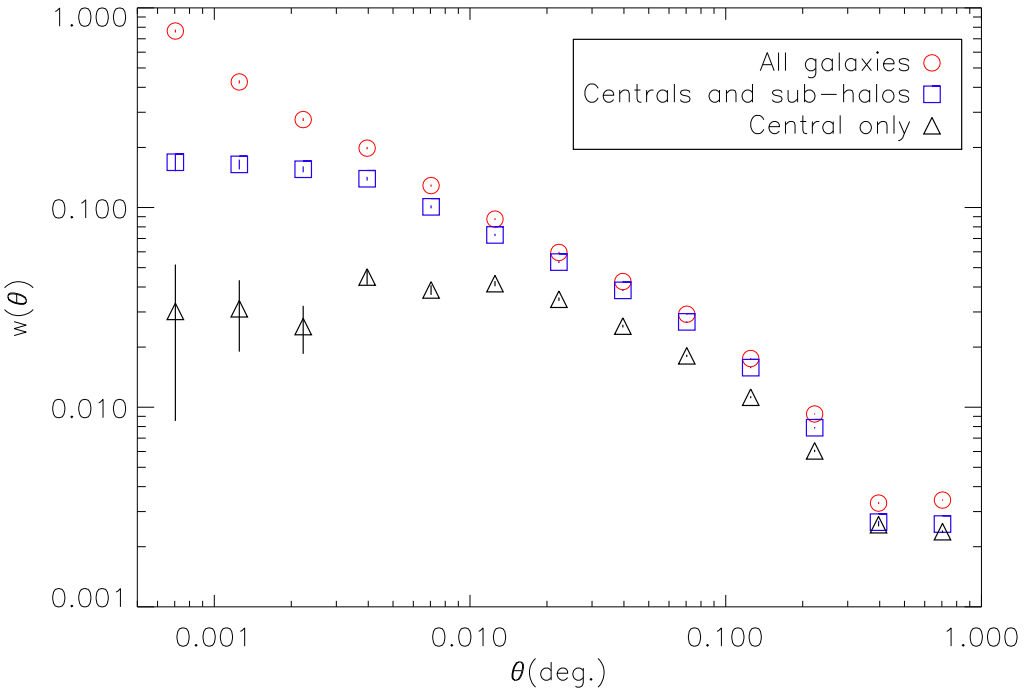

Figure 5.8. Correlation function measurements for a subset of galaxies from the mock observations of Kitzbichler & White (2007). The galaxies used here lie in the redshift range $1 < z < 1.5$ and have $K < 23$. They are further defined in the mock catalogue by whether they are the central galaxy of their parent halo, the main galaxy in a sub halo, or a satellite. Measurements shown are for centrals only (blue triangles), centrals plus main sub-halo galaxies (red squares) and all galaxies (black circles).

matter multiplied by the HOD-weighted bias. The HOD is typically a power-law above some minimum mass, and hence as the number of satellite galaxies increases the HOD-weighted bias also increases. The influence of the HOD is shown explicitly in the following example. A control sample is taken from the Kitzbichler & White (2007) model of all galaxies with $K < 23$ within the redshift range $1 < z < 1.5$. The model galaxies are defined as either parent-halo central galaxies, the central galaxy of a sub-structure or a satellite. Figure 5.7 shows how these galaxies are distributed amongst parent halos of different masses. The count is much higher for low-mass halos as there are many more of them. However, they are typically host to only a single galaxy, while more massive halos host many satellites. Auto-correlation functions were computed for each of these samples, which are presented in Figure 5.8. The absence of satellite galaxies (including those which are sub-halo central galaxies) has a clear impact at small scales, but as was suggested would be the case, there is also a depression of the large-scale amplitude. This amplitude depression is a factor ~ 1.5 and so is, however, unlikely to be the only cause of the difference in clustering between the model and observed sBzK samples.

Under the assumption of one galaxy per halo, the difference in amplitudes on large scales suggests that the bias of model sBzK galaxies is wrong. Such a situation may arise if the galaxies formed in lower mass halos in the model are brighter than in the real universe and hence are included in the $K_{AB} < 23$ magnitude cut. Indeed,

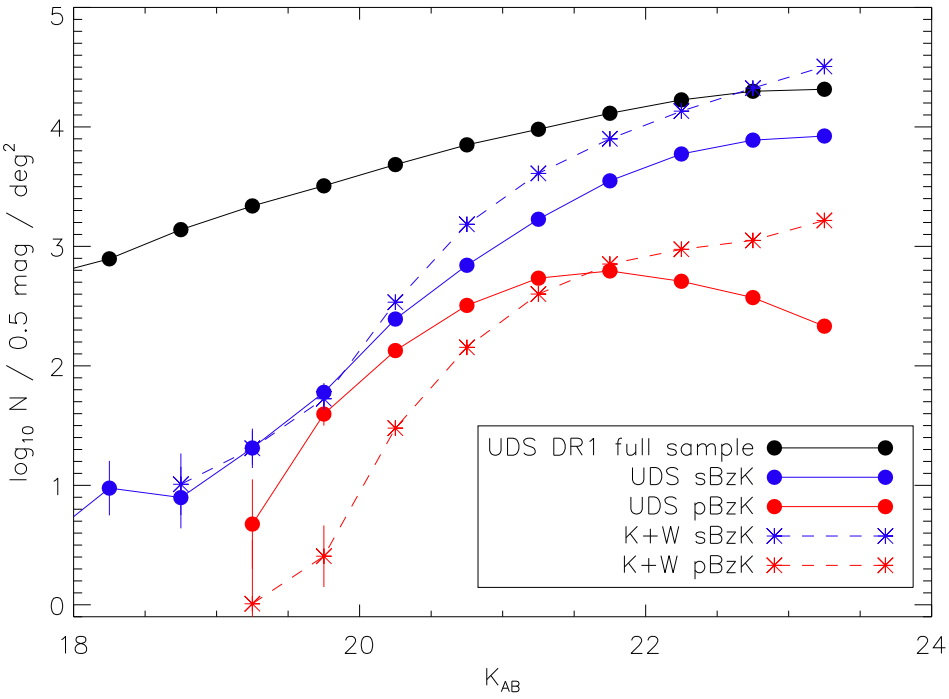

Figure 5.9. The number counts of the BzK-selected galaxies from the mock observations of Kitzbichler & White (2007) (labelled K+W) and the UDS samples. There is a clear over-abundance of star-forming objects (sBzK) and faint passive objects (pBzK). The brighter passive objects also appear to be ~ 0.5 magnitudes too dim.

the K-band number counts of the model galaxies, shown in Figure 2 of Kitzbichler & White (2007), are substantially over-predicted for $K > 20$. At the magnitude limit imposed in this work ($K_{AB} = 23$), their Figure 3 shows that these 'extra' galaxies lie in the range $1 < z < 3$, precisely the range over which BzK-selected galaxies are found. Furthermore the sky densities of the two samples used here are significantly different. The sky density of sBzKs from the UKIDSS UDS is ~ 21200 per square degree, while the density of those from the Kitzbichler & White (2007) model is ~ 47500 per square degree. The number of sBzKs produced by the model is therefore more than twice as many as in the real universe, despite the red satellite problem outlined earlier.

These findings could be easily confirmed by inspection of the number counts. These are plotted alongside the number counts from the UDS BzK-selected galaxies in Figure 5.9. The passive BzK number counts do not show the turn-over which was expected and do not even plateau as seen in the real data. Furthermore the brightest pBzKs appear to be ~ 0.5 magnitudes dimmer in the mock observations. There is a clear over-abundance of star-forming BzKs, as expected from the sky density values. The number counts are consistent with the hypothesis that halos of too small mass are being permitted to form relatively bright objects at high redshift. The clustering measurements of BzK-selected galaxies have therefore highlighted two known problems which are shared by a number of S-A models: the over-production of stars in modest halos and the early quenching of star-formation in satellite galaxies.

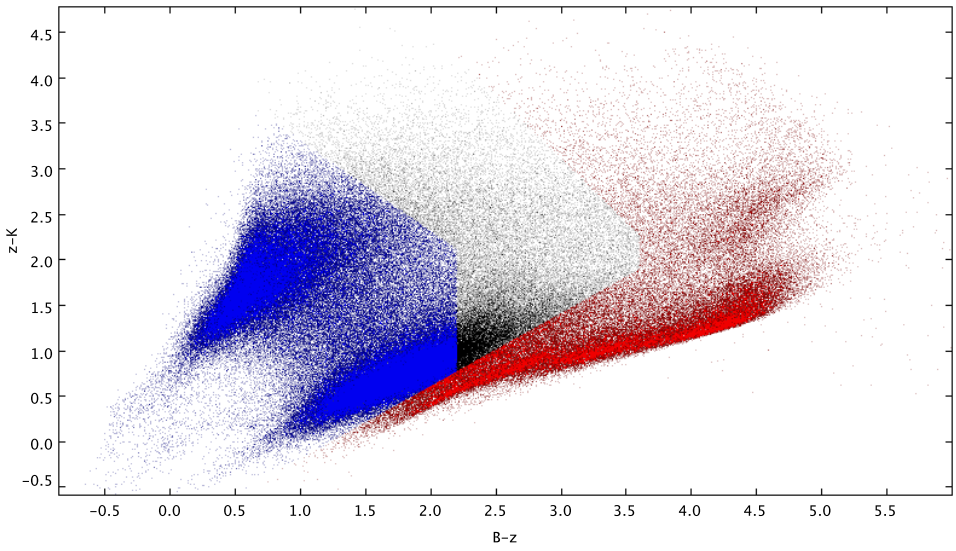

Figure 5.10. Selection diagram for the most conservative passive sample (red points) and star-forming sample (blue points) from chapter 3, applied to the mock catalogues of Kitzbichler & White (2007) (c.f. Figure 3.9). In the model, the $(B-z)-(z-K)$ plane is more fully populated by model galaxies than in the original selection, resulting in a small number of galaxies with ambiguous selection. In such cases the galaxies were assigned to the passive sample, but assigning them to either the star-forming or unused sample (black points) made no qualitative difference to the results.

The following section will further examine the clustering properties of star-forming and passive galaxies, covering the $0 < z < 3$ redshift range.

5.4 Clustering of passive and star-forming galaxies

In chapter 3 a simple set of boundaries were constructed to preferentially select a conservative passive sample and an actively star-forming sample. These boundaries are used in conjunction with the mock observations to create samples comparable with those of chapter 3. As stated in the previous section, the two filter sets (UDS/SXDS used in the observations and Johnson/SSDS used in the models) produce very similar colours. These selection boundaries applied to the model galaxies are shown in Figure 5.10. Blue points are those of the star-forming sample, red are passive galaxies and the black points are galaxies which have an intermediate degree of star-formation. The black points are not used in any of the subsequent analysis. The bluer (in $B-z$) part of the passive galaxy selection is more fully populated than in the observational data. As a result, the boundaries may not be defined precisely enough in this region to separate the passive and star-forming galaxies effectively. It should be noted that galaxies in this region of the diagram are expected to lie in the lowest redshift bin of the sample. The measurements have been repeated in this $0 < z < 0.5$ bin, with the selection boundary moved down by 0.1 in $z-K$ colour, and again by assigning the galaxies which change from passive to star-forming by this

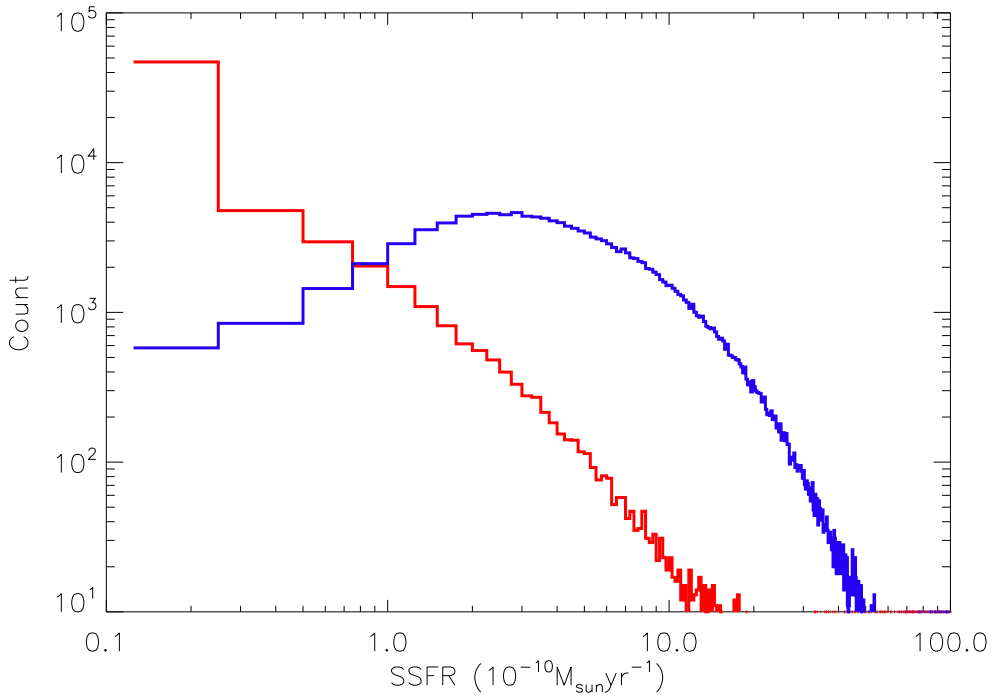

Figure 5.11. Histogram of the specific star-formation rates (SSFR) of the conservative passive and star-forming samples (red and blue respectively) for the model galaxies. The two samples are clearly different: the passive sample is indeed dominated by galaxies with extremely low SSFRs, while the star-forming sample have typical SSFR$\sim 3\times10^{-10}\mathrm{M}_{\odot}\mathrm{yr}^{-1}$. A typical galaxy in the star-forming sample will therefore double its stellar mass in less than a Hubble time.

process to the black sample instead. There were no qualitative differences between the three cases.

The observed samples were defined as passive or star-forming based on the minimum of knowledge about their star-formation histories. The model galaxies, on the other hand, have been tracked since the first collapse of their host halo. Hence in principle any desired method of separating star-forming from passive galaxies can be performed. In Figure 5.11 the specific star-formation rates (SSFR) for the two samples have been plotted. The simple photometric selection conditions appear to do a reasonably good job of separating those galaxies which are forming a significant amount of stars from those which are not. The mean SSFR of the star-forming sample is clearly greater than that of the passive sample. This argument is circular to some degree, as the same stellar population code is used to produce the model galaxy colours and the templates which are fit to the real galaxies' photometry. However, the models have been shown to reproduce a number of statistical properties of low-redshift galaxies (De Lucia & Blaizot 2007) and also reproduce the features of the $(B-z)-(z-K)$ plane without any further optimisation.

From the model star-forming and passive samples, correlation lengths were computed in the same way as those in chapter 3, with the exception that the cosmology appropriate to the Millennium simulation was used during de-projection. The vastly greater number of objects in the simulated data set allows measurements to be made

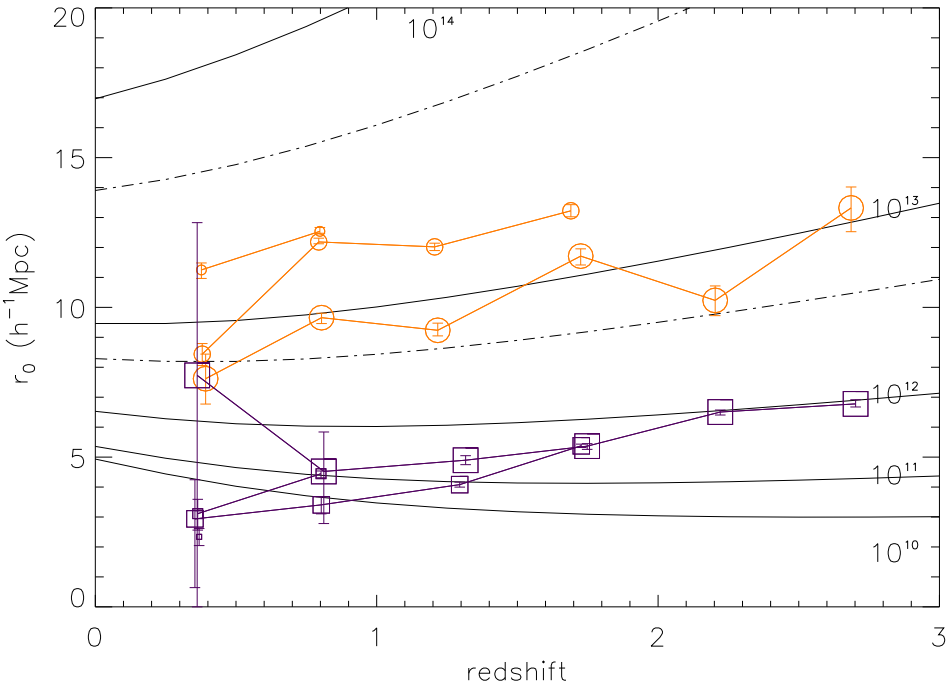

Figure 5.12. Correlation lengths derived for model galaxies using the passive and star-forming sample definitions of chapter 3. The symbols and colours have the same meanings as in Figure 3.7. The behaviour found in chapter 3 for the UDS galaxies is reproduced by the model and the separation of passive and star-forming galaxies is even more pronounced.

of passive sub-samples at greater redshift and of bright star-forming samples at low redshift. The apparent oscillatory evolution in clustering of the brightests passive samples is most likely spurious: possibly driven by the mock sample construction method, but most likely simply indicative that there are further sources of error which have not been accounted for. The measurements, together with the Mo & White (2002) halo mass predictions are shown in Figure 5.12. The cosmologies differ slightly between the observational measurements and those of the model galaxies hence the precise values of r_0 may disagree. However, Ω_m has only a small influence on clustering strength and the general trends should not depend on the underlying cosmology. These trends are discussed in the following.

In the observed galaxies, the passive galaxies were more strongly clustered for a given K-band luminosity for all sub-samples at $z < 1.5$. This behaviour is clearly reproduced by the model galaxies, and the distinction between the passive and star-forming samples is even more pronounced. The trend for dimmer passive galaxies to be more strongly clustered than brighter sub-samples at $z \sim 1$ is also clearly seen. Due to the over-efficient termination of star-formation in satellite galaxies in the mock samples, it might be expected that this luminosity inversion of clustering strength in passive galaxies would be exaggerated. However, there is no evidence for an excessive enhancement, and the implied halo masses of all passive sub-samples at any given redshift are very similar.

A second point of note in the results of the observed galaxies was that the clus-

tering strength of the star-forming galaxies clearly declined over the whole redshift range, $0 < z < 3$. This was interpreted as a manifestation of downsizing, as for a fixed rest-frame K-band luminosity the implied halo masses were found to be smaller at lower redshift. Reproduction of downsizing was among the main motivations for the modifications that De Lucia & Blaizot (2007) made to the model of Croton et al. (2006). However, they were primarily interested in downsizing in brightest cluster galaxies, and not $\sim L^*$ star-forming galaxies. Nevertheless, a smooth decrease in r_0 is found for star-forming galaxies in the models, and furthermore the absence of a strong luminosity dependence at fixed redshift is also well matched to the observations. The downsizing behaviour described here is slightly more pronounced in the observed sub-samples. However, this conclusion is drawn particularly from the highest redshift points and these are among the least certain, both in terms of the measured $w(\theta)$ fits and photometric redshift reliability. This should not be taken as a failure of the models then, but rather a point of possible tension which requires further investigation.

The last conclusion drawn from the observational results was the highly tentative possibility that the clustering strengths of passive and star-forming galaxies of a given luminosity become equal around $2 < z < 2.5$. This possibility is clearly disfavoured by the model galaxies. Equal K-band luminosities for passive and star-forming galaxies actually implies a more stellar-massive sample of passive galaxies. However, the lack of strong luminosity segregation in star-forming galaxies suggests that the order of magnitude difference in the implied host-halo masses is due to passivity, rather than mass segregation. Both this point, and the previous point could have a common cause. To be in better agreement with the observed clustering strengths, $\sim L^*$ star-forming galaxies would need to be hosted in more massive dark matter halos at $z \sim 3$. This in turn suggests that star-formation is still too efficient at high redshift, which may also be the same cause of the $1 < z < 3$ number counts excess. Highly energetic feedback has already been implemented to match the low-redshift luminosity function, and further suppressing the star-formation at high redshift may cause difficulties in matching the observations which find massive, passive galaxies in place by $z = 2$ (e.g. Zirm et al. 2007).

An alternative solution may be related to the over-production of passive galaxies seen in the BzK measurements. Perhaps those star-forming galaxies in massive halos are being turned too efficiently into passive galaxies irrespective of whether they are central or satellite galaxies. Allowing the central galaxies of massive halos to maintain a low level of star-formation may better replicate the observational results of chapter 3. It is clear from the measurements above that there is a preferred halo mass for passive galaxies. The radio mode feedback which is used to quench star-formation in massive galaxies requires a hot static halo to be formed. The cooling time in more massive dark matter halos is longer (White & Rees 1978; White & Frenk 1991), so at any epoch only the more massive halos will have been able to establish a hot static halo. This simple reasoning explains why there may be a preferred halo mass scale for passive galaxies. Hydrodynamic simulations have shown that cold gas streams on plunging orbits can penetrate to the centre of the halo (Kereš et al. 2009). If the tentative results of chapter 3 are confirmed, it may be that a star-formation fuel source such as this is required to match the distributions of passive galaxies.

5.5 Conclusions

Through the use of publicly-available mock catalogues generated from a semi-analytic model, the previous results of this thesis have been used to highlight some known issues with galaxy formation models and suggest possible new ones. The well-known over-production of red satellites and enhancement of K-band number counts were shown to impact clearly on the clustering measurements of BzK-selected galaxies. Hence, such measurements will be of great use in testing the next wave of improved models.

In addition to highlighting known issues, the unique measurements of clustering strength presented in chapter 3 have been used in a preliminary investigation to identify two further possible conflicts between models and observation. The down-sizing effect in high-redshift star-forming galaxies appears to be under-estimated in the model and the highly tentative equivalence of clustering strength between star-forming and passive galaxies is not seen. Neither of these are confirmed problems for the models at present, but may become so as further data are taken. As the body of data from observations increases, the constraints that the models must adhere to become ever tighter. Until recently, however, there have been very few such constraints from an unbiased high-redshift galaxy sample. The measurements presented throughout this thesis will therefore form a strong baseline for future model developments.

Chapter 6

Conclusions and further work

6.1 Conclusions

At the outset of this thesis, the bimodality of the galaxy population and some of the questions which it raises were outlined. Foremost among the open questions that are addressed by this body of work is when and where, statistically, the passive galaxy population is built up. I shall briefly describe the conclusions drawn throughout this thesis which are relevant to this central problem, before mentioning the other areas of galaxy evolution which are impacted upon by these results. Finally I shall look ahead to the directions in which this work may be extended in the future.

6.1.1 On the transition to passivity

In the low-redshift universe, galaxy bimodality is very strong. The vast majority of the (K-band) brightest galaxies are red and passively evolving. The use of 'passively evolving' typically refers to galaxies which have low specific star formation rates and have had so for some time ($\sim$ 1Gyr or longer). The recent star-formation history of these galaxies implies that their spectral energy distributions have a strong 4000Å break feature. Most of the work in this thesis relies primarily on photometric data and we lack detailed information regarding the star-formation history. 'Passive' in the context of this thesis therefore refers to a galaxy which, from its photometry, is expected to have a strong 4000Å break. At higher redshift there is evidence for a weakening of the galaxy bimodality. The red sequence becomes less distinct and the number of passive galaxies declines. As each galaxy must have been actively star-forming at some point in its history, it would seem reasonable that there is an early epoch at which the bimodality is absent. Making use of a unique new data set, the work presented in this thesis has sought to identify the epoch over which the bimodality is largely established, and furthermore to define the typical mass of dark matter halos which are hosts to passive galaxies as a function of redshift.

In chapter 2 a simple 2-colour selection technique was used to identify passive and star-forming galaxies over a wide redshift range. Originally designed to select galaxies within the range $1.4 < z < 2.5$, the true redshift ($z_{\rm phot}$) range is somewhat broader. However, it is also very peaked and so is largely representative of the $1 < z < 2$ range. To a magnitude limit of $K < 23$, the passively selected galaxies were found to be more strongly clustered than the star-forming galaxies. This measurement was interpreted as being due to a difference in the host halo masses: passive galaxies are hosted by more massive halos over this redshift range. The halo masses implied for these passive galaxies are in excess of $10^{13}{\rm M}_{\odot}$. Such halos would

be expected to evolve into massive groups or clusters by $z = 0$. The star-forming sample on the other hand are hosted by halo with masses an order of magnitude smaller. This was clear evidence that the bimodality seen at low-redshift was already established to some degree. However, the brightest star-forming galaxies were also very strongly clustered - at least as strongly as the passive galaxies. This second finding is precisely what we would expect to see if the passive population was in the process of being built up. Over the redshift range $1 < z < 2$ there is a mixture of passive and star-forming galaxies which, between them, will become dominant galaxies in systems which host passive galaxies by $z = 0$.

Leading on from this finding, in chapter 3 a more detailed study of the relative and absolute clustering properties of passive and star-forming galaxies was carried out. There was the possibility that the difference in clustering found in chapter 2 was due to the fact that passive galaxies are an intrinsically bright population at these redshifts. Brighter galaxies are found in more massive halos and so may cluster more strongly for that reason alone. Passive and star-forming samples were broken down into volume-limited sub-samples in luminosity and redshift of widths $\Delta M_K = 1$ and $\Delta z = 0.5$ respectively. For all sub-samples measurable, passive galaxies were found to cluster more strongly. Furthermore, the dependence of luminosity on clustering strength within the star-forming sub-samples was found to be very mild. The clear conclusion is that passivity is a stronger indicator of clustering strength (and hence host halo mass) than luminosity. Although a passive galaxy is more massive than a star-forming galaxy of equal luminosity, the lack of strong luminosity segregation suggests that this result will stand if sub-samples are constructed by stellar mass. The evolution of clustering scale length with redshift was also found to possibly differ between the two samples. Passive galaxies maintain roughly constant clustering strength across the whole redshift range accessible. It is difficult to say with certainty why this may be the case, and the measurements themselves will require further data to confirm this finding. However, if it is confirmed it implies passive galaxies are hosted by similar mass halos irrespective of redshift ($\mathrm{M_{halo}} \simeq 10^{13}\mathrm{M_{\odot}}$). The r_0 values of the star-forming samples increased steadily at higher redshifts. Such evolution implies that the clustering strengths of passive and star-forming galaxies of a given luminosity will become equal at some point within the redshift range $2 < z < 3$. This fascinating possibility is just beyond the capabilities of the current data set. If the trends were to be extrapolated further, then it implies that the star-forming galaxies may be the more clustered population. In fact this is highly unlikely as it would require those star-forming galaxies to become less clustered once they became passive.

Part of the trouble in probing the $z > 2$ universe is the need to break the galaxy sample into small sub-samples to compare the behaviour of galaxies with different properties. This requirement can be partially avoided by using marked correlation functions, and this is the approach taken in chapter 4. In chapter 3 it was found that of the red galaxies, passive galaxies cluster more strongly than red, dusty, star-forming galaxies. Marked correlation functions were used to test this finding, but due in part to the difficulty of defining a suitable 'passivity' mark, the results were inconclusive. For the same reasons the measurements were unable to sufficiently highlight any possible dependence of clustering on passivity at $z > 2$. At $z < 2$ however, it is clear that galaxy clustering depends more strongly on passivity and colour than K-band luminosity.

Finally, in chapter 5 the mock galaxies from a recent semi-analytical model were

used in combination with the observation measurements to suggest areas where our understanding of galaxy formation is incomplete. Comparison of $w(\theta)$ from the model galaxies and observed galaxies for BzK-selected samples highlighted two known problems. The first is the over-production of faint, high redshift galaxies, seen in chapter 5 as a factor ~ 2 discrepancy in the sky density of star-forming BzK galaxies and a depression of the clustering amplitude. The second is the over-efficient transition from blue to red when a galaxy becomes a satellite in a massive halo. The effect of these two issues was to exaggerate the difference in clustering amplitude of the passive and star-forming BzKs. The more detailed study presented in chapter 3 is ideally suited to comparison with the mock galaxies as no such measurements were available when the model was constructed. The r_0 measurements of star-forming and passive sub-samples at $1 < z < 3$ are therefore a genuine prediction by the model. The results of the mock galaxies and the observed galaxies were found to broadly agree. Even the implied halo masses were in reasonable agreement. However, the equivalence of star-forming and passive galaxy clustering by $z = 3$ that was suggested by the observations was not seen in the mock galaxy results. If this is found to be a real failing of the current model and not just driven by a lack of observational data, then the implications for our understanding of galaxy evolution could be substantial.

6.1.2 Downsizing seen through clustering

The work presented in this thesis has implications for the theory of galaxy formation beyond the central theme of the build up in galaxy bimodality. The observation that the most massive galaxies were the first to have formed their stars was termed 'downsizing' by Cowie et al. (1996). This finding has since been confirmed and shown to manifest itself in different ways by a great number of authors. The work presented in this thesis is no exception and evidence for downsizing can be found in many of the results presented. The number counts of the passive BzK-selected galaxies turn over at $K_{AB} \sim 22$, indicating a lack of faint passive galaxies at their epoch. This result is confirmed by the luminosity functions shown in section 2.6.1: passive BzKs are among the brightest galaxies at their redshift. As they are already passive their stars much have been formed at some even earlier epoch, meaning that downsizing is in evidence even at $z > 1$. Furthermore their host halos are among the most massive, implying group or cluster mass halos by $z = 0$. These findings are all consistent with the massive group and cluster-dominant galaxies at low redshift being the first to form their stars. There is further evidence for downsizing in the correlation length measurements in chapter 3. Star-forming galaxies of a given luminosity are hosted by more massive halos at higher redshift. This trend is also seen in the measurements of the semi-analytic galaxies. However, the evolution appears to be milder in the mock sample. Whether this is due to limitations of the data or an issue with the model requires further investigation.

6.2 Future work

6.2.1 Short-term follow-up

The central result of this thesis is in chapter 3, where it was found that passive galaxies are more strongly clustered than star-forming galaxies for $z < 1.5$. For simplicity, ease of comparison with future work and robustness, the galaxies were separated in to sub-samples by rest-frame K-band luminosity. It could be argued

that a galaxy's stellar mass is a more fundamental quantity than luminosity. As galaxies fade over time after they have finished forming the bulk of their stars, passive galaxies are more stellar massive than an equally luminous star-forming galaxy. Although the luminosity segregation (and hence mass segregation) in clustering is very mild, it cannot be entirely ruled out that passive and star-forming galaxies of equal stellar mass are equally clustered. There are data already in hand which will be used in the near future to further improve the photometric redshifts of the UDS galaxies and produce reliable stellar masses. At that time, the analysis presented in chapter 3 could be repeated, though I fully expect the result to stand.

Although the passive galaxy clustering was probed out to $z = 2$ in chapter 3, the lack of sufficient sources made firm conclusions difficult to draw at such a great distance. A lack of depth restricted how faint a sample could be, while the brighter samples were restricted by the survey volume. This latter restriction will not be overcome by the UDS alone, but in combination with other survey data in the longer-term future. However, it will be possible to push the luminosity limit up to a magnitude fainter over the next 1-2 years. Such a gain in depth will not affect the bright samples, but enable the study of sub-L^* populations out to $z = 3$. The clustering, and indeed number, of faint passive systems at such redshifts will be fascinating results.

The marked-correlation function technique presented in chapter 4 has great potential to not only confirm the findings of the auto-correlation function analyses, but to also enable us to learn about environmental effects as a function of scale. The short introductory study presented in this thesis does not make use of all of the information in the MCF. In computing de-projected correlation lengths a constant slope of $\delta = 0.8$ was assumed throughout. These lengths were then used to learn how the clustering strength depends on a quantity such as colour or passivity. Such an assumption is not necessary in marked-correlation functions, as the dependence of the mark is found directly from the MCF at any scale. MCFs have not been used extensively in the literature. Comparison between two studies is not as straight forward as comparing two values for r_0. Furthermore, there are hidden biases, such as the implicit assumption that the redshift distribution of the sample does not depend upon the mark. However, coupled with the following area of future research, marked-correlation functions will be of great use in the future.

The halo model is the foundation of much of the current theoretical work. The halo occupation distribution branch of this framework allows us to gain further details of a population's host halos from clustering statistics. Applying the technique to the $w(\theta)$ measurements computed for this thesis should produce increased insight into how and where the passive galaxy population builds up. Questions such as whether there is a minimum mass for a galaxy to become passive, or whether the difference in clustering between passive and star-forming galaxies is due in part to very different occupation distributions may be answered by use of the HOD technique. Furthermore, application of HOD to the marked-correlation functions may allow us to take account of mark-dependent redshift distributions. I am unaware of any current work using a MCF and HOD together, but the potential gains make it a path well worth pursuing.

In addition to gains in photometry and developments in theoretical work, the coming year will see the release of the UDSz. The UDSz is a spectroscopic survey, targeting almost 3000 galaxies in the UDS field, with the majority pre-selected by z_{phot} to lie at $z > 1$. These data will have a great impact upon the reliability of

the photometric redshifts, and all of the quantities derived from them, but perhaps the greatest scientific gains will come from the data themselves or the follow-up observations that they will precipitate. This thesis has been a good step forward in our understanding of the statistical build up of the passive population, but the mechanism remains a mystery. Whether it is due to the action of an AGN or some other means, it is likely that targeted observations or detailed knowledge of small-scale environments will be required to confirm our theories.

6.2.2 Long-term outlook

At several points in this thesis, the need for additional fields comparable to the UDS has been mentioned. The new VISTA instrument has now begun taking data and there are two projects in particular using it that are worth mentioning in the context of this work. The first of these two is the UltraVISTA survey. Covering an area similar to the UDS, it is planned to go even deeper in K-band depth than the UDS. This new field, together with the UDS and the cosmos field (which now also has deep K-band data), will allow the effects of sample variance to be fully quantified. Moreover the additional numbers of sources will reduce the variance of the measurements due to Poisson errors. Such an advancement may not seem too important, but it is likely that much better constraint of errors will be required to get the most out of the HOD analysis proposed above. The second project running on VISTA is the VIDEO survey. The intention is for this survey to be slightly shallower than the current UDS depth, but over $\sim 12\mathrm{deg.}^2$. As most passive galaxies at any redshift are among the rarest objects, survey area is highly important. This survey therefore may be able to push the present analysis to brighter luminosities, of which there are simply too few in the UDS for reliable measurements.

The SXDS/UDS field has become one of the most important fields for extra-galactic research worldwide. This status was reinforced recently when time was awarded to use the Hubble space telescope to observe a significant fraction of the field with WFC 3 in the J and H-bands. These data, once taken, will enable the study of morphologies of the same galaxies used for analysis in this thesis. The possible links between morphology, colour and star-formation of an unbiased sample at high redshift is sure to be an area of intense research. The UDS is also one of the key fields for for the HERMES survey utilising the recently launched Herschel satellite. Together with the SCUBA-2 data which are planned for the field, these data will provide an independent method of identifying high redshift dusty star-forming galaxies. Once all of the approved observations have been made, the combination of depth, area, range and quality of data available in the UDS field will be unprecedented.

In recent years, observational results have been the source of much of our increased understanding of the universe. Studies such as those presented in this thesis break new ground and present new challenges for galaxy formation models. The addition of ingredients in models that are required to match observations in turn motivates further data collection. In chapter 5, a possible disagreement between observed measurements and model predictions was highlighted. There must first be additional evidence to confirm whether this disagreement is indeed genuine or simply a result of poor statistics. But if confirmed, it would then fall to those working on such models to find why the discrepancy exists. This process is ongoing, and in each instance our knowledge of the workings of the universe increases. The real long-term outlook is that there is no end in sight, and there will be more to learn

for many years to come.

Appendix A

Appendix

In this appendix I detail some tests of how the choice of mark impacts upon the resulting marked-correlation function. Though useful in interpreting the results of the marked-correlation functions, there remains much to be done before a complete understanding can be achieved.

A.1 Binary marks

The simplest choice of mark is to separate two populations, such as the red and blue samples in the previous chapter, and give one sample a mark, $M = 1$, and the other $M = 0$. Choosing marks in this way is not the most instructive but allows simple comparison between the results of the measurements and analytic solutions.

Writing the marked correlation function in terms of pair counts we have

$$M(\theta) = \frac{W(\theta) + 1}{w(\theta) + 1} = \frac{1 + (WW - 2WR_W + R_W R_W)/R_W R_W}{1 + (DD - 2DR + RR)/RR}. \tag{A.1}$$

The subscript 'W' here represents the fact that the random objects are given a mark from the overall distribution. Mark-weighted pairs can be expanded as follows:

$$WW = D_1 D_1 + D_1 D_0 + D_0 D_1 + D_0 D_0, \tag{A.2}$$

where the subscripts refer to the mark of the contributing galaxies. Clearly only the $D_1 D_1$ pairs will contribute to the sum. Following a similar argument for the WR_W and $R_W R_W$ pair counts, we have

$$W(\theta) + 1 = 1 + (D_1 D_1 - 2D_1 R_1 + R_1 R_1)/R_1 R_1. \tag{A.3}$$

The weighted correlation function is therefore simply the auto-correlation function of those galaxies which have been given a mark $M = 1$.

Using this information it is possible to construct predictions for the marked correlation functions of galaxy samples from the UKIDSS UDS. In Chapter 2, correlation function amplitudes were fit to the pBzK and sBzK selections, assuming a fixed slope of $\delta = -0.8$. Following the same procedure for a composite population, comprising both sBzK and pBzK galaxies, the amplitude was found to be $A = 2.234 \times 10^{-3}$. A marked correlation function of the composite sample, assigning a mark of $M = 1$ to pBzKs and $M = 0$ to sBzKs should follow

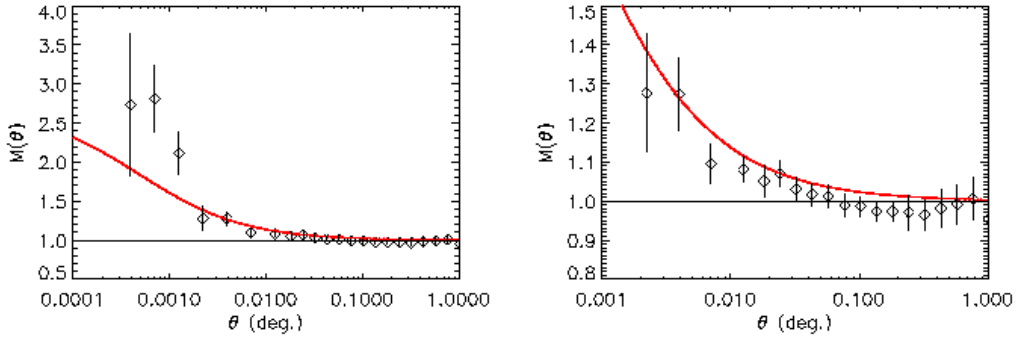

Figure A.1. Marked correlation function for BzK-selected galaxies where pBzKs are given a mark of M=1 and sBzKs are given M=0. The red line is the prediction from power-law fits to the auto-correlation functions (see text). The data are the same in each panel, but the plotting ranges have been altered for clarity.

$$M(\theta) = \frac{1 + (5.99 \times 10^{-3})\theta^{-0.8}}{1 + (2.23 \times 10^{-3})\theta^{-0.8}} \tag{A.4}$$

at large scales ($\theta > 0.01$ deg.). This line, together with the actual measurements for $M(\theta)$ are shown in Figure A.1. Over the scales that were used to fit the power-law to the auto-correlation functions, the measurements are consistently below the prediction. Furthermore, at the very smallest scales there is a significant enhancement over the prediction. This second point alone would suggest that the pBzK sample is more influenced by multiple occupation of halo than the combined sample.

The is one part of the correlation function which has been neglected up until now: the bias due to finite field size. This bias is corrected for by an integral constraint, the value of which is determined from the survey geometry, correlation function slope and amplitude. The first and second of these have been kept constant, but the amplitude of the pBzK ACF is almost three times that of the combined sample. Rather than include this bias correction in the data, the prediction can be modified as so:

$$M(\theta) = \frac{1 + (5.99 \times 10^{-3}(\theta^{-0.8} - C))}{1 + (2.23 \times 10^{-3}(\theta^{-0.8} - C))}, \tag{A.5}$$

where the numerical value of C is ~ 2.5 (see Chapter 2). The modified prediction is shown in Figure A.2. The agreement between the measurements and the prediction is improved, and the vast majority of the points are now consistent with the power-law fit to 1σ, as expected.

However, the agreement can be improved even further. In Figure A.3 the power-law slope for the pBzK ACF has been allowed to vary. The slope is now $\delta = -1$ and the amplitude $A = 2.0^{-3}$. For a slope of $\delta = -1$ the constant $C' = 3.35$ and hence

$$M(\theta) = \frac{1 + (2.0 \times 10^{-3}(\theta^{-1.0} - C'))}{1 + (2.23 \times 10^{-3}(\theta^{-0.8} - C))}. \tag{A.6}$$

This is only tentative evidence, but suggests that with larger samples the slope of

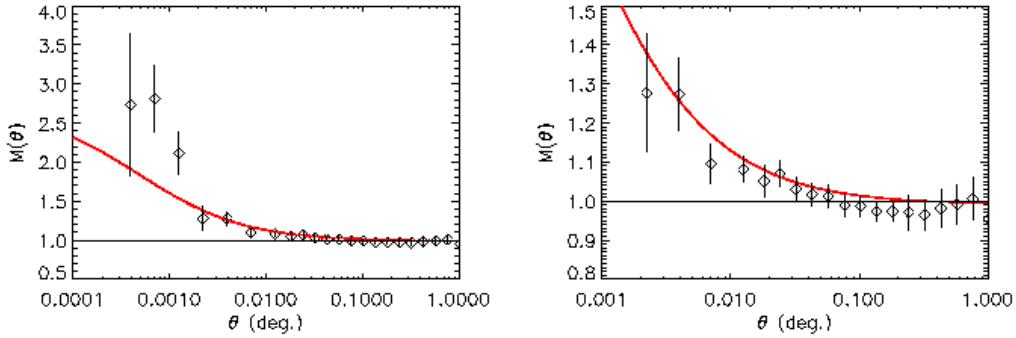

Figure A.2. Same as Figure A.1, but in this case the prediction has been modified to take account of the integral constraint.

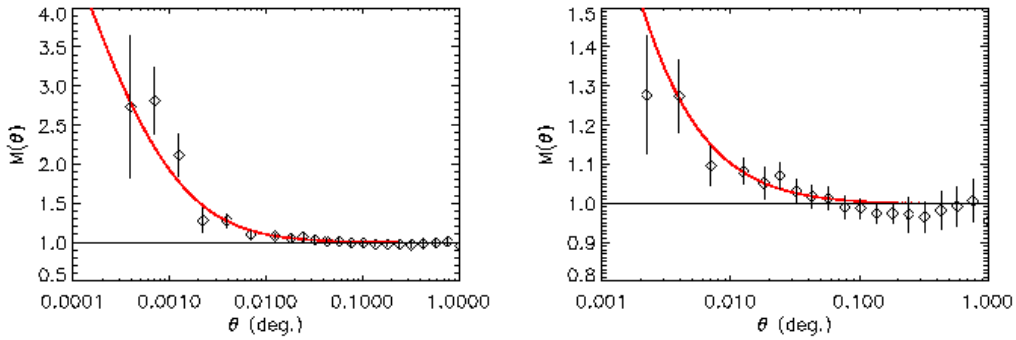

Figure A.3. Same as Figure A.2, but the power-law slope and amplitude of the pBzK auto-correlation function have been changed to $\delta = -1$ and $A = 2.0 \times 10^{-3}$ respectively.

the pBzK auto-correlation function will be found to be steeper than the canonical $\delta = -0.8$. It should be noted that McCracken et al. (2010) find a steeper slope than $\delta = -0.8$ for their pBzK sample. However, they fit over the entire angular range of their data, and are therefore possibly including halo effects. A steeper slope clearly fits the single halo regime much better in the marked correlation function too.

A.2 Discrete and continuous marks

The marking scheme used in the previous subsection is generally not one that is used as it offers no great benefit over using auto-correlation function. In this subsection I shall extend the previous discussion to more realistic scenarios, beginning with a discrete case in which the less clustered galaxies are now also given a mark.

Using the same samples as previously (the pBzK and sBzK samples from Chapter 2), the strongly clustered sample have been given a mark of $M = 4$, and the more weakly clustered sample a mark of $M = 1$. These values are entirely arbitrary but are used in this case to highlight the behaviour of the marked correlation function. The MCF for these marks is shown in Figure A.4. In this case the measurements are slightly suppressed with respect to the earlier mark scheme. The sBzK galaxies now

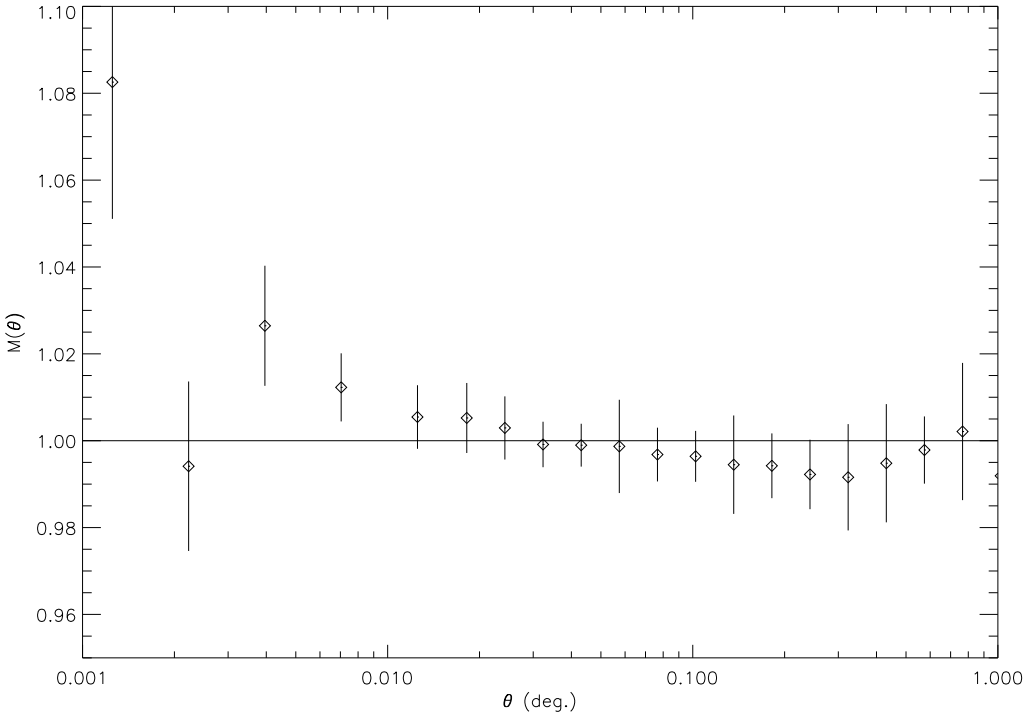

Figure A.4. MCF of the combined pBzK and sBzK sample, with pBzKs given a mark $M = 4$ and sBzKs a mark $M = 1$.

contribute to pair counts and are hosted in some of the same halos as the pBzKs. In fact the only points significantly above unity are at scales that suggest they are due to pairs of galaxies within single halos. However, it should be stressed that this is simply a test case without a physically motivated mark.

Though discrete marks may be useful in certain circumstances (for example, if one wished to test whether earlier type spirals had any environmental dependence, one could assign a mark of $M = 1$ to Sd type spirals up to $M = 4$ for Sa types), to get the most out of a MCF, continuous distributions are used. In the following two distributions are used. The first comprises two Gaussians with a small amount of overlap, designed to mimic a colour distribution but exaggerated so that comparison can be made with the previous discrete case. The pBzK sample were given a random mark from one of the Gaussians, while the sBzK marks were drawn from the other. The two resulting mark distributions, together with the overall distribution are shown in Figure A.5.

The marked correlation function resulting from this mark distribution is shown in Figure A.6. As the mark distribution is blurred with respect the the discrete case, so is the MCF: the halo-scale points that are greater than unity are now barely significant.

The second generated distribution was designed to approximate the logarithm of a quantity such as stellar mass. The mark distributions of the pBzKs and sBzKs are shown in Figure A.7. These marks were used in two ways: firstly in a MCF performed as before (left-hand panel of Figure A.8); secondly they were used as the power for an alternative mark, $M' = 10^M$, which was used to produce the MCF

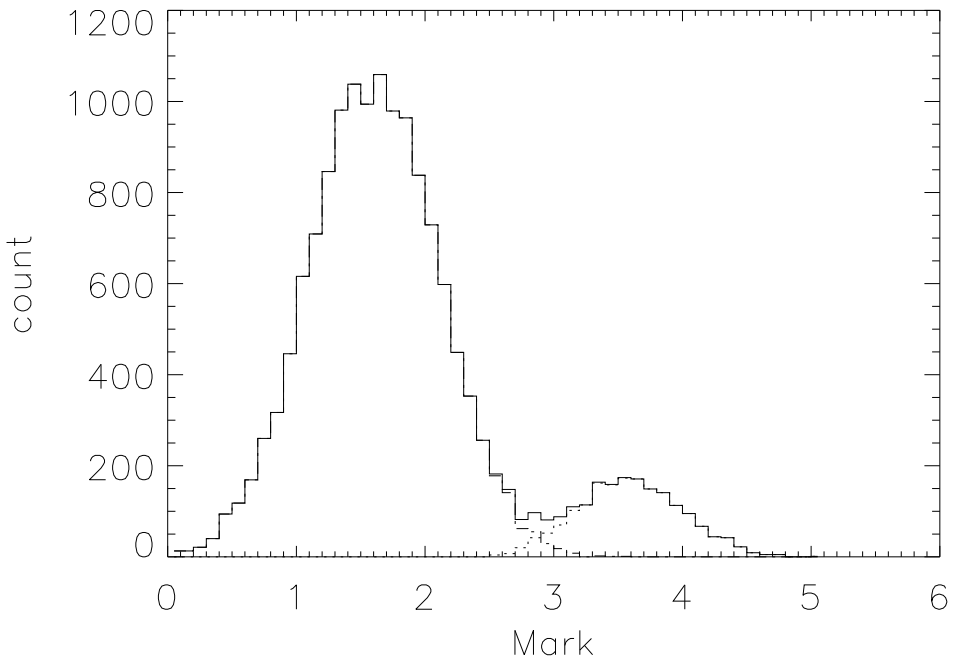

Figure A.5. Histogram of marks of pBzKs (dotted line) and sBzK (dashed line) with the overall distribution (solid line). These marks were intended to approximate a colour distribution but exaggerated to occupy a similar range to the discrete case.

in the right-hand panel of Figure A.8. In this way it is possible to investigate how robust the MCF is against taking logs of the mark of interest.

The first thing to notice is the similarity between the left-hand panel of Figure A.8 and Figure A.6, which suggests that a modest change in the shape of the mark distribution does not greatly effect the marked correlation function. The two panels of Figure A.8, however, do differ slightly. This is not entirely surprising given that the range of M' is $0 < M^{prime} < 100,000$ (rather than $0 < M < 4$) and the mark distribution is greatly distorted. However, the differences are mainly in the less certain points and they are still more similar to each other than the binary example given earlier (though note the difference in scales on the y-axes). The tests run here generally confirm the statement of Skibba et al. (2009) that MCF measurements are relatively insensitive to such rescaling.

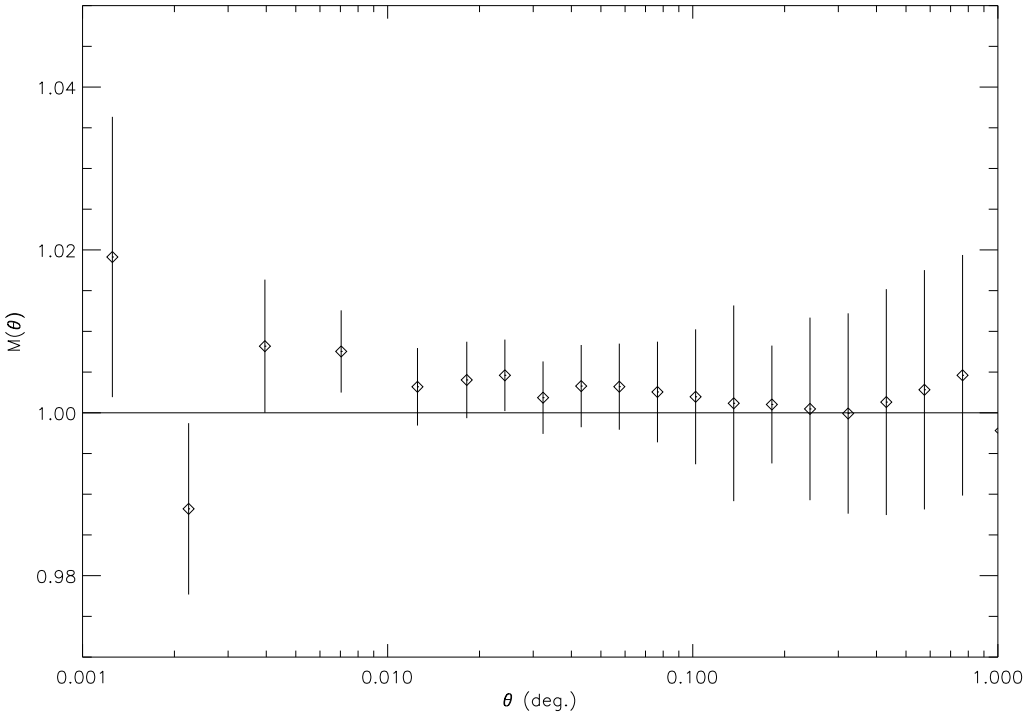

Figure A.6. MCF of the 'U-B' marked BzK-galaxies using the distribution in Figure A.5.

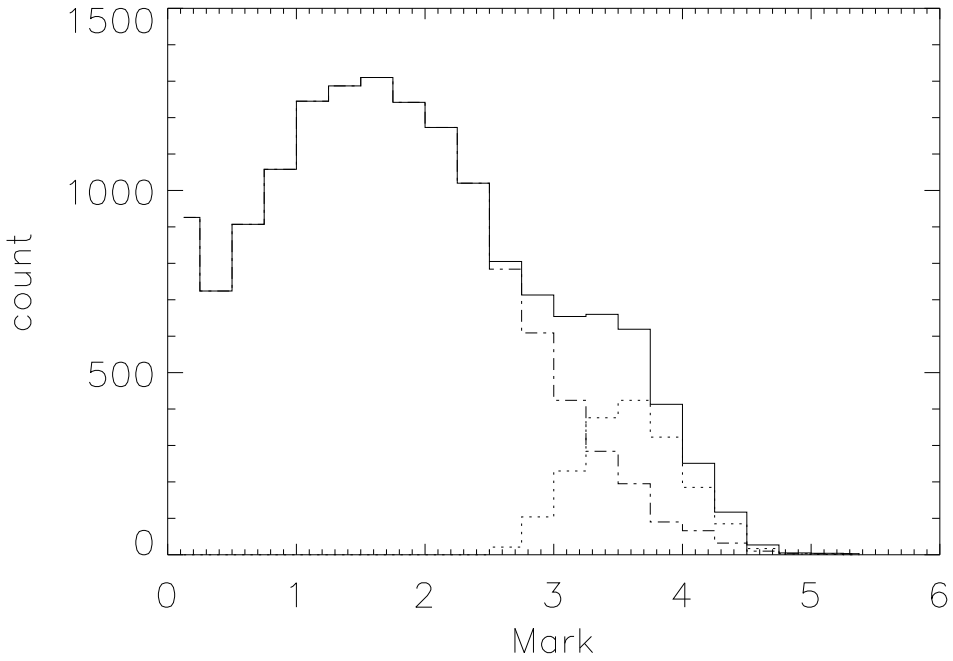

Figure A.7. Histogram of marks designed to approximate a stellar mass distribution. The line-styles have the same meanings as in Figure A.5.

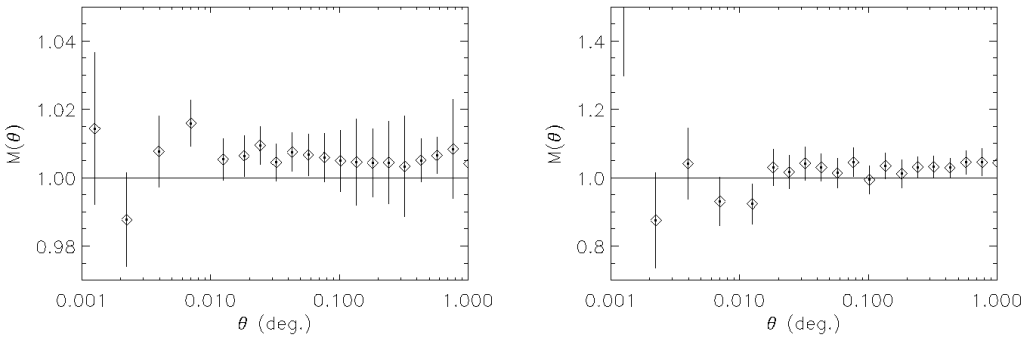

Figure A.8. Left: Marked correlation function using the mark shown in Figure A.7. **Right:** MCF using $M' = 10^M$ (see text).

Bibliography

Abell G. O., 1965, ARA&A, 3, 1

Adelberger K. L., Steidel C. C., Shapley A. E., Hunt M. P., Erb D. K., Reddy N. A., Pettini M., 2004, ApJ, 607, 226

Alimi J., Valls-Gabaud D., Blanchard A., 1988, A&A, 206, L11

Arribas S., Colina L., Clements D., 2001, ApJ, 560, 160

Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, ApJ, 615, L101

Bamford S. P., Nichol R. C., Baldry I. K., Land K., Lintott C. J., Schawinski K., Slosar A., Szalay A. S., Thomas D., Torki M., Andreescu D., Edmondson E. M., Miller C. J., Murray P., Raddick M. J., Vandenberg J., 2009, MNRAS, 393, 1324

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191

Baum W. A., 1962, in G. C. McVittie ed., Problems of Extra-Galactic Research Vol. 15 of IAU Symposium. pp 390–+

Beisbart C., Kerscher M., 2000, ApJ, 545, 6

Bell E. F., Wolf C., Meisenheimer K., Rix H., Borch A., Dye S., Kleinheinrich M., Wisotzki L., McIntosh D. H., 2004, ApJ, 608, 752

Bennett C. L., Banday A. J., Gorski K. M., Hinshaw G., Jackson P., Keegstra P., Kogut A., Smoot G. F., Wilkinson D. T., Wright E. L., 1996, ApJ, 464, L1+

Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38

Benson A. J., Kamionkowski M., Hassani S. H., 2005, MNRAS, 357, 847

Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Davé R., Frenk C. S., Jenkins A., Katz N., Lacey C. G., 2003, ApJ, 593, 1

Bernstein G. M., 1994, ApJ, 424, 569

Binney J., Tremaine S., 1987, Galactic dynamics

Blain A. W., Chapman S. C., Smail I., Ivison R., 2004, ApJ, 611, 725

Blanc G. A., Lira P., Barrientos L. F., Aguirre P., Francke H., Taylor E. N., Quadri R., Marchesini D., Infante L., Gawiser E., Hall P. B., Willis J. P., Herrera D., Maza J., 2008, ApJ, 681, 1099

Boerner G., Mo H., Zhou Y., 1989, A&A, 221, 191

Bolzonella M., Miralles J., Pelló R., 2000, A&A, 363, 476

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Bower R. G., Lucey J. R., Ellis R. S., 1992, MNRAS, 254, 601

Brammer G. B., Whitaker K. E., van Dokkum P. G., Marchesini D., Labbé I., Franx M., Kriek M., Quadri R. F., Illingworth G., Lee K., Muzzin A., Rudnick G., 2009, ApJ, 706, L173

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bundy K., Ellis R. S., Conselice C. J., Taylor J. E., Cooper M. C., Willmer C. N. A., Weiner B. J., Coil A. L., Noeske K. G., Eisenhardt P. R. M., 2006, ApJ, 651, 120

Burles S., Nollett K. M., Turner M. S., 2001, ApJ, 552, L1

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Carlberg R. G., Cowie L. L., Songaila A., Hu E. M., 1997, ApJ, 484, 538

Carroll S. M., Press W. H., Turner E. L., 1992, ARA&A, 30, 499

Casali M., Adamson A., Alves de Oliveira C., Almaini O., Burch K., Chuter T., Elliot J., Folger M., Foucaud S., Hambly N., Hastie M., others 2007, A&A, 467, 777

Chabrier G., 2003, PASP, 115, 763

Charlot S., Fall S. M., 2000, ApJ, 539, 718

Cimatti A., Cassata P., Pozzetti L., Kurk J., Mignoli M., Renzini A., Daddi E., Bolzonella M., Brusa M., Rodighiero G., Dickinson M., Franceschini A., Zamorani G., Berta S., Rosati P., Halliday C., 2008, A&A, 482, 21

Cimatti A., Daddi E., Renzini A., 2006, A&A, 453, L29

Cimatti A., Mignoli M., Daddi E., Pozzetti L., Fontana A., Saracco P., Poli F., Renzini A., Zamorani G., Broadhurst T., Cristiani S., D'Odorico S., Giallongo E., Gilmozzi R., Menci N., 2002, A&A, 392, 395

Cirasuolo M., McLure R. J., Dunlop J. S., Almaini O., Foucaud S., Simpson C., 2010, MNRAS, 401, 1166

Cirasuolo M., McLure R. J., Dunlop J. S., Almaini O., Foucaud S., Smail I., Sekiguchi K., Simpson C., Eales S., Dye S., Watson M. G., Page M. J., Hirst P., 2007, MNRAS, 380, 585

Coil A. L., Newman J. A., Cooper M. C., Davis M., Faber S. M., Koo D. C., Willmer C. N. A., 2006, ApJ, 644, 671

Coil A. L., Newman J. A., Croton D., Cooper M. C., Davis M., Faber S. M., Gerke B. F., Koo D. C., Padmanabhan N., Wechsler R. H., Weiner B. J., 2008, ApJ, 672, 153

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

Cole S., Norberg P., Baugh C. M., Frenk C. S., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Cross N., others 2001, MNRAS, 326, 255

Cole S., Percival W. J., Peacock J. A., Norberg P., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn J., others 2005, MNRAS, 362, 505

Coleman G. D., Wu C., Weedman D. W., 1980, ApJS, 43, 393

Coles P., Lucchin F., 1995, Cosmology. The origin and evolution of cosmic structure

Conselice C. J., 2003, ApJS, 147, 1

Conselice C. J., O'Neil K., Gallagher J. S., Wyse R. F. G., 2003, ApJ, 591, 167

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Coppin K., Chapin E. L., Mortier A. M. J., Scott S. E., Borys C., Dunlop J. S., Halpern M., Hughes D. H., Pope A., Scott D., Serjeant S., Wagg J., Alexander D. M., Almaini O., others 2006, MNRAS, 372, 1621

Cowie L. L., Gardner J. P., Hu E. M., Songaila A., Hodapp K., Wainscoat R. J., 1994, ApJ, 434, 114

Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839

Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006, MNRAS, 365, 11

Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, ApJ, 617, 746

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

De Lucia G., Poggianti B. M., Aragón-Salamanca A., Clowe D., Halliday C., Jablonka P., Milvang-Jensen B., Pelló R., Poirier S., Rudnick G., Saglia R., Simard L., White S. D. M., 2004, ApJ, 610, L77

Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, A&A, 350, 381

Dressler A., 1980, ApJ, 236, 351

Drory N., Bender R., Feulner G., Hopp U., Maraston C., Snigula J., Hill G. J., 2004, ApJ, 608, 742

Dunlop J. S., McLure R. J., Kukula M. J., Baum S. A., O'Dea C. P., Hughes D. H., 2003, MNRAS, 340, 1095

Dye S., Warren S. J., Hambly N. C., Cross N. J. G., Hodgkin S. T., Irwin M. J., Lawrence A., Adamson A. J., Almaini O., Edge A. C., Hirst P., Jameson R. F., Lucas P. W., van Breukelen C., others 2006, MNRAS, 372, 1227

Einasto M., 1991, MNRAS, 252, 261

Ellis R. S., Abraham R. G., Dickinson M., 2001, ApJ, 551, 111

Erb D. K., Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., Hunt M. P., Moorwood A. F. M., Cuby J., 2003, ApJ, 591, 101

Faber S. M., Jackson R. E., 1976, ApJ, 204, 668

Fabian A. C., 1999, MNRAS, 308, L39

Fabian A. C., Sanders J. S., Taylor G. B., Allen S. W., Crawford C. S., Johnstone R. M., Iwasawa K., 2006, MNRAS, 366, 417

Farrah D., Lonsdale C. J., Borys C., Fang F., Waddington I., Oliver S., Rowan-Robinson M., Babbedge T., Shupe D., Polletta M., Smith H. E., Surace J., 2006, ApJ, 643, L139

Font A. S., Bower R. G., McCarthy I. G., Benson A. J., Frenk C. S., Helly J. C., Lacey C. G., Baugh C. M., Cole S., 2008, MNRAS, 389, 1619

Foucaud S., Almaini O., Smail I., Conselice C. J., Lane K. P., Edge A. C., Simpson C., Dunlop J. S., McLure R. J., Cirasuolo M., Hirst P., Watson M. G., Page M. J., 2007, MNRAS, 376, L20

Franx M., Labbé I., Rudnick G., van Dokkum P. G., Daddi E., Förster Schreiber N. M., Moorwood A., Rix H., Röttgering H., van de Wel A., van der Werf P., van Starkenburg L., 2003, ApJ, 587, L79

Furusawa H., Kosugi G., Akiyama M., Takata T., Sekiguchi K., Tanaka I., Iwata I., Kajisawa M., Yasuda N., Doi M., Ouchi M., Simpson C., Shimasaku K., Yamada T., Furusawa J., Morokuma T., Ishida C. M., others 2008, ApJS, 176, 1

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Gebhardt K., Bender R., Bower G., Dressler A., Faber S. M., Filippenko A. V., Green R., Grillmair C., Ho L. C., Kormendy J., Lauer T. R., Magorrian J., Pinkney J., Richstone D., Tremaine S., 2000, ApJ, 539, L13

Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146

Grazian A., Salimbeni S., Pentericci L., Fontana A., Nonino M., Vanzella E., Cristiani S., de Santis C., Gallozzi S., Giallongo E., Santini P., 2007, A&A, 465, 393

Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385

Guth A. H., Jagannathan K., 1998, American Journal of Physics, 66, 94

Guzman R., Lucey J. R., Bower R. G., 1993, MNRAS, 265, 731

Hamilton A. J. S., 1993, ApJ, 417, 19

Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, ApJ, 374, L1

Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039

Hartley W. G., Lane K. P., Almaini O., Cirasuolo M., Foucaud S., Simpson C., Maddox S., Smail I., Conselice C. J., McLure R. J., Dunlop J. S., 2008, MNRAS, 391, 1301

Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75

Hayashi M., Shimasaku K., Motohara K., Yoshida M., Okamura S., Kashikawa N., 2007, ApJ, 660, 72

Heinis S., Milliard B., Arnouts S., Blaizot J., Schiminovich D., Budavári T., Ilbert O., Donas J., Treyer M., Wyder T. K., McCracken H. J., others 2007, ApJS, 173, 503

Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454

Hopkins P. F., Bundy K., Hernquist L., Wuyts S., Cox T. J., 2010, MNRAS, 401, 1099

Huang J., Glazebrook K., Cowie L. L., Tinney C., 2003, ApJ, 584, 203

Hubble E. P., 1936, Realm of the Nebulae

Hughes D. H., Serjeant S., Dunlop J., Rowan-Robinson M., Blain A., Mann R. G., Ivison R., Peacock J., Efstathiou A., Gear W., Oliver S., Lawrence A., Longair M., Goldschmidt P., Jenness T., 1998, Nature, 394, 241

Ilbert O., Arnouts S., McCracken H. J., Bolzonella M., Bertin E., Le Fèvre O., Mellier Y., Zamorani G., Pellò R., Iovino A., Tresse L., Le Brun V., Bottini D., others 2006, A&A, 457, 841

Imai K., Pearson C. P., Matsuhara H., Wada T., Oyabu S., Takagi T., Fujishiro N., Hanami H., 2008, ApJ, 683, 45

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jing Y. P., 1998, ApJ, 503, L9+

Jorgensen I., Franx M., Kjaergaard P., 1996, MNRAS, 280, 167

Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576

Kauffmann G., Heckman T. M., White S. D. M., Charlot S., Tremonti C., Brinchmann J., Bruzual G., Peng E. W., Seibert M., Bernardi M., Blanton M., others 2003, MNRAS, 341, 33

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

King A., 2005, ApJ, 635, L121

Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2

Kochanek C. S., Pahre M. A., Falco E. E., Huchra J. P., Mader J., Jarrett T. H., Chester T., Cutri R., Schneider S. E., 2001, ApJ, 560, 566

Kodama T., Yamada T., Akiyama M., Aoki K., Doi M., Furusawa H., Fuse T., Imanishi M., Ishida C., Iye M., Kajisawa M., Karoji H., Kobayashi N., Komiyama Y., others 2004, MNRAS, 350, 1005

Kong X., Daddi E., Arimoto N., Renzini A., Broadhurst T., Cimatti A., Ikuta C., Ohta K., da Costa L., Olsen L. F., Onodera M., Tamura N., 2006, ApJ, 638, 72

Lacey C., Cole S., 1993, MNRAS, 262, 627

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lane K. P., Almaini O., Foucaud S., Simpson C., Smail I., McLure R. J., Conselice C. J., Cirasuolo M., Page M. J., Dunlop J. S., Hirst P., Watson M. G., Sekiguchi K., 2007, MNRAS, 379, L25

Lawrence A., Warren S. J., Almaini O., Edge A. C., Hambly N. C., Jameson R. F., Lucas P., Casali M., Adamson A., Dye S., Emerson J. P., others 2007, MNRAS, 379, 1599

Le Fèvre O., Vettolani G., Garilli B., Tresse L., Bottini D., Le Brun V., Maccagni D., Picat J. P., Scaramella R., Scodeggio M., Zanichelli A., others 2005, A&A, 439, 845

Lee K., Giavalisco M., Gnedin O. Y., Somerville R. S., Ferguson H. C., Dickinson M., Ouchi M., 2006, ApJ, 642, 63

Lejeune T., Cuisinier F., Buser R., 1997, A&AS, 125, 229

Lemson G., Virgo Consortium t., 2006, ArXiv Astrophysics e-prints

Lilly S. J., Le Fèvre O., Renzini A., Zamorani G., Scodeggio M., Contini T., Carollo C. M., Hasinger G., Kneib J., Iovino A., Le Brun V., Maier C., Mainieri V., others 2007, ApJS, 172, 70

Lilly S. J., Longair M. S., 1984, MNRAS, 211, 833

Limber D. N., 1954, ApJ, 119, 655

Limber D. N., 1957, ApJ, 125, 9

Lintott C. J., Schawinski K., Slosar A., Land K., Bamford S., Thomas D., Raddick M. J., Nichol R. C., Szalay A., Andreescu D., Murray P., Vandenberg J., 2008, MNRAS, 389, 1179

Lonsdale C. J. e. a., 2003, PASP, 115, 897

Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163

Madau P., 1995, ApJ, 441, 18

Madgwick D. S., Hawkins E., Lahav O., Maddox S., Norberg P., Peacock J. A., Baldry I. K., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Cole S., others 2003, MNRAS, 344, 847

Magliocchetti M., Cirasuolo M., McLure R. J., Dunlop J. S., Almaini O., Foucaud S., de Zotti G., Simpson C., Sekiguchi K., 2008, MNRAS, 383, 1131

Magliocchetti M., Maddox S. J., 1999, MNRAS, 306, 988

Magliocchetti M., Porciani C., 2003, MNRAS, 346, 186

Magliocchetti M., Silva L., Lapi A., de Zotti G., Granato G. L., Fadda D., Danese L., 2007, MNRAS, 375, 1121

Magorrian J., Tremaine S., Richstone D., Bender R., Bower G., Dressler A., Faber S. M., Gebhardt K., Green R., Grillmair C., Kormendy J., Lauer T., 1998, AJ, 115, 2285

Mancini C., Daddi E., Renzini A., Salmi F., McCracken H. J., Cimatti A., Onodera M., Salvato M., Koekemoer A. M., Aussel H., Floc'h E. L., Willott C., Capak P., 2010, MNRAS, 401, 933

McCarthy P. J., Le Borgne D., Crampton D., Chen H., Abraham R. G., Glazebrook K., Savaglio S., Carlberg R. G., Marzke R. O., Roth K., Jørgensen I., Hook I., Murowinski R., Juneau S., 2004, ApJ, 614, L9

McCarthy P. J., Persson S. E., West S. C., 1992, ApJ, 386, 52

McCracken H. J., Capak P., Salvato M., Aussel H., Thompson D., Daddi E., Sanders D. B., Kneib J., Willott C. J., Mancini C., Renzini A., Cook R., others 2010, ApJ, 708, 202

McCracken H. J., Ilbert O., Mellier Y., Bertin E., Guzzo L., Arnouts S., Le Fèvre O., Zamorani G., 2008, A&A, 479, 321

McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, MNRAS, 395, 2196

Meisenheimer K., Wolf C., 2002, Astronomy and Geophysics, 43, 030000

Meneux B., Guzzo L., de La Torre S., Porciani C., Zamorani G., Abbas U., Bolzonella M., Garilli B., Iovino A., Pozzetti L., Zucca E., Lilly S. J., Le Fèvre O., Kneib J., Carollo C. M., others 2009, A&A, 505, 463

Meneux B., Le Fèvre O., Guzzo L., Pollo A., Cappi A., Ilbert O., Iovino A., Marinoni C., McCracken H. J., Bottini D., Garilli B., Le Brun V., Maccagni D., Picat J. P., Scaramella R., Scodeggio M., Tresse L., others 2006, A&A, 452, 387

Mewe R., Kaastra J. S., Liedahl D. A., 1995, Legacy

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

Mo H. J., White S. D. M., 2002, MNRAS, 336, 112

Nesvadba N. P. H., Lehnert M. D., De Breuck C., Gilbert A. M., van Breugel W., 2008, A&A, 491, 407

Neyman J., Scott E. L., 1952, ApJ, 116, 144

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19

Norberg P., Baugh C. M., Hawkins E., Maddox S., Madgwick D., Lahav O., Cole S., Frenk C. S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., others 2002, MNRAS, 332, 827

Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020

Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peebles P. J. E., 1980, The large-scale structure of the universe

Peebles P. J. E., 1982, ApJ, 263, L1

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perlmutter S., Aldering G., Goldhaber G., Knop R. A., Nugent P., Castro P. G., Deustua S., The Supernova Cosmology Project 1999, ApJ, 517, 565

Pickles A. J., 1998, PASP, 110, 863

Plummer H. C., 1911, MNRAS, 71, 460

Porciani C., Magliocchetti M., Norberg P., 2004, MNRAS, 355, 1010

Postman M., Geller M. J., 1984, ApJ, 281, 95

Pozzetti L., Mannucci F., 2000, MNRAS, 317, L17

Press W. H., Schechter P., 1974, ApJ, 187, 425

Quadri R. e. a., 2007, AJ, 134, 1103

Raymond J. C., Smith B. W., 1977, ApJS, 35, 419

Reeves J. N., O'Brien P. T., Ward M. J., 2003, ApJ, 593, L65

Reiss D. J., Germany L. M., Schmidt B. P., Stubbs C. W., 1998, AJ, 115, 26

Roche N. D., Almaini O., Dunlop J., Ivison R. J., Willott C. J., 2002, MNRAS, 337, 1282

Romer A. K., Collins C. A., Böhringer H., Cruddace R. G., Ebeling H., MacGillivray H. T., Voges W., 1994, Nature, 372, 75

Salpeter E. E., 1955, ApJ, 121, 161

Schechter P., 1976, ApJ, 203, 297

Schmidt M., 1968, ApJ, 151, 393

Sekiguchi K., Akiyama M., Furusawa H., Simpson C., Takata T., Ueda Y., Watson M. W., The Sxds Team 2005, in A. Renzini & R. Bender ed., Multiwavelength Mapping of Galaxy Formation and Evolution pp 82–+

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Sheth R. K., Tormen G., 2002, MNRAS, 329, 61

Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385

Silk J., Rees M. J., 1998, A&A, 331, L1

Simon P., 2007, A&A, 473, 711

Skibba R. A., 2009, MNRAS, 392, 1467

Skibba R. A., Bamford S. P., Nichol R. C., Lintott C. J., Andreescu D., Edmondson E. M., Murray P., Raddick M. J., Schawinski K., Slosar A., Szalay A. S., Thomas D., Vandenberg J., 2009, MNRAS, 399, 966

Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080

Smail I., Ivison R. J., Blain A. W., 1997, ApJ, 490, L5+

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311

Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, ApJ, 600, L171

Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504

Spergel D. N., Bean R., Doré O., Nolta M. R., Bennett C. L., Dunkley J., Hinshaw G., Jarosik N., Komatsu E., Page L., Peiris H. V., Verde L., Halpern M., Hill R. S., Kogut A., Limon M., others 2007, ApJS, 170, 377

Spergel D. N., Verde L., Peiris H. V., Komatsu E., Nolta M. R., Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S. S., Page L., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2003, ApJS, 148, 175

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629

Stanford S. A., Eisenhardt P. R., Dickinson M., 1998, ApJ, 492, 461

Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2003, ApJ, 592, 728

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17+

Stott J. P., Smail I., Edge A. C., Ebeling H., Smith G. P., Kneib J., Pimbblet K. A., 2007, ApJ, 661, 95

Stoyan D., 1984, Math. Nachr.

Stoyan D., Stoyan H., 1994, Fractals, Random Shapes and Point Fields

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Swinbank A. M., Chapman S. C., Smail I., Lindner C., Borys C., Blain A. W., Ivison R. J., Lewis G. F., 2006, MNRAS, 371, 465

Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, ApJ, 621, 673

Toft S., van Dokkum P., Franx M., Labbe I., Förster Schreiber N. M., Wuyts S., Webb T., Rudnick G., Zirm A., Kriek M., van der Werf P., Blakeslee J. P., Illingworth G., Rix H., Papovich C., Moorwood A., 2007, ApJ, 671, 285

Toomre A., Toomre J., 1977, Violent Tides Between Galaxies. pp 271–+

Trager S. C., Faber S. M., Worthey G., González J. J., 2000, AJ, 119, 1645

Tremonti C. A., Heckman T. M., Kauffmann G., Brinchmann J., Charlot S., White S. D. M., Seibert M., Peng E. W., Schlegel D. J., Uomoto A., Fukugita M., Brinkmann J., 2004, ApJ, 613, 898

Tremonti C. A., Moustakas J., Diamond-Stanic A. M., 2007, ApJ, 663, L77

Tully R. B., Fisher J. R., 1977, A&A, 54, 661

Valls-Gabaud D., Alimi J., Blanchard A., 1989, Nature, 341, 215

van Dokkum P. G. e. a., 2006, ApJ, 638, L59

Warren S. J., Hambly N. C., Dye S., Almaini O., Cross N. J. G., Edge A. C., Foucaud S., Hewett P. C., Hodgkin S. T., Irwin M. J., Jameson R. F., et al. 2007, MNRAS, 375, 213

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Navarro J. F., Evrard A. E., Frenk C. S., 1993, Nature, 366, 429

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879

Wolf C., Gray M. E., Meisenheimer K., 2005, A&A, 443, 435

York D. G. e. a., 2000, AJ, 120, 1579

Zehavi I. e. a., 2005, ApJ, 630, 1

Zheng Z., Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Davé R., Frenk C. S., Katz N., Lacey C. G., 2005, ApJ, 633, 791

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

Zirm A. W., van der Wel A., Franx M., Labbé I., Trujillo I., van Dokkum P., Toft S., Daddi E., Rudnick G., Rix H., Röttgering H. J. A., van der Werf P., 2007, ApJ, 656, 66

Zucca E., Ilbert O., Bardelli S., Tresse L., Zamorani G., Arnouts S., Pozzetti L., Bolzonella M., McCracken H. J., Bottini D., Garilli B., Le Brun V., Le Fèvre O., Maccagni D., Picat J. P., Scaramella R., others 2006, A&A, 455, 879