On the Distribution of Mass and Light in the Universe

Henry Joy McCracken

July 26, 2011

Contents

Chapter 1

Introduction

1.1 A survey of surveys: measuring the distribution of galaxies

Take a cursory glance at the night sky on a dark moonless night far from any city and one startling fact is immediately apparent: the stars in the sky are not scattered at random across the heavens but instead are *clustered*. For a long time the significance of this structure was lost on human minds, until it was realised that this beautiful veil stretching over half the sky was in fact the disk of our own galaxy.

In the northern hemisphere, only one object beyond our galaxy is visible to the naked eye, the great spiral galaxy of Andromeda. But there are as many galaxies in the visible universe as there are stars in our own Milky Way. Once the confusion over which other fuzzy blobs in the night sky were in our galaxy and which lay beyond it was resolved early in the 20th century, astronomers began in earnest to produce galaxy catalogues and investigate their properties. Could the distribution of galaxies on the sky tell us about the Universe in the same way that stars tell us about our own galaxy?

In a the early 1930s, Edwin Hubble made one of the first maps of the galaxy distribution on the sky, the first measurement of how many galaxies there visible in each interval of apparent brightness and the first determination of the constant of proportionality between the rate of recession of an object and its distance, so laying the foundations for a new discipline, that of *observational cosmology*. Hubble demonstrated that the number of galaxies on each photographic plate follows a *lognormal* distribution. He took took care to ensure that the sensitivity variation between each different photographic plates were correctly accounted for, and his sample was considerably more homogeneous than those of his competitors. Although it was not realised at the time, this lognormal distribution discovered by Hubble was amongst the first evidence that at intermediate scales, the distribution of galaxies was not uniform. Clusters of galaxies were well known, but were the galaxy clusters themselves clustered? Could the structures observed be due obscuration in our own galaxy?

Answering these question satisfactorily would take at least another half-decade. Although photographic plates made it easy to survey large areas of

the sky, making qualitative analyses of this plate data without computers was extremely laborious and time-consuming. Earlier efforts to catalogue the sky using plates (such as the "Carte du Ciel" project started in Paris at the end of the 19th century) had foundered in part due to an underestimate of just how much work was really involved. However, in the 1950s, using the wide field Carnegie "Astrograph" telescope, capable of simultaneously taking images through two separate filters, Shane and Wirtanen made a detailed map of the observed galaxy distribution on large scales (Shane & Wirtanen, 1954). Over the space of ten years, they counted the number of galaxies in cells on more than a thousand photographic plates. They were able to derive contour plots of the galaxy distribution and demonstrate conclusively that could not be explained by a Poisson distribution.

In parallel, statistical tools were developed to characterise the galaxy distribution. One of the most popular and enduring metrics has been the two-point galaxy correlation function w, which describes the excess number of pairs as a function of angular separation compared to a random distribution. This had already been applied to smaller object catalogues such as galaxy clusters. At the end of the 1960's, it was shown that the amplitude of w as a function of angular separation in the Shane and Wirtanen catalogue could be well approximated by a power-law (Totsuji & Kihara, 1969) with a characteristic scale r_0.

By the 1970s, advances in computer technology made it possible to reanalyse the Lick galaxy catalogue (Shane & Wirtanen, 1967) at the original resolution of $10' \times 10'$ (Groth & Peebles, 1977). For the first time, a measurement of the "Scaling relation" for the observed clustering of galaxies was made: progressively fainter galaxy slices have lower amplitudes at a fixed angular separation. This removed any lingering doubts that the observed clustering signal could be caused by galactic obscuration and was in fact a signal of cosmological origin. Using a mathematical relation discovered in the 1950's (what is known today as the "Limber formula" (Limber, 1953)) one could calculate the characteristic scale of clustering at the depth of the survey. In the 1950's, measuring correlation functions on large galaxy catalogues such as Shane and Wirtanen's was impossible due to lack of computing power. But by the 1990's, automatic plate scanners combined with object detection software made it possible to create million-galaxy catalogues covering a significant fraction of the sky. This in turn led to for the first time a precise determinations of the angular dependence of w out to large scales(Maddox et al., 1990).

However, surveys such as these only tell us about the galaxy distribution projected along the line of sight. Galaxies are most certainly not standard candles, and inferring an object's distance from it's apparent luminosity is not possible, or at least not advisable. A true map of the Universe needs distance information. What was really required was a *redshift survey*. Unfortunately, measurement of galaxy redshifts requires a prohibitive amount of telescope time. Nevertheless, at the end of the 1970's and the beginning of the 1980's, the first systematic redshift surveys with well-defined selection limits were started, and galaxy distances were measured one by one, with surveys comprising a few hundred objects taking years to accumulate. One such ground-breaking survey, the Center for Astrophysics Redshift Survey, aimed to measure radial velocities

(and hence redshifts) for all galaxies brighter than 14.5 magnitudes at high galactic latitude. This survey (Huchra et al., 1983) provided the first three-dimensional picture of what the local Universe looked like, revealing a bubbles, sheets and voids (de Lapparent et al., 1986). Subsequent surveys such as the LCRS revealed in greater detail the sponge-like structure of the local Universe (Shectman et al., 1996). These works also provided compelling evidence that the Universe is homogenous on scales larger than $100h^{-}1$ Mpc.

At the same time, "pencil-beam" redshift surveys were beginning to provide much deeper slices of the Universe although the small transverse size of these surveys made early results difficult to interpret (Broadhurst et al., 1990). Lilly et al. (1995) in the landmark Canada France Redshift survey measured redshifts for galaxies selected galaxies systematically in a magnitude-limited sample and made the first measurements on how the clustering pattern has evolved from the present day to high redshift (Fevre et al., 1996). At the same time, deep photometric surveys such as the Hubble Deep Field Williams et al. (1996) project and the William Herschel deep field (Metcalfe et al., 2001) provided very deep multicolour observations of the distant universe well beyond the spectrosopic limit.

In the last ten years, the Sloan digital sky survey (SDSS) and the two-degree-field surveys (2df) have provided redshift catalogues an order of magnitude larger. The 2dF survey measured the distortion of the clustering pattern of galaxies due to infall (Peacock et al., 2001), a key signature of the formation of structures by gravitational instability. The now-complete 2dF and Sloan surveys have provided extremely detailed information on how galaxies are distributed in the local Universe and how this clustering depends on galaxy type and environment.

But what is the origin exactly of the distribution of galaxies seen in these surveys, and how does it relate to the underlying cosmological model? What can galaxy clustering tell us about the evolution of galaxies and the composition of the Universe?

1.2 The development of our cosmological model

With the advent of the theory of General Relativity, it became possible for the first time to imagine a mathematical theory which could describe the formation and evolution of the Universe. Starting from the cosmological principle, namely that the Universe is homogeneous and isotropic on large scales, one can derive a set of equations which describe the evolution of matter and radiation over the age of the Universe. The most natural solution for the field equations in the case of a homogeneous and isotropic Universe is that the Universe is expanding. However, Einstein was uncomfortable with this result and decided to add an additional term, the cosmological constant, in order to ensure that the Universe remained static. It is worth remembering that at the time that General Relativity was formulated, astronomers were uncertain even which of the nebulae were part of our own galaxy, and which lay beyond it. A static Universe seemed to be the most reasonable solution.

Meanwhile Vesto Slipher, at Flagstaff Observatory in the deserts of Arizona, was making the first spectroscopic observations of "nebulae" – the generic term people used at the time to describe unresolved objects. In 1912 he measured the radial velocity of Andromeda, and found that it was *blueshifted*: it was moving towards our galaxy. By 1917 he had measured many further nebulae and found all the other objects had positive recessional velocities. Edwin Hubble combined these measurements with his spectroscopic measurements with Cepheid variable distances to show that there was a positive correlation between distance and recessional velocity. The more distant objects were receding more rapidly from us. The obvious interpretation of which (although one that Hubble did not immediately draw himself) was that the Universe was expanding.

If the Universe was expanding, then there must have been a time in the past when it was much hotter and denser than today. If this was the case, then the Universe should be bathed in radiation from this ancient explosion, as suggested by George Gamov in the 1940s. Could this be detectable? At the time, microwave receivers were gradually increasing in sensitivity. Two radio astronomers, Penzias and Wilson, attempting to map the galaxy were irritated by a uniform signal coming from every direction in the sky. Talking to astronomers from Princeton, Dicke and Peebles, they realised that this signal was in fact the relic radiation from the beginning of the Universe, redshifted to microwave wavelengths. This was not actually the first time this signal was observed: as an interesting historical aside, more than twenty years previously, Canadian astronomer Andrew McKellar, examining the transitions of diatomic molecules, had determined the temperature of the background as 2.3K. But the arrival of war in Europe and McKellar's publication in a low-visibility journal meant that his results did not get the attention that the deserved, and it was two decades before the significance of this measurement was fully appreciated. Later experiments, including most spectacularly the Cosmic Microwave Background explorer (COBE) satellite provided a very precise measurement of the spectrum of the relic radiation and showed that, as predicted, it matched very closely that of a black body.

How did these structures grow? On very large scales, gravitation is the most dominant force: unlike other forces, it is never repulsive. So a picture in which structures are "seeded" by small initial perturbations in an otherwise homogeneous and isotropic Universe seemed like a promising approach. Perturbations will grow providing they are in structures more massive than a characteristic mass (known as the "Jeans mass"). If there were small inhomogeneities in Gamov's cloud they could grow in time under the action of gravity. Such power spectrum for these perturbations (which describes the amplitude of perturbations depends on angular scale) was suggested by the Russian mathematician Zel'dovich, who thought that the largest structures should form first. But there were still doubts concerning the exact composition of the Universe.

In the 1930s, Swiss astromer Fritz Zwicky measured radial velocities of galaxies in the Coma cluster. Using the "virial theorem", which provides a way to estimate the mass of a relaxed system based on velocity measurements, he derived the dynamical mass of the cluster. He found, surprisingly, that this derived mass was several hundred times greater than one would expect if all

the mass was locked up in visible galaxies. Zwicky suggested that some unseen "dark matter" might account for the missing mass.

Paradoxically, this result was forgotten for most of the next three decades. In the mean-time, Dutch astronomers were making increasingly more precise measurements of galaxy rotation curves using 21 centimetre measurements of neutral hydrogen in the outskirts of galaxies. On the other side of the Atlantic, astronomers were also measuring optical rotation curves for the first time. However, it was computer simulations of disk formation which suggested that galaxies might in fact be embedded inside massive dark haloes. Ostriker & Peebles (1973) using numerical simulations found that disks would only remain stable if they were embedded inside large haloes. Since no such haloes were observed around nearby galaxies then these haloes must be dark "dark".

At the time, most astronomers thought that the "dark matter" required to form disks, flatten the rotation curves and explain the high velocities inside clusters was probably ordinary matter such as dead stars. However, the observed amount deuterium, created during the big bang, set quite severe limits on the total amount of baryonic material in the Universe – and it was quite uncomfortably small. It seemed that whatever the dark matter was, it was not a material similar to ordinary matter.

By the 1980's computers were powerful enough to carry out simulations with sufficiently large numbers of particles and with sufficient mass resolution to follow for the first time the evolution of structures in the Universe. Davis et al. (1985) carried out a series of numerical simulations which traced in detail the evolution of structures in the Universe, and they compared their results to the distribution of galaxies observed in the first deep galaxy redshift surveys. They were able to show that if dark matter consisted of relativistic particles such as neutrinos it was not capable of producing enough clustering on large scales compared to observations. The problem is essentially that neutrinos decouple too soon from the radiation field. They also concluded that if the matter density of the was sufficiently high (and the Universe was "closed", then galaxy formation had to be *biased*: in other words, galaxies could only form at the highest peaks in the dark matter distribution.

In the early 1990s, the COBE satellite provided a measurement of the spectrum of the microwave background: it matched almost perfectly that of a black body at a temperature of 2.73 K. After carefully subtracting background components from the microwave map of the sky, astronomers detected small residual perturbations of the microwave map, the seeds which would lead into galaxies, and thus providing an essential verification of the model proposed by Zel'dovich and others twenty years previously.

By the mid 1990's, the cold dark matter model of structure formation had become established and was the only model which could broadly reproduce observations of the growth of structures. However, the exact values of the cosmological parameters were not known to high precision. Although the inflationary model preferred a flat model with $\Omega = 1$ most observations at the time were consistent with either flat or "open" models. The open models preferred a value of the Hubble constant of $\sim 70\mathrm{km}^{-1}/\mathrm{s}$, which removed the potential conflict with the measured ages of globular clusters which would be older than age of

the Universe in these models. Finally, in the late 1990s, cosmology underwent another radical change. Measurements of distant supernovae indicated that the expansion of the Universe was accelerating. The origin of this accelerating Universe is uncertain, and one of the key tasks of cosmology in the coming years will be to understand the origin of nature of this acceleration.

1.3 The growth of structure and the formation of galaxies

In the last decade, data from higher-precision microwave background experiments combined with other cosmological probes such as baryon acoustic oscillations, supernovae or gravitational lensing have determined to very high precision our current cosmological model: a flat, dark energy-dominated Universe in which dark matter provides around $\sim 30\%$ of the Universes' energy content. It is against the background of this model that the work described here was carried out.

In our current paradigm of galaxy formation, structures consisting of haloes and filaments of dark matter grow from tiny temperature excesses in the early Universe. In turn, stars and galaxies form through the cooling and condensation of gas at the centre of these virialised haloes of dark matter (White & Rees, 1978). These luminous objects in haloes are situated at the peaks of the underlying dark matter distribution. The galaxy formation process is thus driven by a combination of the gravitational interaction between dark matter and ordinary matter and the action of gas physics, all of which takes place against the background of an expanding Universe with a non-zero cosmological constant. In this acceleration-dominated cold dark matter model, ("lambda CDM") structures grow "hierarchically", small haloes forming first and merging to form larger ones. At the same time, as haloes merge, galaxies merge.

It is worth returning to our "survey of surveys" and reviewing the current state-of the art. In the local Universe, the Sloan York et al. (2000) and two-degree-field Colless et al. (2001) surveys have now provided a detailed three-dimensional picture of the distribution of galaxies. Figure 1.1 shows the distribution of galaxies mapped by the seventh data release of the Sloan digital sky survey ("DR7"); a rich pattern of voids and filaments is evident. The DR7 release provides spectroscopic redshifts for more than million galaxies and photometric redshifts for more than ten times as many. These very large samples have enabled extensive investigations of galaxy properties as a function of type and luminosity (Norberg et al., 2002) and have provided broad support for the " biased" picture of galaxy formation. In this scenario luminous galaxies form at the peaks of the underlying density field; more luminous galaxies live in the centers of more luminous dark matter haloes. It is relatively straightforward to explain the present-day large scale clustering amplitudes starting from the microwave background power spectrum and this biased galaxy formation picture.

However, the SDSS-main sample only extends to $z \sim 0.2$; this narrow red-

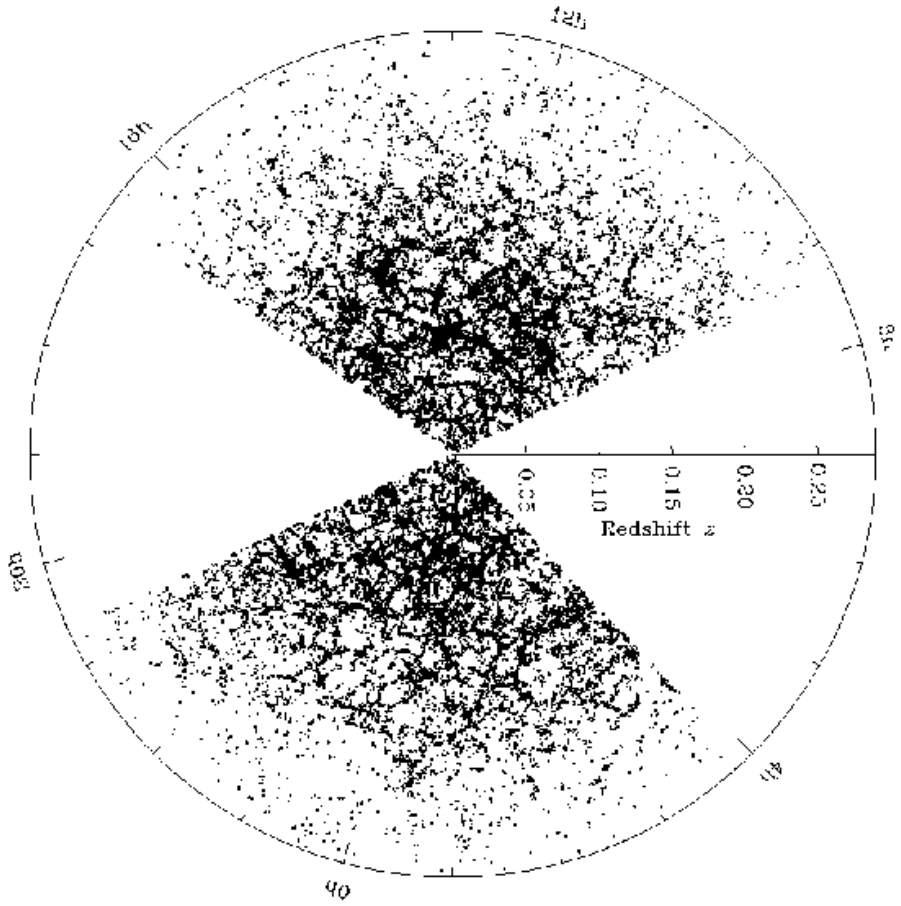

Figure 1.1 The large scale distribution of galaxies in the local universe, as probed Sloan Digital Sky Survey 7th data release (Zehavi et al., 2010).

shift baseline means that it is quite difficult to detect evolutionary changes in the galaxy population within the survey (although certain tracer populations such as luminous red galaxies can be observed to higher redshifts, it is hard to follow the evolution of normal galaxies like our Milky Way out to cosmologically significant distances). Since the mid 1990s, “Pencil beam“ spectroscopic surveys using 4-8m class telescopes and deep imaging surveys have revealed a

Universe at higher redshifts quite different from the one found in the SDSS and 2DF. What is the relationship between what has been observed in these surveys and the local Universe observed by the SDSS?

The rate at which stars form in the Universe increases rapidly as one probes further back into the history of the Universe and reaches a peak at around $z \sim 1-2$; the same effect is seen in the evolution of the "mass function", or how much stellar mass exists per cubic megaparsec at each mass interval. Most baryons already seem to be locked up in stars by redshift $z \sim 1-2$. Furthermore, the number density of passively evolved massive galaxies at intermediate redshifts is much larger than might naively expected from models in which smaller haloes merge to form larger ones. A related question is the "bimodality" seen in galaxy properties in the local Universe, where we see a clear division between red, evolved galaxies and blue, star-forming populations. What is the origin of this bimodal distribution? How did star-forming galaxies become old, evolved and passive? Although the basic White & Rees model of hierarchical merging of dark matter haloes is probably correct, it seems clear now additional mechanisms are required to regulate star-formation in massive and less-massive haloes and suppress the formation of low mass galaxies. The key open question underlying all this is the how the relationship between mass and light evolves from $z \sim 3$ to the present day.

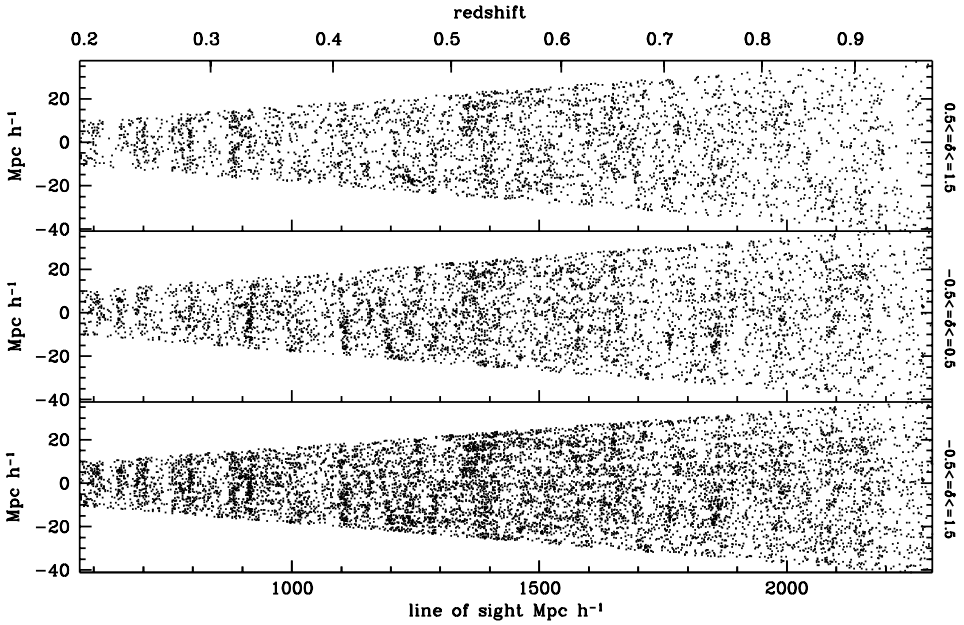

Figure 1.2 The structure of the Universe at intermediate redshifts, as seen by the VIMOS-VLT wide survey. From Garilli et al. (2008).

1.4 An outline of this work

Developments in instrumentation have now made it possible to construct catalogues capable of at least partially addressing some of these questions. The key is to assemble samples of galaxies large enough to make divisions by type, absolute magnitude, redshift and stellar mass. Furthermore, one must have a sufficiently large wavelength baseline to permit coverage of a large redshift range. Control over systematic errors is essential: whilst early pencil-beam surveys were limited by shot noise effects and sample variance, in today's surveys control over systematic errors become essential.

My work over the past ten years has concentrated on attempts to *measure* and to *understand* the evolution of galaxy clustering and galaxy evolution over a significant fraction of the lifetime of the Universe. The *measurement* component of my work has involved the construction of large catalogues of galaxies based mostly on photometric surveys. An important aspect of this has been the development and exploitation of photometric redshift techniques which have enabled dramatic increase in the amount of useful information which can be extracted from photometric surveys given a suitable spectroscopic training set. (In addition to the papers cited here, it is worth mentioning that these catalogues I have developed have been used in hundreds of published articles). To *understand* the observed clustering pattern of galaxies requires comparisons with models, and in Chapter 4 I will present a description of a phenomenological model of galaxy clustering, the "halo model"

To summarize a selection of major results:

- Galaxy clustering at intermediate redshifts. Using the Canada-France Deep Fields survey (CFDF), one of the first wide field ($1\,\mathrm{deg}^2$) deep multi-colour surveys carried out with CCD detectors, the projected clustering of galaxies and its variation with apparent colour was investigated. This survey provided precise measurements of projected field galaxy clustering at $z \sim 1$ (McCracken et al., 2001). With Sebastien Foucaud, we applied a simple colour cut to the CFDF catalogues to select lyman break galaxies at $z \sim 3$, showing for the first time evidence that the clustering amplitude of Lyman break galaxies depends on absolute luminosity, an important test of the biased model of galaxy formation. Part of this data set was also used to make the first detection of correlated ellipticities or cosmic shear (Van Waerbeke et al., 2000)).

- Galaxy clustering in the CFHTLS-deep (McCracken et al., 2008). The four CFHTLS deep fields provided one of the first opportunities to carry out a detailed study of the dependence of the clustering properties of the field galaxy population as a function of redshift, type and luminosity from $z \sim 0$ to $z \sim 1$.

- The universe at intermediate redshifts and near-infrared selected surveys. In an early paper (McCracken et al., 2000) not described in this thesis, I constructed a series of near-infrared galaxy samples in the Herschel deep

field and used them to investigate the evolution of galaxies at intermediate redshifts: in 2010, I presented a new, very-wide field near-infrared catalogue covering the entire COSMOS field (McCracken et al., 2010). Near-infrared data provides two important advantages: accurate measurements of stellar masses, and the possibility to measure photometric redshifts in the all-important $1 < z < 2$ redshift range.

- In collaboration with J. Coupon and M. Kilbinger, we developed an analytic model of clustering of galaxies, the "halo model" and applied it to the $\sim 10^6$ galaxies in the CFHTLS-wide (Coupon et al., 2011). This groundbreaking work provides the most accurate measurements to date of the dependence of galaxy clustering properties on type, luminosity and redshift. By analysing our measurements in the framework of the halo model, we were are able to derive the halo mass where star-formation is most efficient and its dependence on redshift.

- Developement, testing and exploitation of photometric redshifts (Ilbert et al., 2006; Coupon et al., 2009). Photometric redshifts have become a valuable technique to extract additional information from purely photometric surveys. Starting from precise photometric catalogues extracted from the Canada-France legacy survey, I have participated in the test and development of photometric redshift techniques which have been applied to large ground-based catalogues.

In this thesis I will focus on three topics which will allow me to bring together the main themes of my research: galaxy clustering measurements made in the CFHTLS deep fields; our measurements of the passive galaxy population in the Universe at intermediate redshifts in the COSMOS survey; and the evolution galaxies inside haloes in COSMOS and CFHTLS in the framework of the halo model. The second and third chapters are a concise summary of (McCracken et al., 2008) and (McCracken et al., 2010); chapter 4 represents a synthesis of all the techniques developed during my research and also suggests a signpost for the future directions I intend my work to take. It is based in large part in the results presented in (Coupon et al., 2011) and a second paper currently in preparation.

Chapter 2

Galaxy clustering with photometric redshifts

2.1 Making precise photometric redshift measurements

In the beginning, galaxy clustering measurements were made on samples selected by apparent magnitude. With the availability of multi-colour data it became possible to search for objects in broad redshift ranges by looking for "breaks" in the spectral features. For example, in star-forming galaxies no photons are emitted bluewards of 912Å: this effect combined with the opacity of the Universe at ultraviolet wavelengths makes this feature a useful signature of high redshift galaxies. In fact, this technique had been tried in the early 1990s to locate high redshift galaxies (Guhathakurta et al., 1990) but without the possibility of spectroscopic confirmation on 8-m telescopes the results were inconclusive. It was only with the advent of the 8m Keck telescopes could specially selected sub-samples be spectroscopically verifered permitting in the 1990s large numbers of galaxies at distant redshifts to be rapidly accumulated (Steidel et al., 1996). Conversely, a similar approach can be applied to early-type galaxies by searching for the 4000Å break (Adelberger et al., 2005).

Spectral breaks can be approximately located using only three bands. But if more filters are available, one can actually estimate an approximate redshift for the galaxy, using the "photometric redshift" technique. Although there are many variations, the most common method used is to compare predicted fluxes for a set of galaxy templates at a range of redshifts with the observed fluxes; an early example of this technique can be found in Baum (1962), in which the authors find a approximate redshift for a nearby cluster using this method. In addition, finding the best-fitting template will give provide an estimate of the galaxy's spectral type, absolute magnitude and stellar mass.

Of course, there are a large number of details which make the practical computation of photometric redshifts much more difficult than this outline would suggest. The range over which photometric redshifts can be determined depends strongly on the filter set used: with only optical bands, it is hard to measure redshifts in the critical (and interesting) $1 < z < 2$ redshift range,

where features move out of visible wavebands. The Lyman-break and Balmer-break features, the most prominent characteristics in galaxy spectra, can easily be confused if observations are not sufficiently deep in ultraviolet and near-infrared bands, resulting in what is known as "catastrophic" failure where a galaxy at low redshift is erroneously assigned a high-redshift solution.

A more insidious and difficult-to-control problem is that of photometric calibration problems and redshift-dependent systematic errors. Until the mid-2000's, photometric redshifts had in general only been used with very deep, extremely high quality space-based data like the Hubble deep field (Williams et al., 1996), which had excellent, uniform photometric calibration.

With the advent of MegaCam (Boulade et al., 2000) and other wide-field cameras it became relatively easy to collect photometry over a broad spectral range (MegaCam, combined with the MegaPrime top-end is one of the few detector–telescope combinations sensitive in u to z bands) covering an entire degree field-of-view with a well-sampled point-spread function. Combined with telescope "queue-scheduing", in which observations are only taken under a well-defined set of atmospheric conditions, this makes possible collecting extremely homogenous and well-calibrated photometric data. However, even with data of this quality, calibration using a spectroscopic training set remains essential. Without spectroscopic redshifts, it is impossible to assess the effect of systematic errors which may depend on magnitude, spectral type and redshift.

In 2006 we (myself, Olivier Ilbert and Stephane Arnouts) created a very large "training set" for photometric redshift computations by combining ultra-deep *ugriz* photometric measurements from the CFHTLS deep fields with $\sim$ 3,000 spectroscopic redshifts measured by the VIRMOS VLT deep survey team (Fèvre et al., 2005b). Although galaxies in this survey had initially been selected using deep imaging taken using the CFH12k camera (McCracken et al., 2003) this field was also covered by the "D1" field of the Canada France Legacy Survey. I had assisted in processing this data as part of my *Tache de Service* at the TERAPIX[1] data centre. This represented the largest spectroscopic training set ever used until then to compute photometric redshifts at intermediate redshifts. The combination of high-quality photometric measurements from the Canada-France legacy survey with a very large spectroscopic training set was completely unique.

Our results were interesting. Previously, several groups had already computed photometric redshifts for large ground-based galaxy samples, but without sufficiently large training sets – typically, only a few hundred spectroscopic redshifts were available – and this made it difficult to assess the robustness of the resulting photometric redshifts. We found that our "blind" photometric redshifts, computed without using any information supplied by the spectroscopic training set, showed redshift-dependent offsets as illustrated by Figure 3 from Ilbert et al. (2006). Without our large training set, these systematic errors would not have been detected.

The origin of these systematic offsets is a combination of many factors. In Ilbert et al. (2006) we found that by comparing photometric fluxes measured

[1] http://terapix.iap.fr

for galaxies with known spectroscopic redshifts with predicted fluxes based the galaxy spectra we could "correct" our best fitting templates by determining a fixed offset. These corrected templates provided much better photometric redshift estimates, even for galaxies which had not been used in calculating the template flux offsets. We speculated that a combination of incorrect knowledge of the telescope response function and perhaps residual aperture offsets in the galaxy photometry were the origin of these systematic offsets. The large training set also allowed an investigation of how photometric redshifts errors depended on galaxy type and redshift. Not surprisingly, earlier galaxy types, showing a strong break, had more accurate photometric redshifts; conversely, bluer galaxy types had lower photometric redshift accuracy. In particular, determining photometric redshifts can be very challenging for starburst galaxies, which may or may not have an emission line feature. In addition, differences in point-spread-functions between different bands can produce a redshift-dependent offset. In the CFHTLS, queue-scheduled observations help to minimise this effect in the deep fields, although it remains problematic in the Wide (which we expect to address in future releases of the CFHTLS wide).

Another important aspect of this work was determining the redshift ranges and realistic redshift accuracies which could be attainable using well-calibrated photometric data. Once again, various works had attempted to estimate this using simulations which turned out (in general) to completely neglect systematic photometric errors. In general, these simulations relied simply taking a set of known templates (sometimes even employing the same templates used in the photometric redshift determination!) adding noise and distributing galaxies at random redshifts. Such a procedure generally neglects the effects of insidious systematic errors discovered in Ilbert et al..

The photometric redshifts described in this paper were distributed publicly and have been now been cited in almost 200 works. In a subsequent paper, Jean Coupon (a doctoral student at the IAP) under the direction of myself and Olivier Ilbert, computed photometric redshifts for the Wide part of the CFHTLS survey; these were also subsequently released to the public Coupon et al. (2009).

Once we had demonstrated that we understood the limitations of our photometric redshift measurements and the redshift ranges over which they could be utilised, they were ready to be exploited scientifically. In the following section, I will present a resume of the results presented in McCracken et al. (2008).

2.2 Clustering properties of a volume-limited galaxy sample

2.2.1 Sample selection

The key question we would like to address is how the distribution of galaxies depends on their intrinsic properties and how this distribution changes with look-back time. As we have seen, surveys like the SDSS have provided a detailed picture of how galaxies are distributed in the local Universe and how

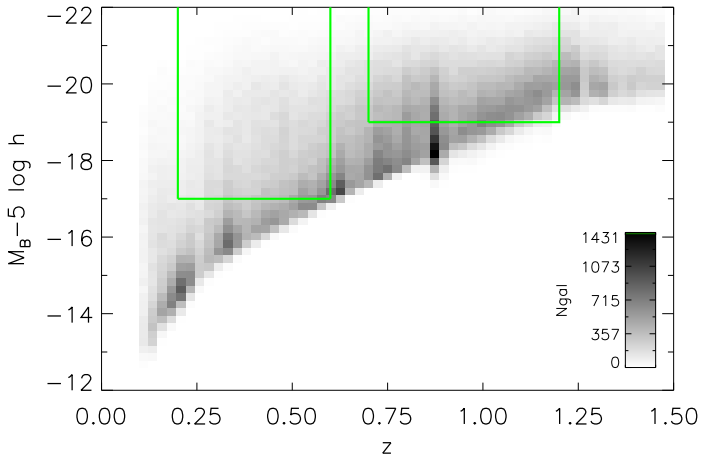

Figure 2.1 Gray-scaled histogram showing the distribution of galaxies as a function of absolute magnitude and redshift for four CFHTLS fields for all galaxy types and for an apparent magnitude limit of $i^* < 24.5$.

clustering properties depend on spectral type and intrinsic luminosity. Ideally, we would like to make similar measurements at $z \sim 1$. However, even with a high-multiplex instrument like VIMOS, accumulating enough spectroscopic redshifts to carry out studies of galaxy clustering as a function of redshift is time-consuming. For this reason, photometric redshifts calibrated with a sufficiently large training set (so that systematic effects are well understood) is a promising technique to investigate galaxy clustering at high redshifts; distance measurements to a very large number of galaxies can be made in a relatively modest investment of telescope time.

In McCracken et al. (2008) we used the photometric redshift catalogue derived in Ilbert et al. to make a series of galaxy samples containing galaxies selected by spectral types, rest-frame luminosities and redshifts in the four deep fields of the CFHTLS. An important aspect of these samples is that they are *volume-limited*, which is to say that galaxies of the same type in the sample are visible at both high and low redshift ends of the sample bin. In Figure 2.1 we shows a two-dimensional image of the distribution of objects in the absolute magnitude-redshift plane. Galaxies in this plot have been selected by apparent magnitude; if one were simply to take all galaxies between two redshift intervals, at progressively higher and higher redshifts each slice would be dominated by galaxies of higher instrinsic luminosity.

Inside this absolute luminosity / redshift plane galaxies were classified using multi-colour information in a similar way to other works in the literature (for example Lin et al., 1999; Wolf et al., 2003; Zucca et al., 2006). For each galaxy the rest-frame colours were matched with four templates from Coleman et al. (1980). These four templates have been optimised using the VVDS spectro-

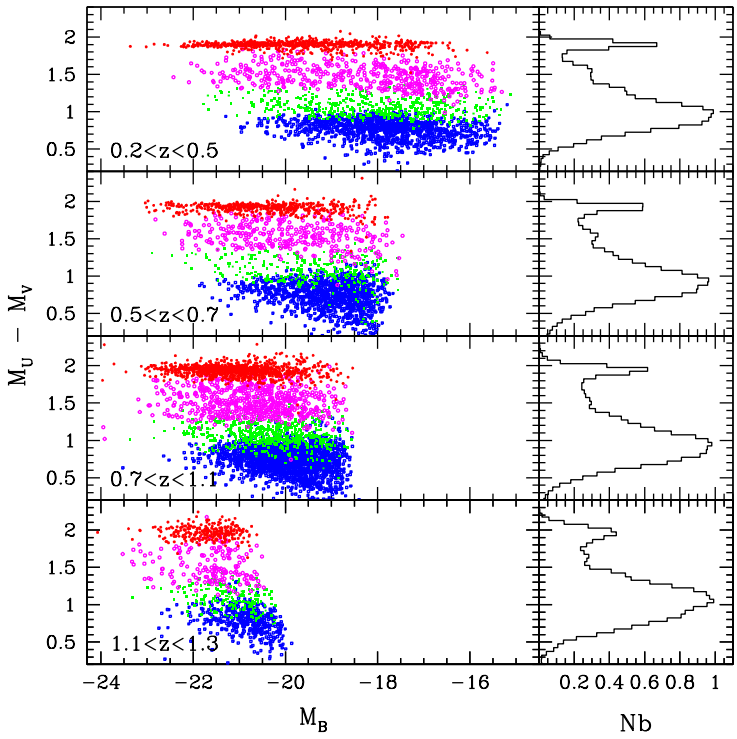

Figure 2.2 Rest frame $M_U - M_V$ colour as a function of B-band absolute magnitude for the CFHTLS D1 deep field. Each panel from top to bottom corresponds to the redshift bins used in this paper. The points show the four different best-fitting spectral types. Red, magnenta, green and blue points correspond to Coleman et al. Ell, Sbc, Scd, and Irr templates. The right-hand panels show the colour histograms for each redshift slice.

scopic redshifts, as described in Ilbert et al., and are presented in Fig.2 of this work. Galaxies have been divided in four types, corresponding to the optimised E/S0 template (type one), early spiral template (type two), late spiral template (type three) and irregular template (type four). Type four includes also starburst galaxies. We show in Figure 2.2 the rest-frame colour distribution of the galaxies for each type. Type one galaxies comprise most of the galaxies of the red peak of the bimodal colour distribution. The other types are distributed in the blue peak. The rest frame colours of galaxies become smoothly bluer from type one to type four respectively.

2.2.2 Clustering amplitudes as a function of type, luminosity and colour

We derive projected clustering amplitudes w as a function of angular scale θ from the samples described above by using the Landy & Szalay (1993) estimator:

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

where DD, DR and RR are the number of data–data, data–random and random–random pairs with separations between θ and $\theta+\delta\theta$. These pair counts are appropriately normalised; we typically generate random catalogues with ten times higher numbers of random points than input galaxies.

The limited precision of our photometric redshifts is an important point to consider. In a given redshift interval, $z_1 < z < z_2$ it is possible that a given galaxy may be in fact lay outside this redshift interval. To account for this, we employ a weighted estimator of $w(\theta)$, as suggested by Arnouts et al. (2002). In this scheme we weight each galaxy by the fraction of the galaxy's probability distribution function enclosed by the interval $z_1 < z < z_2$. Our measured correlation functions are corrected for the 'integral constraint' which arises from the fact that the mean density of the sample must be estimated from the sample itself:

$$C = \frac{1}{\Omega^2} \int \int \omega(\theta) d\Omega_1 d\Omega_2 \tag{2.2}$$

Error bars at each angular bin are computed from the field-to-field variance for each sample in the four CFHTLS deep fields. Once we have these *projected* measurements in hand, we can derive *spatial* clustering measurements at the effect redshift of each slice by making use us of photometric redshift information using Limber's equation (Limber, 1953), which we cite here for completeness:

$$A_w = r_0^{\gamma}(z_{\mathrm{eff}}) \sqrt{\pi} \frac{\Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)} \frac{\int_{z_1}^{z_2} g(z) (dN/dz)^2 dz}{[\int_{z_1}^{z_2} (dN/dz) dz]^2} \tag{2.3}$$

Thus, given a measurement A_w of the correlation function amplitude for the redshift slice under consideration and a knowledge of that slice's redshift distribution we can derive $r_0(z_{\mathrm{eff}})$.

It is worth mentioning the major assumption underlying this approach, namely that the correlation function can be adequately described by a simple power-law shape. As we will see in Chapter 4, this is at best an approximation and one consequence of this is that the resulting best-fitting amplitude and derived correlation amplitudes actually depend on the scales on which we carry out our fits. For the sample presented here, this is not problematic because the samples which are expected to have the largest deviation from a power-law are those which have the smallest numbers of galaxies (the pure type 1 and type 2 samples) and thus the dominant source of error comes from field-to-field variance.

The results for galaxy samples selected by luminosity in two redshift bins are summarised in Figure 2.3. These figures show the best-fitting correlation amplitude as a function of absolute luminosity for the three different samples (early, late and the full galaxy population) in two redshift ranges ($0.2 < z < 0.6$ and $0.7 < z < 1.1$) considered in this Section.

Several results are apparent: firstly, at all absolute magnitude slices and in

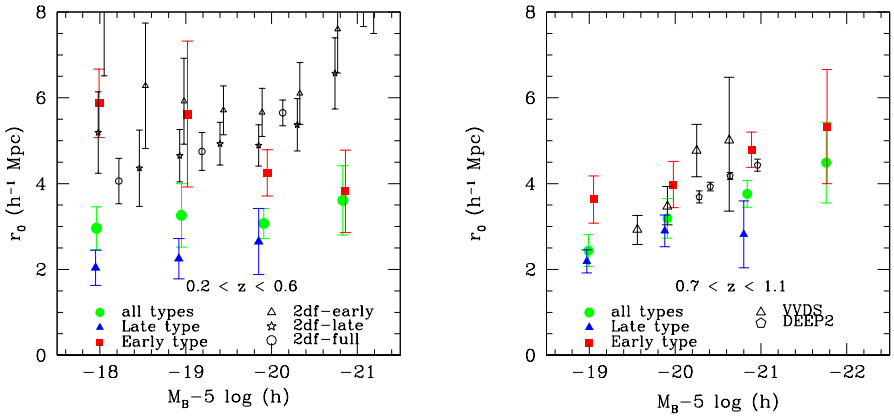

Figure 2.3 Left panel: The comoving correlation length r_0 as a function of median absolute luminosity and type for objects with $0.2 < z < 0.6$. Filled circles show the full galaxy population. In addition to type selection, the galaxy sample is selected in one-magnitude bins of absolute luminosity. Triangles and squares represent the late and early-type populations respectively. Right panel: measurements from our $0.6 < z < 1.1$ sample.

both redshift ranges, early-type galaxies are *always* more strongly clustered (higher values of r_0) than late-type galaxies. Secondly, we note that clustering amplitude for the late-type population is constant, remaining fixed at $\sim 2h^{-1}$ Mpc over a large range of absolute magnitudes and redshifts. The behaviour of the early-type population is more complicated. For the $0.2 < z < 0.6$ bin, we see some evidence that as the median luminosity increases, the clustering amplitude of this population *decreases*, from around $\sim 6h^{-1}$ Mpc for the faintest bins, to $\sim 5h^{-1}$ Mpc. We note that the difference in clustering amplitude between the early and late populations is smaller for the higher-redshift bin. We also note that the clustering amplitudes we derive for our blue and full-field galaxy populations are considerably lower than those reported by Norberg et al. at lower redshifts.

Figure 2.4 shows the clustering amplitudes of bright early-type galaxies; filled squares indicate galaxies of types one and two, and open squares represent a pure type-one sample. As expected, the clustering amplitudes of the pure type one population (with overall redder rest-frame colours) are higher than the combined type one and two samples. (We have also measured the clustering amplitude of the pure type four population and find that in this case clustering amplitudes are lower than the combined sample of types three and four.) Error bars are computed from the field-to-field variance.

Previous clustering measurements were made as a function of either absolute luminosity, type or redshift. For example, Meneux et al. (2006) made measurements in the VVDS spectroscopic redshift survey of the projected correlation

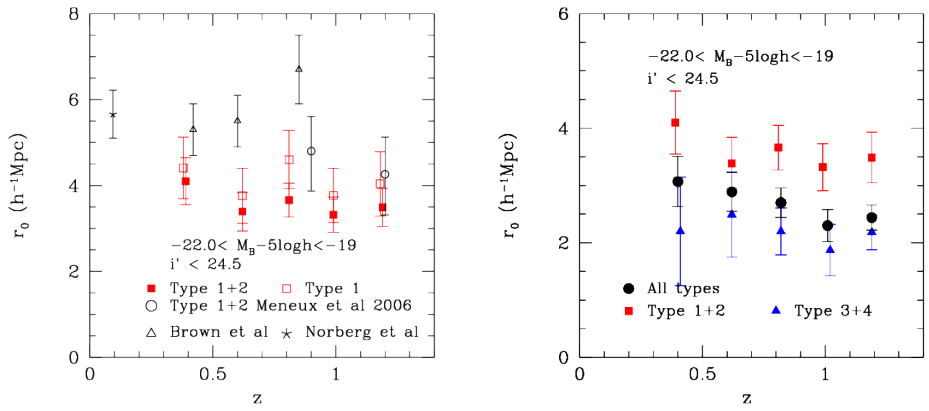

Figure 2.4 Left panel: Clustering amplitude of luminous red galaxies. Open and filled squares show measurements for type one and type one and two combined galaxy samples. Other points show measurements from the literature. Right panel: Redshift dependence of comoving galaxy correlation length r_0 for a series of volume limited samples for early types (squares), late types (triangles) and for the full sample (filled circles).

function w_p for early- and late-type galaxies, using a classification similar to ours. However, in their work galaxies were selected by apparent magnitude; at $z \sim 1$, this means that their rest frame luminosities are comparable to the brightest galaxies in our sample. We compare these $z \sim 1$ measures with our data; they are shown as the open circles in Figure 2.4. The open triangles represent measurements from Brown et al. (2003) who measured clustering of red galaxies selected using three-band photometric redshifts in the NOAO wide survey. Their results are above ours by at least one or two standard deviations.

Overall, broad trend seen here is that the clustering amplitude of bright early-type galaxies is constant with redshift. Figure 17 from McCracken et al. (2008), not shown here, demonstrates that clustering amplitudes for bright galaxies (both red and blue populations) is independent of redshift. The clustering amplitudes for the full galaxy population however at a given fixed absolute luminosity is lower at higher redshifts.

2.2.3 Discussion

A sample of 100,000 photometric redshifts in the CFHTLS legacy survey deep fields has been used to investigate the dependency of galaxy clustering on rest-frame colour, luminosity and redshift.

In rest-frame colour-selected samples at all redshift ranges we consistently find that galaxies with redder rest-frame colours are more strongly clustered than those with bluer rest-frame colours (Figure 2.3). Such an effect has long been observed for galaxies in the local Universe (for example Norberg et al.,

2002; Zehavi et al., 2005; Loveday et al., 1999; Guzzo et al., 1997) and at higher redshifts for samples selected by type and luminosity (Meneux et al., 2006; Coil et al., 2006). Numerical simulations find a similar effect: For example, Weinberg et al. (2004) show that older, redder galaxies are more strongly clustered. This is a generic prediction from most semi-analytic models and hydrodynamic simulations of galaxy formation: older, more massive galaxies formed in regions which collapsed early in the history of the Universe. At the present-day such regions are highly-biased with respect to the dark matter distribution. On the other hand, the luminous field galaxy population, dominated by blue star-forming galaxies at $z \sim 1$, is only weakly biased with respect to the dark matter distribution.

For the brightest ellipticals ($-22 < M_B - 5\log h < -19$) in our survey, we find that their clustering amplitude does not change with redshift (right panel of Figure 2.4), indicating that at $z \sim 1$ the elliptical population must be strongly biased with respect to the underlying dark matter distribution. Comparing our measurements for objects with redder rest-frame colours with those of other surveys, we find similar clustering amplitudes. Furthermore, sub-samples of galaxies with redder rest-frame colours produce even higher correlation amplitudes (left panel of Figure 2.4).

In a second set of selections we considered the dependence of galaxy clustering on luminosity and type in two broad redshift bins: $0.2 < z < 0.6$ and $0.7 < z < 1.1$ (we leave a 'gap' in the range $0.6 < z < 0.7$ to ensure that there is no contamination between high and low redshift ranges). Once again, for the most luminous objects ($M_B - 5\log h \sim -20$) the correlation amplitude is approximately constant between these two redshift bins. However, for fainter red objects, at a fixed absolute luminosity, we see a decline in correlation amplitude between $z \sim 0.4$ and $z \sim 1$; the same is true for samples selected purely by absolute magnitude. We find no evidence for a change in clustering amplitude at the same luminosity for the blue population with redshift.

At $0.2 < z < 0.6$, where we are complete to $M_B - 5\log h < -17$, we find that red galaxies with $M_B - 5\log h \sim -20$ are more strongly clustered than bluer galaxies of the same luminosity. Moreover as the sample rest frame luminosity decreases to $M_B - 5\log h \sim -18$ the clustering amplitude rises from $\sim 4h^{-1}$ Mpc to $\sim 6h^{-1}$ Mpc. A similar effect has been reported in larger, low-redshift samples in the local universe (Swanson et al., 2007; Norberg et al., 2002), where both the Sloan and 2dF surveys have found higher clustering amplitudes for redder objects fainter than L^*. Some evidence for this effect has also been reported in numerical simulations (Croton et al., 2007), which indicate that this behaviour arises because faint red objects exist primarily as satellite galaxies in halos of massive, strongly clustered red galaxies. This means that less luminous, redder objects reside primarily in higher density environments at $z \sim 0.5$. This is in agreement with recent studies of galaxy clusters at intermediate redshift which indicate a rapid build-up of low luminosity red galaxies in clusters since $z \sim 1$ (van der Wel et al., 2007). Our survey is not deep enough to probe to equivalent luminosities at $z \sim 1$.

Conversely for the redshift bin at $0.7 < z < 1.1$ we see that for the full galaxy population more luminous objects are more strongly clustered: $\sim 2h^{-1}$ Mpc for

galaxies with $M_B - 5\log h \sim -19$ and $\sim 4h^{-1}$ Mpc for galaxies with $M_B - 5\log h \sim -21$. At all luminosity bins, galaxies with redder rest-frame colours are always more strongly clustered than bluer galaxies.

Concluding, we may summarise our results as follows: firstly, for samples of galaxies with similar absolute luminosities, galaxies with redder rest-frame colours are *always* more strongly clustered than their bluer counterparts. Secondly, for the bluer galaxy populations, the correlation length depends only weakly on absolute luminosity. At lower redshifts, we find some evidence that redder galaxies with lower absolute luminosities are more strongly clustered. For the *entire* galaxy population (red and blue types combined) we find that as the median absolute magnitude increases, the overall clustering amplitude increases. For our the most luminous red and blue objects, the clustering amplitude does not change with redshift.

The overall picture we draw from these observations is that the clustering properties of the blue population is remarkably invariant with redshift and intrinsic luminosity. In general, galaxies with bluer rest-frame colours, which comprise the majority of galaxies in our survey, have lower clustering amplitudes (typically, $\sim 2h^{-1}$ Mpc) than the redder populations. The clustering amplitude of the blue population depends only weakly on redshift and luminosity. This is consistent with a picture in which bluer galaxy types exist primarily in lower density environments.

In contrast, the clustering amplitude of the low-luminosity red population is lower at higher redshifts. In Figure 2.3 we see that for the luminous ($M_B - 5\log h \sim -20$) red population, the correlation amplitude does not change with redshift. Moreover, at a fixed absolute luminosity, the correlation amplitude of the full galaxy population and the magnitude-selected galaxy population decreases from $z \sim 0.4$ to $z \sim 1.1$, in step with the underlying dark matter distribution.

2.3 These results in context

Since the publication of these results, other authors have confirmed the invariance of the clustering properties of bright early-type galaxies with redshift (Brown et al., 2008; Foucaud et al., 2010), or found low clustering amplitudes for faint blue galaxies in the local universe, comparable to our low-redshift low-luminosity bin (Carlberg et al., 2009). In general, rest-frame colour (or, approximately equivalently, best-fitting spectral type) seems to play a more important role in determining the clustering amplitude of the galaxy population than selection by rest-frame luminosity; the difference in amplitudes between early and late galaxy types is much larger than the difference between luminous and less luminous galaxies of the same spectral type. This fact certainly suggests the important role that environment plays in determining galaxy clustering.

At the time this work was prepared, interpretation of our results was made more complicated by three factors. Firstly, our samples here were selected by absolute luminosity, and it is well known that the characteristic luminosity of the different galaxy populations evolves significantly over $0 < z < 1$ (for blue

galaxy populations, for example there is around two magnitudes of evolution in $M*$ in this redshift range (Zucca et al., 2006)). Secondly, although some numerical simulations have attempted to make predictions for galaxy properties at these redshifts(Kauffmann et al., 1999) mass resolution effects and the limited volumes of the samples (as well often problematic separation between galaxy types) made initial comparisons problematic. Finally, although the number of galaxies in the CFHTLS-deep is large, we are still limited by shot-noise effects in many selection bins. Therefore, to advance in our understanding of these phenomena three requirements need to be met: better selection criteria for galaxies (preferably by stellar mass); better models; and, of course larger data sets.

Qualitatively these trends can be understood in terms of the biased galaxy formation model discussed in the introduction. Baryonic matter, collecting in the centres dark matter haloes, is heated and compressed to form stars and galaxies. The more massive the underlying dark matter halo, the more clustered the tracer galaxy population becomes. The evolution of galaxy clustering over the age of the Universe is driven in large part by the growth of dark matter haloes over cosmic time, and one of the key factors controlling galaxy formation appears to be the mass of the underlying dark matter halo. Therefore, for a *quantitative* understanding of these measurements became clear that it would be necessary to develop a model grounded in the framework of contemporary models of structure formation and which can make predictions at a range of redshifts. Just such a model will be the subject of Chapter 4. Before this, however, we will first turn to the Universe at $z \sim 2$, a crucial epoch in the formation of structures and the assembly of mass in stars into galaxies.

Chapter 3

Passive galaxies and the Universe at intermediate redshifts

3.1 Introduction

Many observations now indicate that between redshifts of one and two, most mass is assembled into galaxies and most star-formation took place. In 1996, when the first plots were made showing how the star-formation rate evolves over the history of the Universe, (Madau et al., 1996; Lilly et al., 1996), there was a large but suggestive gap between redshifts one and two. From redshifts around one to the present day, the star-formation rate rapidly decreased; at the higher redshifts, it increased to reach a maximum at $z \sim 2$ (although admittedly this behaviour is somewhat disputed today thanks to a more accurate treatment of dust at high redshifts). But what happened in the "gap"?

Filling in the details of what exactly happened between $1 < z < 2$ has been challenging, because it means making observations in near-infrared wavelengths as all potentially useful spectral features move out of the rest-frame optical wavebands. In addition, the sky background is very bright in the near-infrared: only a few sites in the world can observe faint extragalactic objects at these wavelengths. As optical charge-coupled detectors have a red cutoff at 1μm (although has improved slightly recent years with the availability of red-sensitive thinned detectors) an alternate, extremely expensive, detector technology needed to be developed. At first these detectors covered only a very small field of view, and it was not until the beginning of the 1990s when large arrays sensitive at near-infrared wavelengths suitable for surveys became available. Early papers, principally using array cameras developed for the Hubble space telescope on telescopes in Hawaii, demonstrated that the numbers of galaxies as a function of magnitude did not now show the "excess" in counts relative to simple models in which galaxies undergo only evolution in luminosity. Understanding this result requires spectroscopic or photometric redshift surveys capable of reaching at least $K_{AB} \sim 20$. In this chapter I will present a summary of work carried out from 2007 to 2010 in the framework of the "COSMOS"

(Scoville et al., 2007) project. The principal objective was to provide wide-field near-infrared data reaching sufficiently deep to reach at least a M_* mass limit at $z \sim 2$.

At redshifts below one various spectroscopic surveys such as the VVDS (Le Fèvre et al., 2005; Ilbert et al., 2005; Pozzetti et al., 2007) DEEP2 (Faber et al., 2007; Noeske et al., 2007) and zCOSMOS-bright (Lilly et al., 2007; Silverman et al., 2009; Mignoli et al., 2009) have mapped the evolution of galaxy and active galactic nuclei (AGNs) populations areas of a few square degrees. There is now general agreement that star formation in the Universe peaks at $1 < z < 2$ and that $\sim 50\% - 70\%$ of mass assembly took place in the redshift range $1 < z < 3$ (Connolly et al., 1997; Dickinson et al., 2003; Arnouts et al., 2007; Pozzetti et al., 2007; Noeske et al., 2007; Pérez-González et al., 2008). Alternatively stated, half of today's stellar mass appears to be in place by $z \sim 1$ (Drory et al., 2005; Fontana et al., 2004). Hierarchical structure formation models can have difficulty in accounting for the large number of evolved systems at relatively early times in cosmic history (Fontana et al., 2006). Furthermore, there is some evidence that around half the stellar mass in evolved or "passive" galaxies assembled relatively recently (Bell et al., 2004). It is thus of paramount importance to gather the largest sample of galaxies possible at $z \sim 2$.

Near-infrared imaging and spectroscopy are essential in the redshift range $1.4 < z < 3.0$, and for the moment the role of environment and large-scale structure at these redshifts remains largely unexplored (Renzini & Daddi, 2009). It is also worth mentioning that in addition to making it possible to select galaxies in this important range, near-infrared galaxy samples offer several advantages compared to purely optical selections (see, for example Cowie et al. (1994)). They allow us to select $z > 1$ galaxies in the rest-frame optical, correspond more closely to a stellar-mass-selected sample and are less prone to dust extinction. As $k-$ corrections in $K-$ band are insensitive to galaxy type over a wide redshift range, near-infrared-selected samples provide a fairly unbiased census of galaxy populations at high redshifts (providing that the extinction is not too high, as in the case of some submillimeter galaxies). Such samples represent the ideal input catalogues from which to extract targets for spectroscopic surveys as well as for determining accurate photometric redshifts.

Cowie et al. (1996) carried out one of the first extremely deep, complete $K-$ selected surveys and made the important discovery that star-forming galaxies at low redshifts have smaller masses than actively star-forming galaxies at $z \sim 1$, a phenomenon known as "downsizing". Stated another way, the sites of star-formation "migrate" from higher-mass systems at high redshift lower-mass systems at lower redshifts. More recent K-selected surveys include the K20 survey (Cimatti et al., 2002) reaching $K_{\rm AB} \simeq 21.8$ and the GDDS survey (Abraham et al., 2004) which reached $K_{\rm AB} \simeq 22.4$ provide further evidence for this picture. The areas covered by these surveys was small, comprising only ~ 55 arcmin^2 and ~ 30 arcmin^2 K20 and GDDS respectively. While Glazebrook et al. (2004) and Cimatti et al. (2008) provided spectroscopic confirmation of evolved systems $z > 1.4$ and provided further evidence for the downsizing picture (Juneau et al., 2005) their limited coverage made them highly susceptible to the effects of cosmic variance. It became increasingly clear that much larger samples of

passively evolving galaxies were necessary.

At $K < 20$ the number of passive galaxies at $z \sim 2$ redshifts is small and spectroscopic followup of a complete magnitude-limited sample can be time-consuming. For this reason a number of groups have proposed and validated techniques based on applying cuts in colour-colour space to isolate populations in certain redshift ranges. Starting with the Lyman-break selection at $z \sim 3$ (Steidel et al., 1996), similar techniques have been applied at intermediate redshifts to select extremely red objects (EROs; Hu & Ridgway (1994)) or distant red galaxies (DRGs; Franx et al. (2003)) and the "BzK" technique used in this paper (Daddi et al., 2004). The advantage of these methods is that they are easy to apply requiring at most only three or four photometric bands; their disadvantage being that the relationships between each object class is complicated and some selection classes contain galaxies with a broad range of intrinsic properties (Daddi et al., 2004; Lane et al., 2007; Grazian et al., 2007). The relationship to the underlying complete galaxy population can also be difficult to interpret (Fèvre et al., 2005a). Ideally, one like to make complete mass-selected samples at a range of redshifts but such calculations require coverage in many wave bands and can depend sensitively on the template set (Pozzetti et al., 2007; Longhetti & Saracco, 2009). Moreover, for redder populations the mass uncertainties can be even larger; Conroy et al. (2009) estimate errors as larger as 0.6 dex at $z \sim 2$.

At $z \sim 1.4$ Daddi et al. (2004) used spectroscopic data from the K20 survey in combination with stellar evolutionary tracks to define their "BzK" technique. They demonstrated that in the $(B - z)$ $(z - K)$ colour-colour plane, star-forming galaxies and evolved systems are well separated at $z > 1.4$, making it possible accumulate larger samples of passive galaxies at intermediate redshifts that was possible previously with simple one-colour criterion. Subsequently, several other surveys have applied these techniques to larger samples of near-infrared selected galaxies. In one of the widest surveys to date, Kong et al. (2006) constructed K-band selected samples over a ~ 920 arcmin^2 field reaching $K_{\rm AB} \simeq 20.8$ reaching to $K_{\rm AB} \simeq 21.8$ over a 320 arcmin^2 sub-field. The exploration of a field of this size made possible to measure the clustering properties of star-forming and passive galaxy sample and to establish that passively evolving galaxies in this redshift range are substantially more strongly clustered than star-forming ones, indicating that a galaxy-type - density relation reminiscent of the local morphology-density relation must be already in place at $z \gtrsim 1.4$. The UKIDSS survey reaches $K_{\rm AB} \sim 22.5$ over a ~ 0.62-deg^2 area included in the Subaru-*XMM Newton* Deep Survey and Lane et al. (2007) used this data set to investigate the different commonly-used selected techniques at intermediate redshifts, concluding most bright DRG galaxies have spectra energy distributions consistent with dusty star-forming galaxies or AGNs at ~ 2. They observe a turn-over in the number counts of passive BzK galaxies.

Other recent works include the MUSYC/ECDFS survey covering ~ 900 arcmin^2 to $K_{\rm AB} \sim 22.5$ over the CDF South field (Taylor et al., 2009), not to be confused with the GOODS-MUSIC catalog of Fontana et al. (2006), which covers 160 arcmin^2 of GOODS-South field to $K_{\rm AB} \sim 23.8$. This K-band selected catalogue, as well as the FIREWORKS catalog by Wuyts et al. (2008), are based

on the ESO Imaging Survey coverage of the GOODS-South field[1] These studies have investigated, amongst other topics, the evolution of the mass function at $z \sim 2$ and fraction of the red sequence galaxies which were already in place at $z \sim 2$.

Finally, one should mention that measuring the distribution of a "tracer" population, either red passive galaxies or normal field galaxies can provide useful additional information on the galaxy formation process. In particular one can estimate the mass of the dark matter haloes hosting the tracer population and, given a suitable model for halo evolution, identify the present-day descendants of the tracer population, as has been done for Lyman break galaxies at $z \sim 3$. A few studies have attempted this for passive galaxies at $z \sim 2$, but small fields of view have made these studies somewhat sensitive to the effects of cosmic variance. COSMOS provides (Scoville et al., 2007) a contiguous 2 deg^2 equatorial field with extensive multi-wavelength coverage is well suited to probing the universe at intermediate redshift.

In this chapter we use a flat lambda cosmology ($\Omega_m = 0.3$,$\Omega_\Lambda = 0.7$) with $h = H_0/100$ km s^{-1} Mpc^{-1}. All magnitudes are given in the AB system, unless otherwise stated.

3.2 The BzK selection

To compare with previous works we wanted our photometric selection criterion to match as closely as possible as the original "BzK" selection proposed in Daddi et al. (2004). As our filter set is not the same as this work we applied small offsets (based on the tracks of synthetic stars), following a similar procedure outlined in Kong et al. (2006).

To account for the differences between our Subaru $B-$ filter and the $B-$ VLT filter used by Daddi et al. (2004) we use this empirically-derived transformation, defining $bz = B_{J_{\rm total}} - z^+_{\rm tot}$, then for blue objects with $bz < 2.5$,

$$bz_{\rm cosmos} = bz + 0.0833 \times bz + 0.053 \tag{3.1}$$

otherwise, for objects with $bz > 2.5$,

$$bz_{\rm cosmos} = bz + 0.27 \tag{3.2}$$

This "$bz_{\rm cosmos}$" quantity is the actual corrected $(B-z)_{\rm AB}$ colour which we use in this paper.

Finally we divide our catalogue into galaxies at $z < 1.4$, stars, star-forming galaxies and passively evolving galaxies at $1.4 < z < 2.5$, by first defining the BzK quantity introduced in Daddi et al. (2004):

$$BzK \equiv (z - K) - (B - z) \tag{3.3}$$

For galaxies expected at $z > 1.4$ star-forming galaxies (hereafter $sBzK$) are selected as those objects with $BzK > -0.2$. Old, passively evolving galaxies

[1] http://www.eso.org/science/goods/releases/20050930./

(hereafter $pBzK$) can be selected as those objects which have

$$BzK < -0.2, (z - K) > 2.5. \tag{3.4}$$

Stars are selected using this criteria:

$$(z - K) < -0.5 + (B - z) \times 0.3 \tag{3.5}$$

Finally, the full galaxy sample consists simply of objects which do *not* fulfill this stellarity criterion. The result of this division is illustrated in Figure 3.1. The solid line represents the colours of stars in the BzK filter set of Daddi et al. using the empirically corrected spectra presented in Lejeune et al. (1997), and it agrees very with our corrected stellar locus.

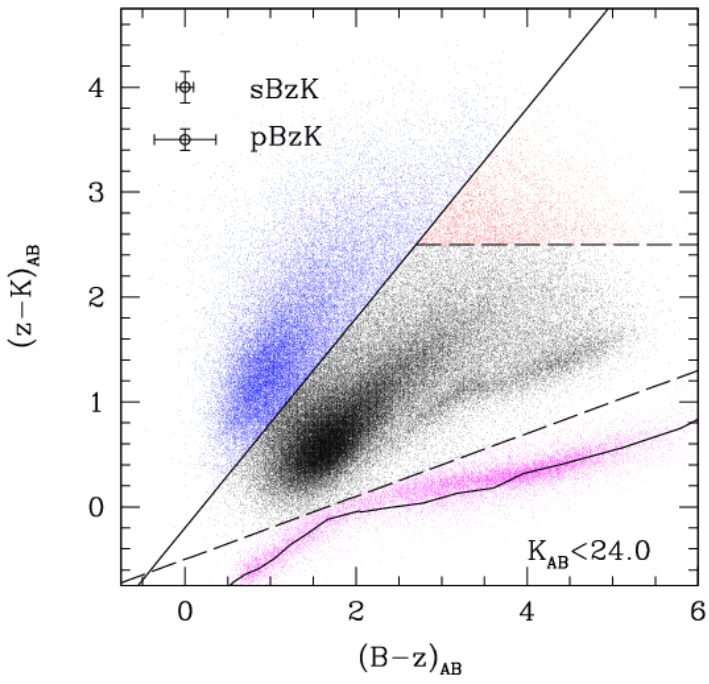

Figure 3.1 The $(B-Z)_{\rm AB}$ vs $(z-K)_{\rm AB}$ diagram for all galaxies in the COSMOS field. Four distinct regions are shown: stars (lower part of the diagram), galaxies (middle), star-forming galaxies (left) and passively-evolving galaxies (top right). The solid line shows the colours of stars in the BzK filter set of Daddi et al. computed using the models of Lejeune et al. (1997)

3.3 Source counts

We now present number counts of the three populations selected in the previous Section.

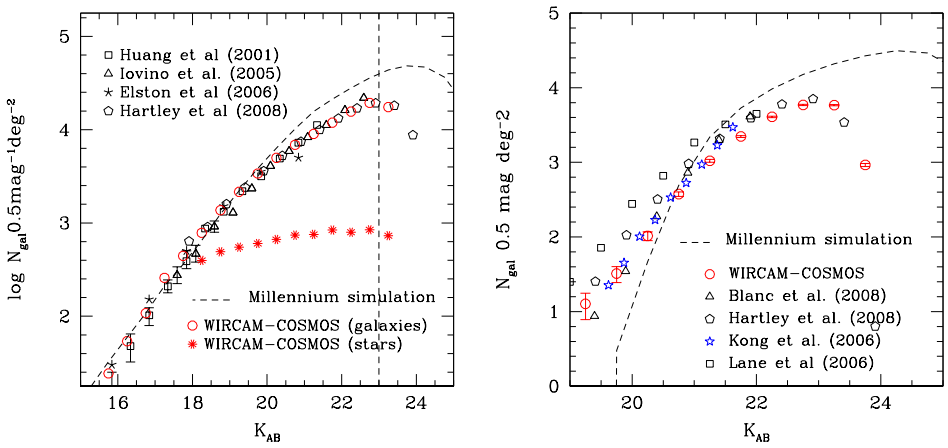

Figure 3.2 Left panel:$K_{\rm s}-$ selected galaxy and star counts from the COSMOS survey (open circles and stars respectively) compared to measurements from recent wide-field near-infrared surveys; right panel: Number counts for star-forming BzK galaxies in the compared to literature measurements and the predictions of the model of Kitzbichler & White (dashed line).

3.3.1 Star and galaxy counts

Figure 3.2 shows our differential galaxy number counts compared to a selection of measurements from the literature. We note that at intermediate magnitudes ($20 < K_s < 22$) counts from the four surveys presented here are remarkably consistent (Elston et al., 2006; Huang et al., 1997; Hartley et al., 2008). At $16 < K_s < 20$ discrepancies between different groups concerning measurement of total magnitudes and star-galaxy separation leads to an increased scatter. At these magnitudes, shot noise and large-scale-structure begin to dominate the number count errors.

The COSMOS-WIRCam survey is currently the only work to provide unbroken coverage over the range $16 < K_s < 23$. In addition, our colour-selected star-galaxy separation provides a very robust way to reject stars from our faint galaxy sample. These stellar counts are shown by the asterisks in Figure 3.2. We note that at magnitudes brighter than $K_{\rm s} \sim 18.0$ our stellar number counts become incomplete because of saturation in the Subaru B image (our catalogues exclude any objects with saturated pixels which preferentially affect point-like sources). Note that even at very faint magnitudes, our stellar counts are remarkably free from contamination by faint galaxies, a point which we will return to later.

3.3.2 $sBzK$ and $pBzK$ counts

Figure 3.2 shows the counts of star-forming BzK galaxies compared to measurements from the literature. We note an excellent agreement with the counts

in Kong et al. (2006) and the counts presented by the MUYSC collaboration (Blanc et al., 2008). However, the counts presented by the UKIDSS-UDS group (Lane et al., 2007; Hartley et al., 2008) are significantly offset compared to our counts at bright magnitudes, and become consistent with it by $K_{\rm s} \sim 22$. These authors attribute the discrepancy to cosmic variance but we find photometric offsets a more likely explanation (see below).

The left pane of Figure 3.3 shows in more detail the zone occupied by passive galaxies in Figure 3.1. Left of the diagonal line are objects classified as star-forming BzK galaxies. Objects not detected in B are plotted as right-pointing arrows with colours computed from the upper limit of their $B-$ magnitudes. This means that in addition to the galaxies already in the $pBzK$ selection box, fainter $sBzK$ with B-band non-detections (shown with the green arrows) may be scattered rightwards into the $pBzK$ region.

Counts for our passive galaxy population including these "additional" objects are represented by the hatched region in the right panel of Figure 3.3. The upper limit for the source counts in this Figure represents the case in which *all* the $(z-K)_s > 2.5$ sources undetected in B are scattered into the $pBzK$ region. Even accounting for these additional objects we unambiguously observe a flattening and subsequent turn-over in the passive galaxy counts at around $K_{\rm s} \sim 22$, well above the completeness limit of either our $K_{\rm s}-$ or $B-$ data in agreement with (Hartley et al., 2008). This turn-over is essentially due to the finite volume probed by our $pBzK$ selection and our counts are in effect a direct probe of the $pBzK$ luminosity function.

This upper limit, however, is a conservative estimate. We have made a better estimate of this upper limit by carrying out a stacking analysis of the objects not detected in $B-$ in both the passive and star-forming regions of the BzK diagram. For each apparent $K_{\rm s}$ magnitude bin in Table we median-combine Subaru $B-$band postage stamps for objects with no B-band detection, producing separate stacks for the star-forming and passive regions of the BzK diagram. In both cases, objects below our detection limit are clearly visible (better than a three-sigma detection) in our stacked images at each magnitude bin to $K_{\rm s}$ 23. By assuming that the mean B magnitude of the stacked source to be the average magnitude of our undetected sources, we can compute the average $(B-z)$ colour of our undetected sources, and reassign their location in the BzK diagram if necessary. This experiment shows that at most only 15% of the star-forming BzK galaxies undetected in $B-$ move to the passive BzK region.

3.3.3 Comparison with semi-analytic models

In Figures 3.2 and 3.3 we show counts of galaxies extracted from the semi-analytical models presented in Kitzbichler & White (2007). To derive counts of quiescent galaxies, we selected all galaxies at $z > 1.4$ in the star-formation rate - mass plane (see, for example Figure 18 from Daddi et al. (2007)) which have star-formation rates less than three times the median value for a given mass. Star-forming objects were defined as those galaxies which did `not` obey

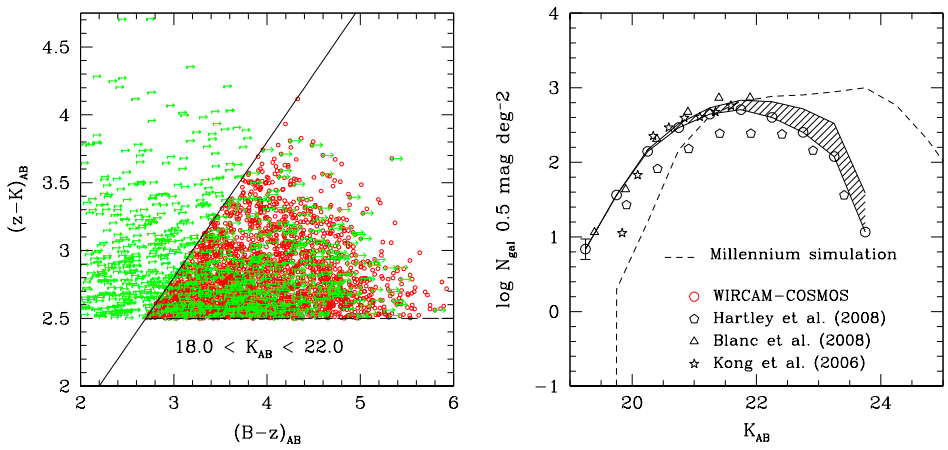

Figure 3.3 Left panel: Selection diagram for the passive BzK population. Objects with rightward-pointing arrows are galaxies plotted at the lower limit of their $(B-z)_{\rm AB}$ colours. Circles are objects selected as $pBzK$ galaxies. Right panel: Differential number counts for the passive BzK population in the COSMOS-WIRCam survey (open circles) compared to measurements from the literature and the predictions of the model of Kitzbichler & White (dashed line). The shaded region represents an upper limit on the number counts of passive $BzKs$ if all star-forming BzKs in left panel were moved into the region of the Figure occupied by the passively evolving population.

this criterion, in this redshift range.

In all plots, the models over-predict the number of faint galaxies, an effect already observed for the $K-$ selected samples investigated in Kitzbichler & White. However, for both our $sBzK$ and $pBzK$ counts, the number of luminous objects $z \sim 2$ is *under-predicted* by the models. It is interesting to compare our results with Figure 7 from Kitzbichler & White, which shows the stellar mass function for their models. At $z \sim 2$, the models both under-predict the number of massive objects and over-predict the number of less massive objects, an effect mirroring the overabundance of luminous $pBzK$ objects with respect to the Kitzbichler & White model seen in our data. A more detailed analysis based on a full mass function analysis (Ilbert et al., 2010) finds a similar discrepancy.

3.4 Photometric redshifts for the $pBzK$ and $sBzK$ population

For many years studies of galaxy clustering at $z \sim 2$ have been hindered by our imperfect knowledge of the source redshift distribution and small survey fields. To overcome these limitations the $2\,\mathrm{deg}^2$ COSMOS field has been observed in over thirty broad intermediate and narrow photometric bands and very precise photometric redshifts are now available (Ilbert et al., 2009). These photometric

redshifts were computed using in addition to the deep Subaru data described in Capak et al. (2007) and new intermediate band data, the K_s data presented in this paper, J data from near-infrared camera WFCAM at the United Kingdom Infrared Telescope and IRAC data from the Spitzer-COSMOS survey (sCOSMOS, Sanders et al. (2007)). These near- and mid-infrared band-passes are an essential ingredient to compute accurate photometric redshifts in the redshift range $1.4 < z < 2.5$, in particular because they permit the location of the 4000 Å break to be determined accurately. Moreover, spectroscopic redshifts of 148 $sBzK$ galaxies with a $\bar{z} \sim 2.2$ from the early zCOSMOS survey (Lilly et al., 2007) have been used to check and train these photometric redshifts.

We have cross-correlated our catalogue with photometric redshifts to derive redshift selection functions for each photometrically-defined galaxy population. Note that although photometric redshifts are based on an optically-selected catalogue, this catalogue is very deep ($i' < 26.5$) and contains almost all the objects present in the K_s-band selected catalogue.

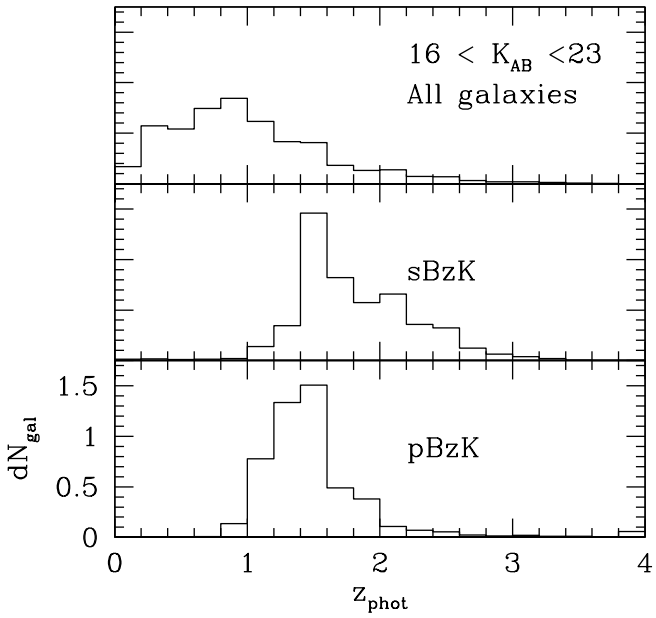

Figure 3.4 The normalised redshift distribution for field galaxies (top panel) $sBzK$ (middle panel) and $pBzK$ galaxies (bottom panel), computed using the 30-band photometric redshifts presented in Ilbert et al. (2009).

Figure 3.4 shows the redshift distribution for all K_s-selected galaxies, as well as for BzK-selected passively-evolving and star-forming galaxies in the magnitude range $18.0 < K_{AB} < 23.0$. We have computed the redshift selection function in several magnitude bins and found that the effective redshift z_{eff}

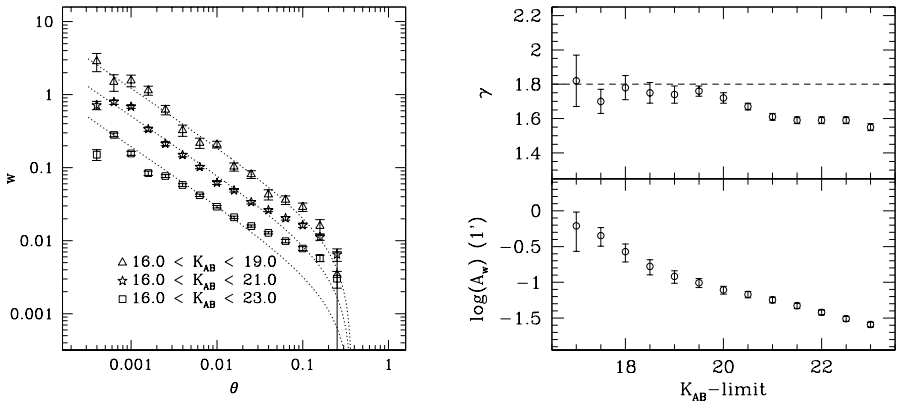

Figure 3.5 Left panel: The clustering amplitude w for galaxies in three slices of apparent magnitude. The dotted line shows a fit to a slope $\gamma = 1.8$ with an integral constraint appropriate to the size of our field applied; right panel, lower: clustering amplitude at $1'$ as a function of K_s limiting magnitude for the full galaxy sample. upper panel: panel: Best fitting slope over entire angular range of our survey ($-3.2 < \log(\theta) < 0.2$).

does not depend significantly on apparent magnitude for the $sBzK$ and $pBzK$ populations. Figure 3.4 illustrates the excellent efficency of the BzK criterion to separate the principal galaxy population at redshifts $1.4 < z < 2.5$ for $sBzK$ galaxies and $1 < z < 2$ for $pBzK$ galaxies.

By using only the blue grism of the VIMOS spectrograph at the VLT, the zCOSMOS-Deep survey is not designed to target $pBzK$ galaxies, and so no spectroscopic redshifts were available to train the photometric redshifts of objects over the COSMOS field. At these redshifts the main spectral features of $pBzK$ galaxies, namely Ca II H& K and the 4000 Å break, have moved to the near infrared. Hence, optical spectroscopy can only deliver redshifts based on identifying the so called Mg-UV feature at around 2800Å in the rest frame. All in all, spectroscopic redshifts of passive galaxies at $z > 1.4$ are now available for only a few dozen objects (Daddi et al., 2005; Glazebrook et al., 2004; Cimatti et al., 2008). We note that the average spectroscopic redshift of these objects ($\bar{z} \sim 1.7$) indicates that the average photometric redshift of $\bar{z} \sim 1.4$ of our $pBzK$ galaxies to the same K_s-band limit may be systematically underestimated. For the medium to short term one has to unavoidably photometric redshifts when redshifts are needed for large numbers of passive galaxies at $z > 1.4$.

3.5 Clustering of near-infrared-selected samples

3.5.1 Field galaxies in magnitude-selected samples

Methods similar to those presented in the previous Chapter were used to determine the angular correlation function w, in logarithmically separated bins. Figure 3.5 shows $w(\theta)$ for galaxies in three magnitude slices ranges. It is clear that the slope of w becomes shallower at fainter magnitudes. At small separations (less than $1''$) w decreases due to object blending. The right panel of Figure 3.5 we shows the dependence of slope γ on $K_{\rm s}$ limiting magnitude. Here we fit for the slope and amplitude simultaneously for all slices. At bright magnitudes the slope corresponds to the canonical value of ~ 1.8; towards intermediate magnitudes it becomes steeper and fainter magnitudes progressively flatter. It is interesting to compare this Figure with the COSMOS optical correlation function presented in Figure 3 of McCracken et al. (2007) which demonstrated that the slope of the angular correlation function becomes progressively flatter at fainter magnitudes. One possible interpretation of this behaviour is that at bright magnitudes our $K_{\rm s}$-selected samples are dominated by bright, red galaxies which have an intrinsically steeper correlation function slope; our fainter samples are predominantly bluer, intrinsically fainter objects with shallower intrinsic correlation function slope.

Finally, it is instructive to compare our field galaxy clustering amplitudes with literature measurements as our survey is by far the largest at these magnitude limits. Figure 3.6 shows the scaling of the correlation amplitude at one degree as a function to limiting $K_{\rm s}$ magnitude, compared a compilation of measurements from the literature. To make this comparison, we have assumed a fixed slope of $\gamma = 1.8$ and converted the limiting magnitude of each catalogue to Vega magnitudes. In general our results are within the 1σ error bars of most measurements, although it does appear that the COSMOS field is slightly more clustered than other fields in the literature, as we have discussed previously (McCracken et al., 2007).

3.5.2 Clustering at $z \sim 2$

We now consider the clustering properties of passive and star-forming galaxy candidates at $z \sim 2$ selected using our BzK diagram.

The upper panel of Figure 3.7 shows the angular correlation functions for our $pBzK$, $sBzK$ and for all galaxies. In each case we apply a $18.0 < K_s < 23.0$ magnitude cut. For comparison we show the clustering amplitude of dark matter computed using the redshift selection functions presented in Section 3.4 and the non-linear power spectrum approximation given in Smith et al. (2003). At intermediate to large scales, the clustering amplitude of field galaxies and the $sBzK$ population follows very well the underlying dark matter. The lower panel of Figure 3.7 shows the bias b, as a function of scale, computed simply as $b(\theta) = \sqrt{(w_{\rm gal}(\theta)/w_{\rm dm}(\theta)}$. Dashed, dotted and solid lines show b values for $pBzK$, $sBzK$ and field galaxies retrospectively (in this case our w measurements have

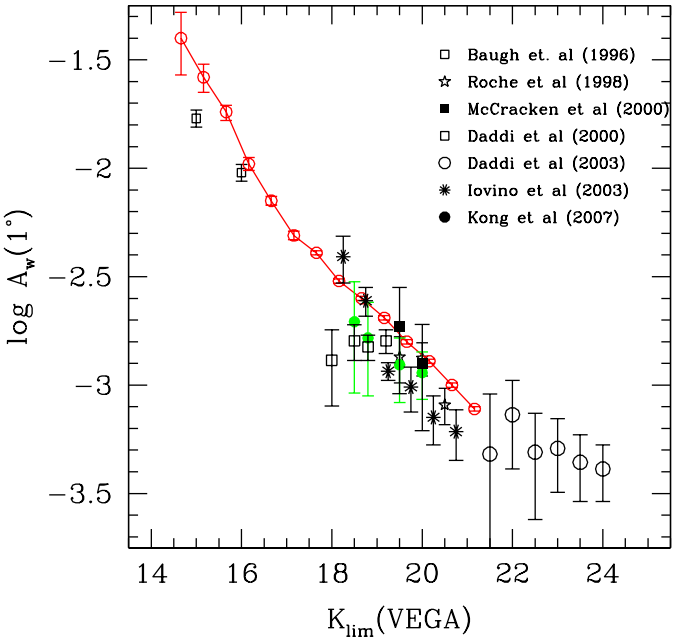

Figure 3.6 Fitted clustering amplitude at 1 degree as a function of $K_{\rm VEGA}$ limiting magnitude (connected open circles), compared to values from the literature.

been corrected for the integral constraint). The bias for the faint field galaxy population is close to 1 indicating that the faint $K_{\rm s}-$ selected galaxy population is indeed a very good tracer of underlying mass.

3.5.3 Spatial clustering

To de-project our measured clustering amplitudes and calculate the co-moving correlation lengths at the effective redshifts of our survey slices we use the photometric redshift distributions presented in Section 3.4.

Given a redshift interval z_1, z_2 and a redshift distribution dN/dz we define the effective redshift in the usual way, namely, $z_{\rm eff}$ is defined as

$$z_{\rm eff} = \int_{z_1}^{z_2} z(dN/dz)dz / \int_{z_1}^{z_2} (dN/dz)dz. \qquad (3.6)$$

Using these redshift distributions together with the fitted correlation amplitudes in presented in Sections 3.5.1 and 3.5.2 we can derive the co-moving correlation lengths r_0 of each galaxy population at their effective redshifts using the usual Limber (1953); Peebles (1980) inversion. We assume that r_0 does not change over the redshift interval probed.

It is clear that our use of photometric redshifts introduces an additional

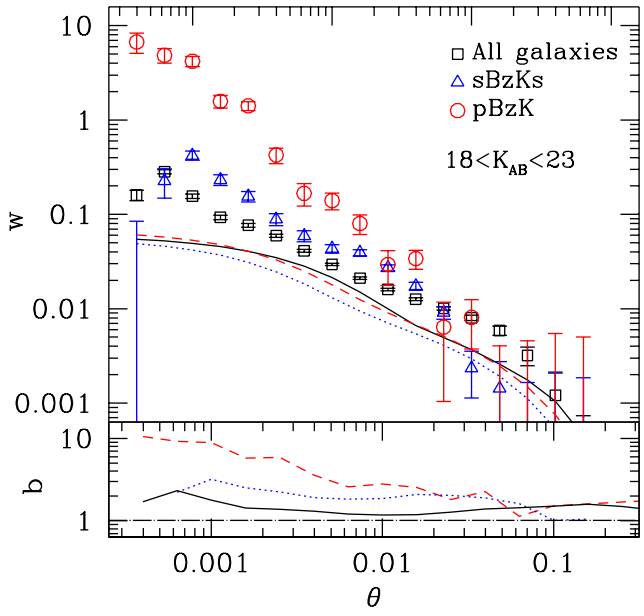

Figure 3.7 Top panel – The amplitude of the galaxy correlation function w for field galaxies, star-forming BzK galaxies and passive BzK galaxies with $18 < K_{AB} < 23$ (squares, triangles and circles). The lines show the predictions for the non-linear clustering amplitudes of dark matter computed using the non-linear power spectrum. Bottom panel: bias, b for $pBzK$, $sBzK$ and field galaxies.

uncertainty in r_0. We attempted to estimate this uncertainty by using the probability distribution functions associated with each photometric redshift to compute an ensemble of r_0 values, each estimated with a different $n(z)$. The resulting error in r_0 from these many realisations is actually quite small, ~ 0.02 for the $pBzK$ population. Of course, systematic errors in the photometric redshifts could well be much higher than this. Figure 9. in Ilbert et al. shows the 1σ error in the photometric redshifts as a function of magnitude and redshift. Although all galaxy types are combined here, we can see that the approximate 1σ error in the photometric redshifts between $1 < z < 2$ is ~ 0.1. Our estimate of the correlation length is primarily sensitive to the median redshift and the width of the correlation length. An error ~ 0.1 translates into an error of ~ 0.1 in r_0. We conclude that, for our $pBzK$ and $sBzK$ measurements, the dominant source of uncertainty is comes from our errors on w.

We note that previous investigations of the correlation of passive galaxies always assumed a fixed $\gamma = 1.8$; from Figure 3.5 it is clear that our slope is much steeper. These surveys, however, fitted over a smaller range of angular scales and therefore could not make an accurate determination of the slope for the

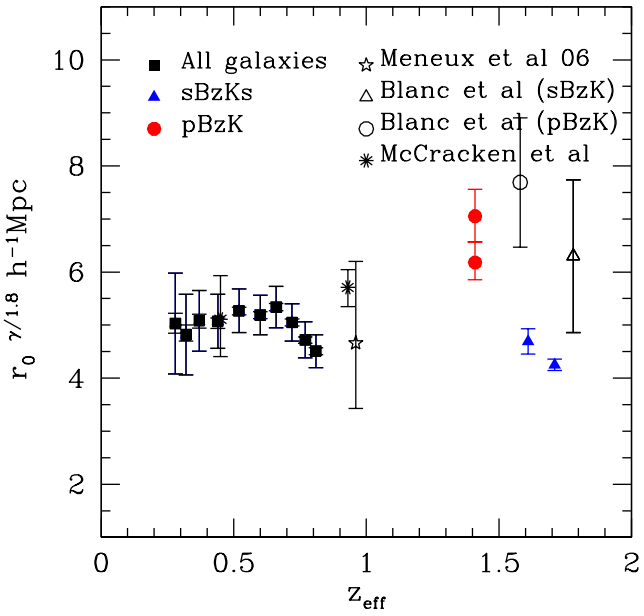

Figure 3.8 The rescaled co-moving correlation length $r_0^{\gamma/1.8}$ as a function of redshift for K_s- selected field galaxies (filled squares), $sBzK$ galaxies (filled triangles) and $pBzK$ galaxies (filled circles). Also shown are results from lower-redshift optically selected red galaxies and higher redshift K_s- selected samples.

$pBzK$ population. In all cases we fit for both γ and A_w. These measurements are plotted in Figure 3.8. At lower redshifts, our field galaxy samples are in good agreement with measurements for optically selected redder galaxies from the VVDS surveys (Meneux et al., 2006) and also with the CFHTLS measurements presented in the previous chapter (McCracken et al., 2008). At higher redshifts, measurements for $pBzK$ and $sBzK$ galaxies are in approximate agreement with the measurements of Blanc et al. (2008). We note that part of the differences with the measurements of Blanc et al. arises from the their approximation of the redshift distribution of passive BzK galaxies using simple Gaussian. We note that a steep slope γ for optically-selected passive galaxies has already been reported at lower redshift surveys (Madgwick et al., 2003).

3.6 Summary

This chapter described counts, colours and clustering properties for a large sample of $K-$ selected galaxies in the 2 deg^2 COSMOS-WIRCam survey. This represents the largest sample of galaxies to date at this magnitude limit. By adding deep Subaru optical data we are able to reliably classify our catalogue into star-forming and quiescent/passive objects using the selection criterion

proposed by Daddi et al. (2004). Counts of field galaxies and star-forming galaxies clearly show a change in slope at $K_{\rm s} \sim 22$. Moreover, our number counts of quiescent galaxies clearly turnover at $K_{\rm s} \sim 22$ an effect which cannot be explained by incompleteness in any of our very deep optical bands. Our number counts of passive, star-forming and field galaxies agree well with surveys with brighter magnitude limits.

For simple magnitude-limited samples the Millennium simulation reproduces very well galaxy counts in the range $16 < K_{\rm AB} < 20$. However, at fainter magnitudes Kitzbichler & White (2007)'s models predict many more objects than are seen in our observations. Furthermore at $K_{\rm s} \sim 22$ the number density of faint passively evolving galaxies in our survey is well below the predictions semi-analytic models of Kitzbichler & White. This indicates that at $z \sim 2$ massive galaxies were already in place.

Cross-matching our catalogue with precise 30-band photometric redshifts calculated by Ilbert et al. we have derived the redshift distributions for each galaxy population. Our passive galaxies have a $z_{med} \sim 1.4$, in approximate agreement with similar spectroscopic surveys of smaller numbers of objects.

Turning to clustering properties, for a $K_{\rm s}-$ selected samples, the clustering amplitude declines in the monotonically towards fainter magnitudes. However, the slope of the best-fitting angular correlation function becomes progressively shallower at fainter magnitudes, an effect already seen in the COSMOS optical catalogues. At the faintest magnitudes, the field galaxy population is only slightly more clustered than the underlying dark matter distribution, indicating that $K_{\rm s}-$ selected samples are excellent tracers of the underlying mass. Star-forming and passive galaxy candidates are more clustered than the field galaxy population. At very small angular scales the passive BzK population is very strongly biased with respect to the dark matter distribution.

Using our photometric redshift distributions, we have derived the co-moving correlation length r_0 for each galaxy class. Fitting simultaneously for slope and amplitude we find a co-moving correlation length $r_0^{\gamma/1.8}$of $\sim 7h^{-1}$ Mpc for the passive BzK population and $\sim 5h^{-1}$ Mpc for the star-forming BzK galaxies at $K_{\rm s} < 22$. Our field galaxy clustering amplitudes are in approximate agreement with optically-selected red galaxies at lower redshifts.

3.7 These results in context

Today, in May 2011, one-and-a-half years since the results described in this Chapter were published, securing spectroscopic redshifts for passively evolving galaxies at $z \sim 2$ remains extremely challenging. The MORICS instrument on Subaru seems to be capable of measuring redshifts for only the very brightest objects ($K_{AB} \sim 18$); other promising instruments have yet to produce results. The addition of thinned red-sensitive CCD detectors to VIMOS on the VLT has enabled the collection of a small number of sources at the lower redshift end of the $1 < z < 2$ window, but higher redshifts remain unexplored. KMOS, a cooled near-infrared spectrograph for the VLT, may provide large samples of galaxies when it comes online in 2011. The most significant progress has come

through improved imaging capabilities. Although the data set described in this paper remains the deepest, largest near-infrared survey of the distant Universe published to date, the VISTA telescope has been successfully operating for the last one and a half years, and the first stacks and catalogues will soon become available. Reaching 1-2 magnitudes deeper than the catalogues described here, these data should provide a much better constraint on the faint end of the galaxy stellar mass function at $z \sim 2$.

Other than a simple comparison with semi-analytic models, the original work presented here made little attempt to place these results in a broader cosmological context. Our intention at the time was to apply our "'halo model" to the object classes outlined here and describe the results in a subsequent paper. While it was reasonably easy to fit clustering measurements of the passive galaxy population to our model, the "Star-forming" comparison sample proved extremely challenging, essentially because of the very broad redshift distribution these galaxies have (see Figure 3.4). The dark matter haloes hosting these galaxies undergo considerable changes between $z \sim 3$ and $z \sim 1$ and fitting a single model to this projected correlation function was problematic. A more promising approach to investigate the changing relationship between mass and light seems to be to consider purely mass-selected galaxy samples, and this is the approach we will adopt in the following hapter.

Chapter 4

Following the evolution of mass and light in the CFHTLS and COSMOS surveys

4.1 A phenomenological model for galaxy clustering

4.1.1 Characterising dark matter haloes

Once it had been demonstrated that predictions from model Universes dominated by cold dark matter were consistent with the observed clustering signal on scales of megaparsecs and larger (Davis et al., 1985), work turned to characterising the characteristics of the dark matter distribution in simulations. It was realised at an early stage that "haloes" of dark matter played an important role in galaxy formation (White & Rees, 1978) and for this reason it became crucial to determine their properties. A series of papers (Sheth & Tormen, 1999; Jenkins et al., 2001; Tinker et al., 2008) provided even more stringent tests of the Press & Schechter (1974) formalism which describes how the number of dark matter haloes depends on the halo mass, whilst others investigated in detail the density profiles of dark matter haloes and how this halo mass profile depends on redshift and cosmology (Navarro et al., 1997; Bullock et al., 2001). Other works investigated the clustering properties of dark matter haloes (Mo & White, 1996). The end result of this work was that by the end of the 1990s, a complete framework existed which could describe analytically the properties of dark matter haloes. Of course, only galaxies can be directly observed. The fundamental question which one attempts to address in constructing a theory of galaxy formation is how dark matter haloes are populated by galaxies and what are the underlying physical processes which drive star formation inside haloes. To do that requires a framework to relate visible galaxies to dark matter.

4.1.2 Relating haloes to galaxies

In "semi-analytic" theories of galaxy formation (Cole et al., 1994) (already discussed in the previous Chapter), starting from the White & Rees paradigm a set of analytic prescriptions are used to determine how galaxies "light up" dark

matter haloes. These rules are used to populate "merger trees" generated using numerical simulations which describe how the number and mass of dark matter haloes evolves with time. Such models (see Baugh (2006) for a recent review) can reproduce a wide variety of properties of the observed galaxy population. In practice, however, implementation is complicated; one must adjust the various galaxy scaling laws to correctly reproduce the observed statistical properties of the Universe at $z \sim 0$, such as the galaxy luminosity function, and there are potentially a large number of adjustable parameters. Alternatively, one can consider simulating the entire galaxy formation process by including both dark and baryonic matter in a large hydrodynamic simulation. While this approach is useful for investigating in detail how gas physics can affect galaxy formation, these simulations do not currently have sufficient mass resolution (due to the large computational requirements) to follow the formation and evolution of galaxies over the lifetime of the Universe.

The "halo model" represents a different, phenomenological to approach to modelling the clustering of galaxies. A distant ancestor of the current halo model can be found in the paper of Neyman & Scott (1952) which imagined galaxies as inhabiting "clumps" although it was not until papers such as those of Seljak (2000); Ma & Fry (2000); Peacock & Smith (2000); Scoccimarro et al. (2001) and others that the halo model in its current incarnation began to emerge. Earlier papers had addressed the clustering of dark matter haloes Mo & White (1996); in the halo model in the link is made with galaxies by making the major simplifying assumption that the number of galaxies in a given dark matter halo is a simple monotonic function of the halo mass. Another important assumption is that the halo occupation function is independent of environment and assembly history of the halo. Although some studies have shown that this "halo assembly bias" can affect galaxy clustering measurements, the amplitude of this effect is, for the moment, much smaller than the size of systematic errors (Croton et al., 2007).

As we shall see, this "halo occupation distribution" (HOD) can take a range of simple functional forms (suggested to some extent by numerical simulations). We assume that the observed correlation function of galaxies is in fact a sum of two terms: a contribution arising from pairs of galaxies inhabiting the same dark matter halo: the "one-halo-term", and a second term arising from pairs of galaxies in separate dark matter haloes (the "two halo term"). Until the end of the 1990s, almost all measurements showed that the galaxy correlation function closely matched a power law with a slope of $\delta \sim 0.8$. However, with the arrival of larger surveys such as the SDSS (Zehavi et al., 2004) it became possible for the first time to detect deviations from this simple power law shape and make a determination of the relative contributions of one-halo and two-halo terms, driving further development of the modelling. One reason why the distinctive one-halo / two halo transition had not been detected previously is that magnitude-limited samples are dominated by fainter, bluer galaxy populations for which the one-halo/two-halo transition is not very pronounced.

The form of the one-halo term depends essentially on the profile of the dark matter halo: the two halo term consists essentially a contribution from the linear part of the dark matter power spectrum; a detailed overview of the

halo model can be found in the recent review of Cooray & Sheth (2002). The transition from one-halo to two-halo regimes is governed by the size of the dark matter halo.

4.1.3 The stellar-mass halo-mass relationship

These halo models are not intended to replace traditional models of galaxy formation, but to complement them. Given galaxy clustering measurements and a determination of the galaxy number density, the halo model can provide (for example) an estimate of underlying dark matter haloes and the fraction of galaxies which are satellites. These quantities can be tracked as a function of redshift. Given a sufficiently large sample, one can determine the characteristics of the hosting haloes for a range of different galaxy populations. Since the physical processes driving galaxy formation and regulating star formation are expected to correlate with the mass of the dark matter haloes which host them, a profitable avenue to pursue in investigating the galaxy formation question is to determine the relationship between the mass in stars M_s and the dark matter halo mass M_h, otherwise known as the Stellar-mass halo-mass relationship (SHR). Thus, given stellar mass estimates, one can estimate the efficiency with which haloes form stars (Moster et al., 2010; Guo et al., 2010; Foucaud et al., 2010).

Studies of the SHR have shown that at $z < 0.1$ the fraction M_s/M_h reaches a maximum of a few percent at $10^{12} M_\odot$, much smaller than the universal baryon fraction of $\sim 20\%$. This mass scale represents the characteristic halo mass where star formation is most efficient. Recent works have suggested that this peak mass shifts to progressively higher masses at higher redshifts (Moster et al., 2010; Conroy et al., 2009). This is a phenomena similar in nature to "downsizing", in which the peak luminosity of galaxies undergoing star formation shifts to progressively lower luminosities at lower redshifts, and is a natural consequence of the fact that the characteristic stellar mass evolves more slowly than the characteristic halo mass. Furthermore, models suggest that the *amplitude* of this peak decreases at increasing redshift; in the model proposed by Conroy & Wechsler, the mass at which baryons are most efficiently converted into stars declines by a factor of 20 from $z \sim 0$ to $z \sim 1$.

In the following sections I will describe the halo model developed in collaboration with M. Kilbinger and J. Coupon and use it to interpret clustering measurements in two different data sets: the COSMOS survey (Scoville et al., 2007) and the CFHTLS. The halo model described here and its extensions to higher orders is expected to form in large part the basis for the doctoral thesis of M. Wolk under the direction of myself and Stephane Colombi. This Chapter contains a selection of results from two articles: Coupon et al., submitted, (Coupon et al., 2011), which considers measurements in the CFHTLS; and McCracken et al. (in preparation) which addresses the COSMOS data. These surveys are complementary; the CFHTLS provides excellent statistics for massive galaxies in luminosity-selected samples over the redshift range $0 < z < 1$; whereas with COSMOS it is possible to construct mass-limited samples reaching from $0 < z < 2$.

4.1.4 Introducing the halo model

Our version of the halo model closely follows the scheme outlined by Berlind & Weinberg (2002) and refined in later papers (Zheng et al., 2005; Tinker et al., 2005). Here we present a simplified outline of our model; full details can be found in appendix B of Coupon et al. (2011). We start by noting that the galaxy correlation function $\xi(r)$ can be expressed as the sum of correlation function arising from pairs in different haloes and from pairs in the same halo:

$$\xi_{gg}(r) = [\xi_{gg}^{1h}(r) + 1] + \xi_{gg}^{2h}(r) \tag{4.1}$$

The one-halo term can be expressed as follows (Kravtsov et al., 2004):

$$1 + \xi_{gg}^{1h}(r) = \frac{1}{2}\bar{n}_g^{-2} \int n(M)\langle(N(N-1)\rangle_M \lambda(r|M) dM \tag{4.2}$$

Similarly, the two halo term computed between two haloes of mass M_1 and M_2 can be expressed as

$$\begin{aligned} \xi_{gg}^{2h} = \xi_{dm}^{lin}(r)\bar{n}^{-2} \int n(M_1) b_h(M_1) \langle N\rangle_{M_1} dM_1 \\ \times \int n(M_2) b_h(M_2) \langle N\rangle_{M_2} \lambda(r|M_1 M_2) dM_2 \end{aligned} \tag{4.3}$$

Here $\xi_{dm}^{lin}(r)$ represents the linear dark matter power spectrum, $\langle N\rangle$ the mean halo occupation and $b_h(M)$ the large scale halo bias. $\bar{n}_g^{-2}$ represents the mean galaxy density and $\lambda(r|M)$ is the convolution of radial profile of galaxies with itself and $\lambda(r|M_1 M_2)$ the convolution of the radial profile of two different galaxies. $n(M)$ represents the mass function of haloes and $\langle(N(N-1)\rangle_M$ the mean number halo pairs.

At large scales, the two halo correlation function reduces to $\xi(r) \sim b^2\xi(r)$. At these scales, much larger than the virial radius, all pairs consist of galaxies in separate haloes, while at smaller scales, most pairs consist of galaxies with the same haloes.

Considerations from both semi-analytic models (Benson et al., 2000) and SPH simulations (Berlind & Weinberg, 2002; Kravtsov et al., 2004) indicate that the galaxy population can be separated into two classes: massive "central" galaxies at the cores of dark matter haloes, which follow the large-scale distribution of dark matter, and "satellite" galaxies which trace the profile of dark matter haloes. The average number of central galaxies, $\langle N_c\rangle_M$ can be modelled as a simple step function with $\langle N_c\rangle_M \sim 1$ for haloes more massive than a minimum halo mass M_{min}, and $\langle N_c\rangle_M = 0$ otherwise. For the satellite galaxy population, the average number of galaxies is well approximated by simple power law of the halo mass $\langle N_c\rangle_M \propto M^\alpha$.

In this work, we model the contributions between the number of satellite galaxies N_c satellite and central galaxies N_s as follows:

$$N(M) = N_{\rm c}(M) + N_{\rm s}(M)\,, \tag{4.4}$$

where

$$N_{\rm c}(M) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{4.5}$$

$$N_{\rm s}(M) = N_{\rm c}(M) \times \left(\frac{M - M_0}{M_1}\right)^{\alpha} \tag{4.6}$$

For an assumed cosmology and dark matter halo profile, our model has five adjustable parameters, $M_{\rm min}, M_1, M_0, \alpha$ and $\sigma_{\log M}$.

Each dark matter halo is modeled as a Navarro, Frenk and White density profile (Navarro et al., 1997). The angular two-point correlation function $w(\theta)$ is computed from the observed photometric redshift distribution and $\xi(r)$ using Limber's equation (Limber, 1954). We use the mass function $n(M, z)$ from Sheth et al. (2001), and we compute the non-linear power spectrum using the (Smith et al., 2003) prescription.

Finally, the number density of galaxies at a given redshift z can be computed as follows:

$$n_{\rm gal}(z) = \int_{M_{\rm low}}^{M_{\rm high}} N(M) n(M, z)\, \mathrm{d}M\,. \tag{4.7}$$

4.1.5 Deduced parameters

From the HOD model we obtain the deduced parameters describing halo properties. The mean galaxy bias $b_{\rm g}$ at redshift z is the mass integral over the halo bias weighted by the number of galaxies,

$$b_{\rm g}(z) = \int_{M_{\rm low}}^{M_{\rm high}} \mathrm{d}M\, b_{\rm h}(M)\, n(M, z) \frac{N(M)}{n_{\rm gal}}(z)\,. \tag{4.8}$$

Similarly, the mean halo mass for a galaxy population is

$$M_{\rm h}(z) = \int_{M_{\rm low}}^{M_{\rm high}} \mathrm{d}M\, M\, n(M, z) \frac{N(M)}{n_{\rm gal}}(z)\,. \tag{4.9}$$

The fraction of central galaxies per halo is

$$f_{\rm c}(z) = \int_{M_{\rm low}}^{M_{\rm high}} \mathrm{d}M\, n(M, z) \frac{N_{\rm c}(M)}{n_{\rm gal}}(z)\,. \tag{4.10}$$

Consequently, the fraction of satellite galaxies is

$$f_{\rm s}(z) = 1 - f_{\rm c}(z)\,. \tag{4.11}$$

4.1.6 Parameter estimation

We use the "Population Monte Carlo" method (PMC) to sample the likelihood space (Wraith et al., 2009), in our implementation `CosmoPMC`[1]. Contrary to the

[1] `http://cosmopmc.info`

widely-used Monte Carlo Markov Chain method, this technique is an adaptive importance-sampling method (Cappé et al., 2004, 2008). The evaluation of the likelihood function, which is the most time-consuming process for most sampling tasks, is easily parallelized in importance sampling with little overhead. Our HOD model is heavily optimised, and the computation of the angular correlation function w in fifteen angular bins is performed in under a second on a desktop computer. The PMC code is run in a multithreaded computing environment using several hundred processors.

For each galaxy sample we simultaneously fit both the projected angular correlation function w and the number density of galaxies $n_{\rm gal}$ by summing both contributions in log-likelihood:

$$\begin{aligned}\chi^2 &= \sum_{i,j}\left[w^{\rm obs}(\theta_i)-w^{\rm model}(\theta_i)\right]\left(C^{-1}\right)_{ij}\left[w^{\rm obs}(\theta_j)-w^{\rm model}(\theta_j)\right] \\ &\quad + \frac{\left[n_{\rm gal}^{\rm obs}-n_{\rm gal}^{\rm model}\right]^2}{\sigma^2_{n_{\rm gal}}}\,. \end{aligned} \tag{4.12}$$

The data covariance matrix C is estimated from jack-knife error estimates. The error on the galaxy number density $\sigma_{n_{\rm gal}}$ contains Poisson noise and cosmic variance.

4.2 Measurements in CFHTLS and COSMOS

Our objective is to understand how the relationship between dark matter and luminous galaxies evolves over the largest possible range of cosmic time. To do this, we will use two separate surveys, the CFHTLS and COSMOS.

4.2.1 Survey characteristics

The COSMOS survey has already been described in Chapter 3. It represents one of the best datasets currently available to study the evolution of clustering in mass-selected samples thanks to its large contiguous field of view (2 deg^2) and multi-wavelength coverage. This makes it possible to compute accurate photometric redshifts from $0 < z < 6$ with low numbers of catastrophic outliers. The addition of deep near-infrared data, described in the previous chapter, means we are able we are able to compute accurate stellar masses.

We will also use data from the CFHTLS wide survey. Covering an effective area of 139 deg^2 in four separate fields of five bands, it is currently the only wide-field survey reaching $z \sim 1$. The CFHTLS remains unmatched in terms of its unique combination of depth, area, wavelength coverage and image quality. Even future, deeper surveys soon to start such as DES will mostly be conducted using red-sensitive detectors, thus missing the crucial near-ultraviolet coverage essential to calculate accurate photometric redshifts over the full redshift range $0 < z < 1.2$. Only CFHTLS has this, with photometric redshifts accurate to $\sim 3\%$ and free from systematic errors to $0 < z < 1$ (Coupon et al., 2009).

Because CFHTLS covers a much larger area than COSMOS one is able to

find many more rarer, more massive haloes and investigate in much greater detail the distribution of galaxies inside dark matter haloes as a function of their average intrinsic properties such as absolute luminosity or rest-frame colour. Although CFHTLS reaches shallower absolute luminosities than COSMOS (and has a median redshift is around $z \sim 0.8$), the much larger area means that in total CFHTLS contains many more objects; to our magnitude limit of $i_{AB} < 22.5$ there are more than 3 million galaxies. Thanks to this, the CFHTLS provides one of the most accurate pictures available of the universe in the redshift range $0 < z < 1$. The survey currently however has one crucial deficiency: the current lack of near-infrared data means that it is not possibly to accurately construct mass-selected samples. In this chapter we will attempt to empirically relate mass to luminosity, but without near-infrared data this remains at best an approximation. Several projects are underway to add near-infrared data to the CFHTLS-wide.

4.2.2 COSMOS and CFHTLS photometric redshifts

The COSMOS photometric catalogues are identical to those used in the previous chapter, with the difference that new $H-$band data has been added, which improves the photometric redshift accuracy in the range $1 < z < 2$ and reduces the numbers of catastrophic outliers. The CFHTLS-wide photometric redshifts are computed based on five broad-band colours following the techniques outlined in Coupon et al. (2009); Ilbert et al. (2006). We use the same template set selection as in the CFHTLS-deep in Chapter 2.

For both surveys, our approach is similar: the best fitting photometric redshift is determined by comparing predicted galaxy colours computed using a set of galaxy templates at a range of redshifts with observed broad-band galaxy colours. However, differences in depths and wavelength coverage between COSMOS and CFHTLS means that the same set of templates are not appropriate for both data sets; for the sparser wavelength sampling of the CFHTLS survey, a narrower range templates provides more accurate photometric redshifts with lower numbers of catastrophic failures. For the dense wavelength coverage of the COSMOS survey containing many intermediate band images as well as deep near-infrared data, a much larger set of templates are used, comprising both empirical and synthetic templates. For full details, see Ilbert et al. (2010). Our typical photometric redshift accuracy, $\sigma_{\Delta_z}/(1+z_{\rm s})$, denoted by the normalised absolute median distribution, is $1.48 \times median(|z_{\rm p} - z_{\rm s}|/(1+z_{\rm s}))$, where $z_{\rm p}$ and $z_{\rm s}$ denotes spectroscopic and photometric redshifts respectively.

4.2.3 Sample construction

For COSMOS, we construct a series of volume-limited samples selected by stellar mass. The left panel of Figure 4.6 shows stellar mass as a function of redshift; the box represents our completeness limit. For the CFHTLS wide, we extract a series of volume-limited samples selected by luminosity and color, in a similar fashion to Chapter 2. The right panel of Figure 4.6 shows the distribution of

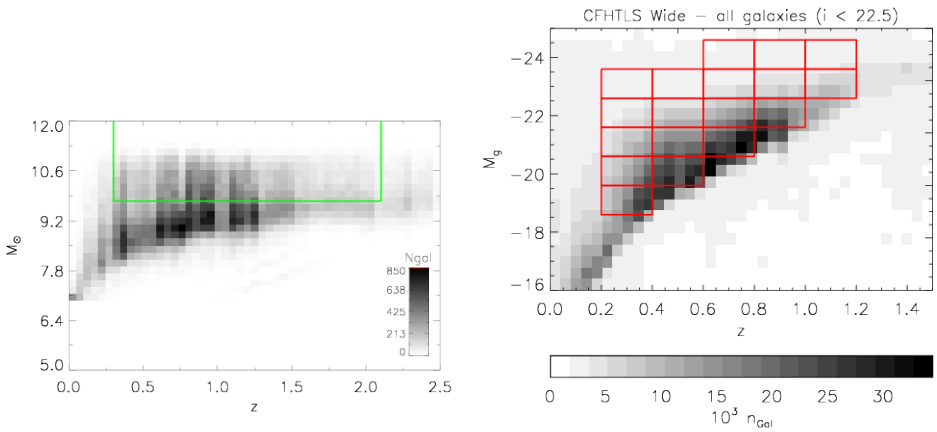

Figure 4.1 Left panel: Stellar mass-redshift plane for the COSMOS data. The box represents the redshift and mass limits used this paper. Inside this box our stellar completeness is 90% or greater. Right panel: luminosity-redshift plane for the CFHTLS-wide data. The series of boxes represents each of the mass-selected volume-limited samples used.

galaxies in the luminosity-redshift plane.

In this luminosity-redshift plane, objects are selected with $i_{AB} < 22.5$ and divided into five redshift bins: $0.2 < z < 0.4$, $0.4 < z < 0.6$, $0.6 < z < 0.8$, $0.8 < z < 1.0$, and $1.0 < z < 1.2$. This large bin width ($\Delta_z = 0.2$) minimises the amount of bin-to-bin contamination due to photometric redshift errors ($\sigma_z < 0.088$). Inside each volume-limited slice we select galaxies by best fitting template, dividing objects in "red" and "blue". This sample selection is shown in Figure 4.2.

4.2.4 Measured angular correlation functions

In Figures 4.3 and 4.4 we show mass-selected angular correlations in four redshift slices in the COSMOS field and luminosity-selected correlation functions for one redshift slice in the CFHTLS wide surveys. The solid line shows the best fitting halo model for each sample. The angular correlation function w is measured in a series of logarithmically separated bins using the Landy & Szalay (1993) estimator as in previous chapters. For the CFHTLS fields, we compute covariance matrices for each galaxy sample using jack-knife error estimates over all fields; for the COSMOS data, we measure jack-knife errors inside the COSMOS field itself by systematically removing one (different) part of the field. This most likely under-estimates the error on large scales, but until larger fields or numerical simulations with sufficient resolution become available, this represents an acceptable compromise.

For the CFHTLS wide samples, the halo model provides a good fit to the

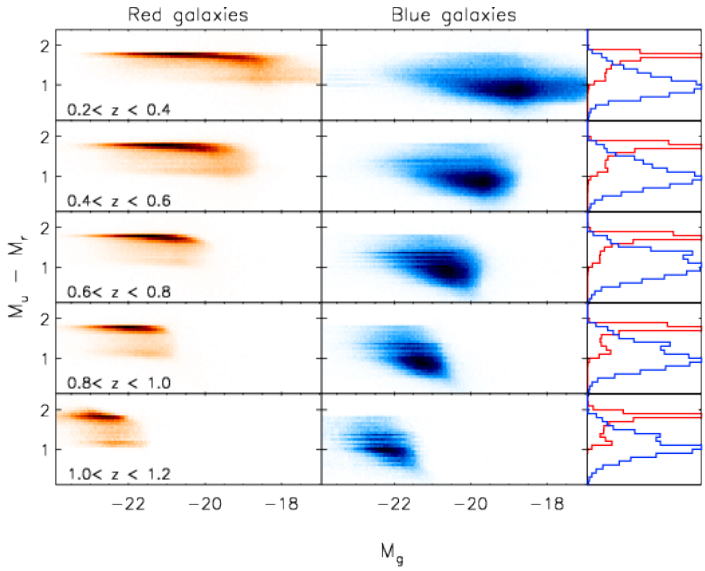

Figure 4.2 Type selection based on best fitting templates for the "red" sample (left) and the "blue" sample (right) in the CFHTLS Wide. The right panel shows the $(M_u - M_r)$ colour as a function of absolute magnitude (M_g) and redshift (from top to bottom) and number counts (right panels) for "red" and "blue" samples.

observations in the three absolute luminosity bins shown here, with χ^2 measurements of $\sim 1-4$ for the full galaxy sample. The right panel of Figure 4.3 shows that the minimum halo mass increases steadily with increasing luminosity threshold. Furthermore, more luminous samples are more dominated by satellite galaxies (the area under the curve is lower for $(N(M)<1)$.

For the COSMOS data, at low redshifts and at scales of several tens of arcminutes, the fit is poor and (at these scales the jack-knife derived error bars are large and so the model is not well constrained). The origin of the poor fit on large scales comes from the presence of several rich clusters at $z \sim 0.7$ in the COSMOS field. This is already evident in Meneux et al. (2009) who showed that at these intermediate redshifts, the COSMOS field is considerably more clustered compared to comparison fields such as the VVDS-02hr survey field (Fèvre et al., 2005b). Note that the amplitude of galaxy clustering increases slowly with redshift; given the fact that dark matter clustering evolves differently a natural consequence of this is that the overall galaxy bias was lower at the present day than it was higher redshifts.

4.2.5 Characteristic mass scales for CFHTLS and COSMOS

For each sample in COSMOS and CFHTLS we compute the five best-fitting halo-model parameters. For the CFHTLS data however we face an additional complication: our samples are selected by absolute luminosity, and not by stellar mass. To make an approximate conversion between average mass-to-light ratio and redshift, we use a relation derived from COSMOS where precise stellar

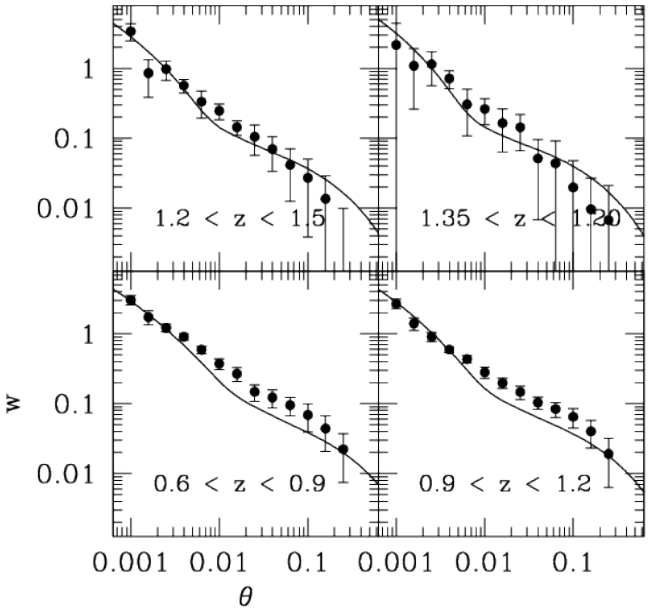

Figure 4.3 Correlation amplitude w as a function of angular scale for four slices in redshift and with stellar mass $M < 10^{10}$ M$\odot$ in the COSMOS field. The solid line shows the best fitting halo model.

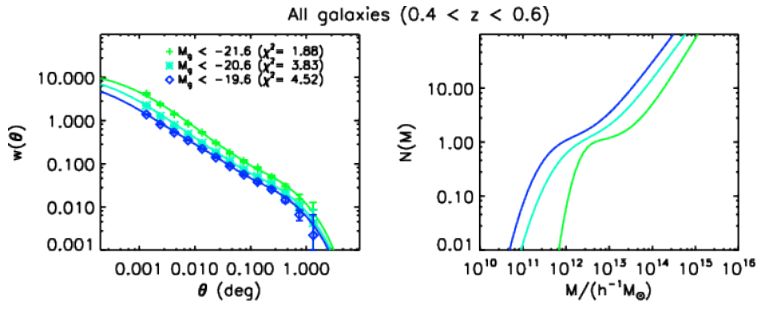

Figure 4.4 Correlation amplitudes w as a function of angular scale for three slices in absolute magnitude for the full galaxy sample in the CFHTLS-wide. The solid lines show the best-fitting halo model. The right panel shows the corresponding best-fitting halo occupation distribution for these slices.

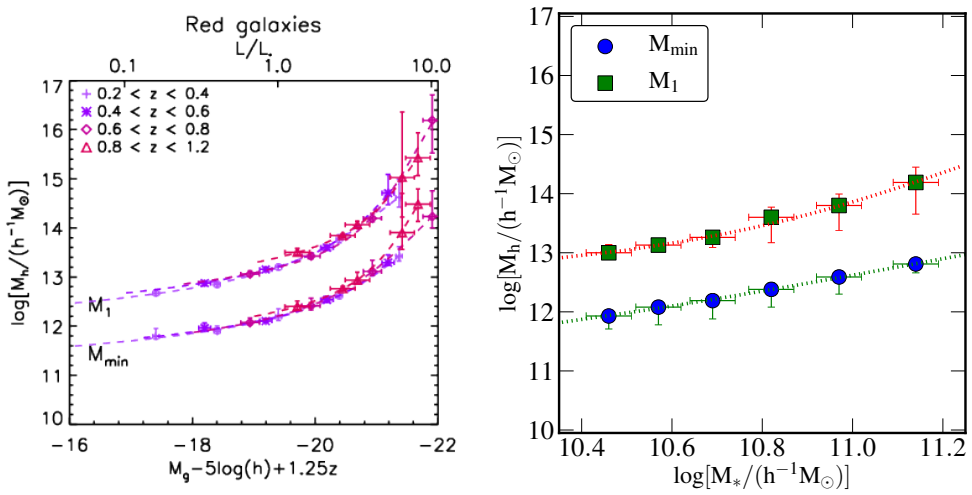

Figure 4.5 Left panel: halo masses M_1 and $M_{\rm min}$ as a function of corrected luminosity threshold for CFHTLS galaxies (all redshifts). Right panel: halo masses as a function of stellar mass for COSMOS galaxies at $z \sim 1$.

masses are available. With the five-band data CFHTLS wide, this relation works well for passive galaxies. For star-forming galaxies, however, the slope of the relation depends on stellar mass, which is unknown. For this reason, here we consider primarily the CFHTLS red galaxy population; for further details, see Coupon et al..

For most samples in both surveys, $\sigma_{\log\ M}$ and M_0, the satellite cut-off, are poorly constrained. We find that the high-mass slope of the halo occupation function, α, is ~ 1, consistent with the expectations from numerical simulations, For very bright samples there is some evidence that α is larger than unity.

The right panel of Figure 4.5 show the best-fitting minimum halo mass $M_{\rm min}$ as a stellar mass for a range of mass thresholds in the COSMOS survey; the left panel shows a similar plot for the CFHTLS selected by threshold luminosity. As we have already seen, on average samples with higher stellar mass thresholds have higher mean clustering amplitudes. A consequence of this is that more clustered populations are hosted by galaxies with higher minimum halo masses; rarer populations are more strongly clustered. We note that at the same mass threshold and redshift range, the measurements from both surveys are consistent. Because the COSMOS survey covers a much smaller area than the CFHTLS the range of effective mass scales it can probe is correspondingly reduced.

In our parametrisation, $M_{\rm min}$ corresponds to the mass for which 50% of galaxies in our sample contain one galaxy. Similarly, M_1 corresponds to the mass required to host at least two galaxies in a given halo. In threshold-limited luminosity samples, such as we have for the CFHTLS, $M_{\rm min}$ corresponds approximately to the luminosity of the central galaxy (Zheng et al., 2005) (or to the stellar mass of the central galaxy, in the case of a mass-limited sample).

How does $M_{\rm min}$ depend on luminosity and halo mass? The curves shown in Figure 4.5 represents the analytic form proposed by Zehavi et al. (2010),

$$L_{\rm cen}/L_* = A\left(\frac{M_h}{M_t}\right)^{\alpha_M} \exp\left(-\frac{M_t}{M_h}+1\right), \tag{4.13}$$

with A, M_t and α_M as free parameters, giving a power-law dependence on halo mass at the high-mass end and an exponential decay at the low-mass end. This expression encapsulates the idea that there exists a "transition halo mass" M_t at which the formation of stars is most efficient, where baryons are most effectively converted into stars, an idea suggested by several different lines of approach, for example from abundance matching measurements (Conroy & Wechsler, 2009; Guo et al., 2010; Moster et al., 2010). In the CFHTLS we fit this relationship in four redshift slices in the range $0.2 < z < 1.2$, shown by the dotted lines in the left panel of Figure 4.5. At $0.2 < z < 0.4$ our best fitting values are $A \sim 0.3$, $\log M_t \sim 11.4$ and $\alpha_M \sim 0.4$ for $M_{\rm min}$, consistent with measurements in the local universe in Zehavi et al.. The parameter M_1 can be thought of as representing the characteristic halo mass required to host two galaxies (the "satellite threshold") and closely follows $M_{\rm min}$ times a constant factor; in general haloes must be more massive in order to host satellites galaxies. Galaxies in the "hosting gap" mass range tend to contain a single, more massive central galaxy (Zehavi et al.).

It is also interesting to consider the redshift evolution of $M_1/M_{\rm min}$. In our CFHTLS $0.2 < z < 0.4$ slice we find (for our full galaxy sample) $M_1/M_{\rm min} \sim 15$, similar to SDSS values. For the COSMOS data at $z \sim 1$, we find that at lower mass thresholds, $\log(\mathrm{M}_*) \sim 10.4$, $M_1/M_{\rm min} \sim 11$; at higher stellar mass thresholds, this increases to ~ 22. Moreover, M_1 and $M_{\rm min}$, measured as the ratio between the different transition masses, decreases towards higher redshifts. This is because M_1 increases more slowly than $M_{\rm min}$. Stated another way, the differences in masses between haloes containing only one galaxy and containing several galaxies becomes smaller at higher redshifts. One possible interpretation of this phenomena is that haloes at higher redshifts may have more recently accreted satellites, a phenomena observed in numerical simulations (Zentner et al., 2005).

In the right panel of panel of Figure 4.5, we now consider similar measurements for $M_{\rm min}$ in the COSMOS survey, this time selected by stellar mass in the redshift range $0.4 < z < 1.7$. Firstly, we note that at the same mass threshold ($M_* < 10.0$ corresponds to $M_g \sim$) we find approximately the same minimum halo masses in both surveys. Secondly, galaxies selected in higher mass thresholds are more strongly clustered. Finally, in $0.4 < z < 1.1$ there is only a weak dependence of $M_{\rm min}$ on redshift, consistent with the results from the CFHTLS. However, beyond $z > 1.1$, $M_{\rm min}$ increases dramatically. Abundance matching arguments suggest that this is a consequence of the rapid decline in the normalisation of the mass function between $1 < z < 2$; our measurements of $M_{\rm min}$ mirrors the rapid build-up in stellar mass observed between $1 < z < 2$.

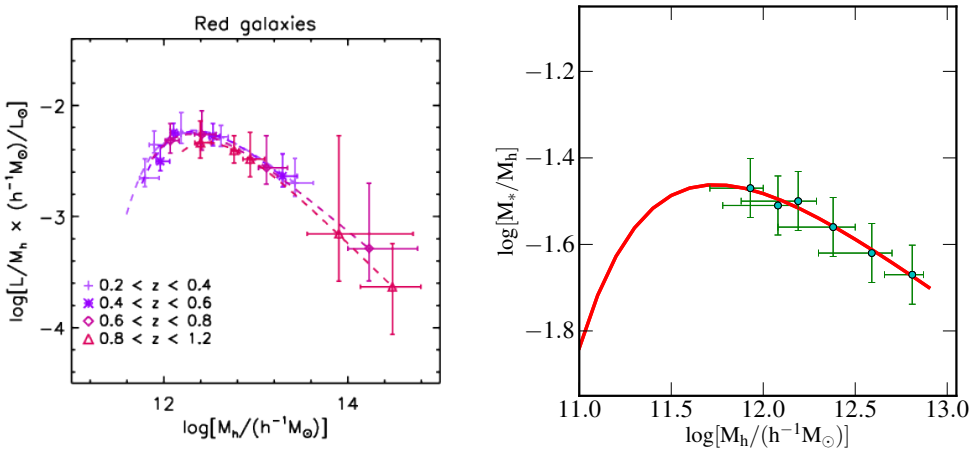

Figure 4.6 The ratio between corrected luminosity and halo mass as a function of halo mass for red galaxies in the CFHTLS data (left panel) and for COSMOS galaxies at $z \sim 1$ (right panel). For the CFHTLS galaxies, we have used an approximate relationship between mass and luminosity to convert our luminosity-selected samples to mass-selected ones.

4.2.6 The mass-to-light ratio of galaxies as a function of halo mass

We can now extend this analysis and consider the ratio of the mass in stars M_* to the halo mass M_h (in this case, the mass of the central galaxy) as a function of halo mass M_h. For the "corrected" luminosity threshold samples of the CFHTLS, this is found by rearranging Equation (4.9) to give

$$\frac{L'_{\rm c}}{L_*} = A\left(\frac{M_{\rm h}}{M_{\rm t}}\right)^{\alpha_M} \exp\left(-\frac{M_{\rm t}}{M_{\rm h}} + 1\right) . \tag{4.14}$$

We can also apply this relationship to the COSMOS data where we are able to measure the stellar mass for each sample. In Figure 4.6 we show results of these fits in CFHTLS for red galaxies over the entire redshift range and for COSMOS at $z \sim 1$ (left and right panels respectively). The CFHTLS measurements clearly constrain the position and amplitude of the peak. The right panel shows that the COSMOS measurements, on the other hand at $z \sim 1$ do not strongly constrain the position of the peak.

The real power in both these surveys lies in our ability to investigate the redshift dependence of the halo model parameters. In Figure 4.7 we show the position of the peak transition mass M_t, derived from fits to Equation (4.13) for a series of redshift slices, for the CFHTLS full sample (not described here), the CFHTLS red galaxy sample and the COSMOS mass-selected sample. In addition we present a compilation of results from the literature. At higher redshifts, our CFHTLS measurements are in agreement with those of Behroozi

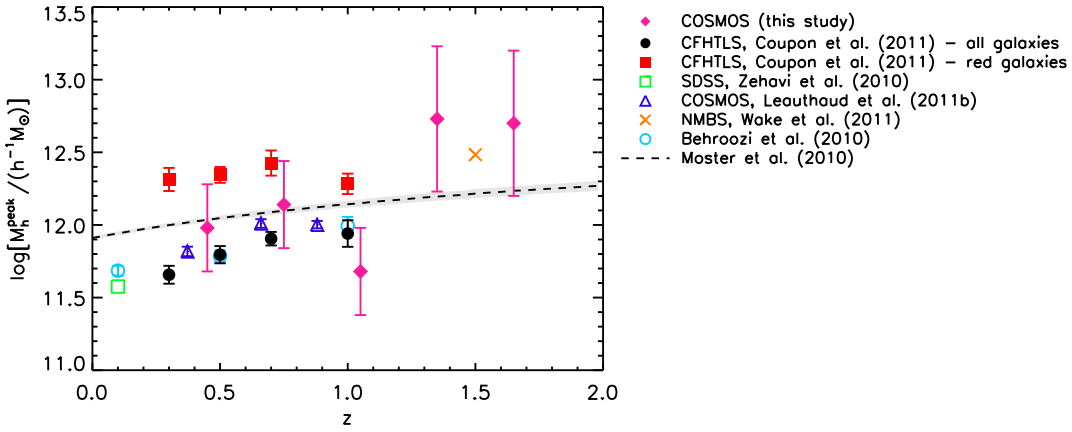

Figure 4.7 Position of the peak $M_{\rm h}^{\rm peak}$ corresponding to $\max(L_{\rm c}'/M_{\rm min})$ for CFHTLS measurements. The dashed line represents results from Moster et al. (2010), where the shaded region shows the $1-\sigma$ uncertainty.

et al. (2010) who used an abundance matching technique to connect dark matter halo mass functions from N-body simulations with observed galaxy stellar mass functions. Both Moster et al. and Behroozi et al.'s abundance matching measurements are based primarily on galaxy stellar mass measurements covering relatively small areas (a few hundred square arcminutes). The CFHTLS full sample differs significantly from Moster et al., although in both cases the two data sets show the same general trend: $M_{\rm h}^{peak}$ shifts gradually towards higher halo masses at higher redshift. In fact, the Moster et al. points do not agree with most other observations and moreover overestimates the position of the peak at low redshift. The measurements from COSMOS, shown as the filled diamonds, are consistent with the CFHTLS at low redshift and at higher redshifts with those of Wake et al. (2010). The measurements from Leauthaud et al. (2011), computed using a joint lensing, abundance and clustering analysis of the COSMOS field, seem to slightly over-estimate the position M_t compared our CFHTLS measurements; one possible explanation of this discrepancy is that the COSMOS has a number of rich structures at $z \sim 0.7$ (Meneux et al., 2009). The measurements from Leauthaud et al. are broadly consistent with our clustering-only determinations.

For the passive galaxy population in the CFHTLS M_t and $M_{\rm h}$ depend only weakly on redshift: measurements of $M_{\rm min}$ at $z \sim 1.0$ we find $\log M_t = 12.4 \pm 0.1$, compared to $\log M_t = 12.1 \pm 0.1$ at $z \sim 1.0$. This is simply because, to first order, central galaxies dominate the total number density for any luminosity; in addition to matching correlation function measurements, halo abundances must match the observed stellar mass function. At $z < 0$, the normalisation ϕ^* of the stellar mass function remains approximately constant (see, for example, Figure 13 of Ilbert et al. (2010)).

The CFHTLS results summarised here represent the first time a single data set covering a statistically significant area has been used to derive the evolution of the peak position as a function of redshift and galaxy type in a self-consistent

manner. The peak position, $M_{\rm h}^{\rm peak}$, can be interpreted as representing the halo mass at which the stellar mass content most efficiently accumulates in haloes, either by star formation or merger processes. The movement of this peak towards higher halo masses at higher redshifts is consistent with a picture in which the sites of efficient stellar mass growth migrate from low-mass to high-mass haloes at higher redshift. This "anti-hierarchical" evolution, frequently referred to as "halo downsizing", resembles closely the scenario first sketched by Cowie et al. (1996), in which the maximum luminosity of galaxies undergoing star-formation declines steadily to the current day. Conversely, the fact that we observe constant transition masses for the red population is also a consequence of the fact that in these more massive haloes, on-going star-formation and mass growth has ceased. These galaxies have undergone essentially passive evolution since $z \sim 1.2$.

Our COSMOS measurements push these measurements to higher redshifts and shows that at $z > 1$ the peak position moves to even higher halo masses. However given the current depth of the near-infrared data in the COSMOS field these results should be treated with some caution. Definitive measurements will be provided once the new, deeper UltraVISTA data is added to the COSMOS photometry.

4.3 Summary of principal results

In this Chapter we have presented a selection of results from our comparison of clustering measurements in the CFHTLS wide and COSMOS surveys with the predictions of an analytic halo model. Some of the main conclusions we draw from these studies are as follows:

1. Clustering strength increases with increasing luminosity or mass threshold, reflecting that bright (or massive) galaxies reside in more strongly clustered massive haloes.

2. At a fixed mass or corrected luminosity threshold, galaxy clustering amplitude increases with increasing redshift. Since the amplitude of dark matter clustering was lower at higher redshifts, this means that the galaxy bias increases steadily towards higher redshifts.

3. By assuming a simple analytic relationship between central galaxy luminosity and halo mass and fitting this to our observations, we find a maximum in the stellar-to-halo mass ratio in the CFHTLS at $z \sim 0.3$, $M_{\rm h}^{\rm peak} = 4.5 \times 10^{11} h^{-1} M_{\odot}$ and $M_{\rm h}^{\rm peak} = 21 \times 10^{11} h^{-1} M_{\odot}$ for the full and red samples, respectively. This transition mass represents the halo mass where baryons were most efficiently converted into stars.

4. In the CFHTLS full sample, $M_{\rm h}^{\rm peak}$ shifts to higher halo masses at higher redshifts. For the red galaxy sample, the peak position evolves less rapidly with redshift. These results can be understood qualitatively from the lack of on-going star-formation in the red galaxy population which means that the stellar mass content in these massive haloes changes very slowly. For

the full galaxy sample, which is expected to contain galaxies still undergoing active star-formation, the shift of the transition mass to progressively more massive haloes at higher redshifts is a manifestation of "anti-hierarchical" galaxy formation. These results also consistent with a scenario in which massive, passive galaxies already fully formed by $z \sim 1.2$. Preliminary results from the COSMOS survey indicate that at $z \sim 2$, $M_{\rm h}^{\rm peak}$ moves to even higher effective halo masses.

4.4 These results in context

It is interesting to consider the results presented here in the context of Chapters 2 and 3. In Chapters 2, we demonstrated how photometric redshifts could be used to make reliable measurement of the dependence of galaxy clustering on luminosity, redshift and colour. In Chapter 3 we showed how near-infrared data makes it possible to extend these measurements to $z \sim 2$ using colour selection techniques. In this Chapter, we have expanded the measurements of Chapter 2 to cover more than three million galaxies in the CFHTLS-wide survey and to mass-selected samples in COSMOS. At $z < 1$, these samples allow us to make precise measurements of the dependence of galaxy clustering on luminosity, colour and redshift. In COSMOS, we can push our measurements to beyond redshift one.

At a fixed redshift, samples selected with brighter luminosity thresholds are more strongly clustered than those with lower luminosity thresholds. An empirical relationship, calibrated in the SDSS survey, is consistent with our observed dependencies of halo mass or clustering strength on luminosity. This relationship follows closely the results shown in Chapter 2 which show that at $L < L^*$ clustering strength depends only weakly on luminosity. For $L > L^*$, clustering strength rapidly increases with threshold sample. Measurements based on mass-selected samples in the COSMOS field confirm this behaviour.

The amplitude of galaxy clustering for a fixed luminosity-threshold sample increases from $z \sim 0.7$ to $z \sim 0.1$, an effect clearly seen in the CFHTLS-wide and also visible in the CFHTLS-deep surveys for the full galaxy samples. Considering the results from our halo model fitting, this increase is simply a consequence of the overall growth of galaxy haloes over cosmic time.

Measurements in the CFHTLS wide confirm the profound difference in clustering properties between red and the full galaxy populations. In general, at the same threshold in absolute luminosity, galaxies with redder rest-frame colours are more clustered than bluer galaxies. With the addition of our analytic halo model, we are able to understand these differences as a simply a consequence of the different dark matter haloes these galaxy populations inhabit. Moreover, from $z \sim 1$ to $z \sim 0$, the observed changes in the clustering properties of the field galaxy population are largely consistent with models in which the parameters of the halo occupation distribution do not change with redshift. Stated another way, at fixed number density $n_{\rm gal}$, the minimum halo mass $M_{\rm min}$ is constant with redshift. These results are discussed in more detail in Coupon et al., but one qualitative conclusion one can draw is that this is evidence for

the limited effect of mergers in the galaxy population at $z < 1$.

Chapter 5

The Future

My masters' thesis, completed in 1995 under the direction of Chris Prichet in Victoria, made use of probably one of the last deep galaxy surveys derived from photographic plate data. At the time sliver halide still had an edge compared to silicon: electronic detectors were restricted to fields of only a few arcminutes on a side. Writing the obligatory "conclusions and future prospects" back then was easy: in this future, we would obviously use wide-field CCD cameras to collect multi-colour data over several square degrees of sky. I had heard about the Sloan digital sky survey, which was still in planning at the time, but work was already underway to pave the prime-focus of the Canada-France Hawaii telescope with silicon, starting with FOCAM, followed in quick succession by the temperamental UH8K camera, CFH12K and now MegaCAM which has been in operation for the last half-decade. MegaCam's crown since been displaced by newer giga-pixel cameras like DeCAM for DES and the Pan-Starrs camera, currently the largest camera on the sky. (The CFH12K camera is still in service and is a central asset of the Palomar Transit Survey.) The progression from single-detector CCDs to wide-field instruments like MegaCAM has required an enormous effort in software development and quality control in order to tame systematic errors. The final release of the CFHTLS, which will be available at the end of 2011, will have an absolute photometric precision over the entire survey of better than 1%. More precise photometric redshifts and well-controlled systematic errors will in all likelihood enable new science, for example making it possible to measure higher orders of the galaxy correlation function out to $z \sim 1$ which could potentially an independent measurement of the evolution of galaxy bias.

A similar progression has taken place in near-infrared detector technology, whose pixel counts lag optical detectors by a decade or so. As part of my thesis project in Durham, surveying the 50 arcmin^2 Herschel Deep Field at the United Kingdom infrared telescope to relatively shallow magnitudes using a 256×256 near-infrared array took more than a few trips up and down the mountain (five night in total). With the arrival of WIRCam at CFHT, four 2048×2048 detectors covering a few hundred arcminutes, meant than one could think about covering the entire 2 deg^2 COSMOS field in a few tens of hours. And very recently, the past year of VISTA data on the COSMOS field, reduced at

TERAPIX, covers the entire COSMOS field to between two or three magnitudes fainter than the COSMOS/WIRCAM data set. VIRCAM, which is an array of 16 2048×2048 detectors, is around three times more efficient than WIRCAM, and is mounted on a telescope dedicated to near-infrared surveys, VISTA. The VISTA public surveys, in conjunction with planned or future optical surveys, will enable us to accumulate very large mass-selected surveys covering hundreds of square degrees and reaching out to $z \sim 2$. I expect the VISTA surveys to occupy a considerable fraction of my time in the coming years.

It's worth reflecting for an instant on this continuous and relentless growth in size and fidelity of our digital surveys of the distant Universe. Astronomy, a science at the frontiers of human knowledge, benefits immensely from the vast capital investments which have been made in semiconductors and the profound advances in solid-state physics over the past century. Today, constructing a single factory to produce a current-generation microprocessor costs billions of dollars; such a vast expenditure would never be made with the unique objective of assuring a means for astronomers to reduce data from array cameras. We are driven by a powerful combination of financial imperatives and our innate curiosity to always consider the next step, the next advance, to always find a new way of doing things better. For centuries in China, the accuracy of water-clocks remained the same, because people considered that the current models were good enough for the task in hand. This attitude is almost completely alien to us today, where continual progress renders everything, from galaxy surveys to mobile telephones, obsolete within a few years. So yes, it is easy to make the prediction that there will be bigger and larger galaxy surveys, with larger wavelength coverage and reacher to deeper depths. But will these technological advances be matched by corresponding theoretical advances capable of fully exploiting these results?

In the last decade, the same technological progress which has driven advances in galaxy surveys have made possible computer models capable of simulating large volumes of the Universe at high mass resolution. These simulations can follow numerically the evolution of galaxies inside dark matter haloes. Of course, if we want the simulation to reach the present epoch, the relevant aspects of gas physics must be put in "by hand", in terms of semi-analytic recipes. As we have seen, some important aspects of the galaxy formation question, for example the way in which star-formation is suppressed in massive haloes, have still to be resolved.

The most effective way of limiting the number of possible solutions open to theorists is to increase the quality and angular coverage of our galaxy surveys, combining precise photometric redshifts, or, ideally, spectroscopic redshifts like the 100,000-galaxy VIPERS redshift survey aims to do. One can imagine extending the phenomenological halo model outlined here to include other observable quantities such as the galaxy stellar mass function or the galaxy-galaxy lensing signal to provide simultaneous constraints on several observables, and this is one direction we expect to take with the new combined COSMOS-UltraVISTA data set which I have been preparing at TERAPIX. Some recent papers in the literature (Peng et al., 2011; jie Peng et al., 2010) have also demonstrated that phenomenological models containing almost no physical processes

can also provide a remarkably accurate description of the evolution of galaxies over a wide range of redshift. Our current implementation of the halo model does not include blue, star-forming galaxies: could some elements from models like those described in Peng et al. be incorporated in our halo model?

In this context, in Winter 2010, in collaboration with colleagues in Saclay and Marseille, I led a French "ANR" proposal "Sagah" which aimed at extending our analytic model with the aim of using it to interpret observations from many different surveys and in particular the growth of stellar mass in the blue galaxy population and the role of active galactic nucleii in galaxy formation. Although this proposal was unsuccessful, we intend to resubmit a revised version at the next call in January 2012. This proposal provides template for the work I intend to carry out the next five years.

All of this takes place against the background of our current cosmological model. Although this model has been spectacularly successful in describing many aspects of the evolution of our Universe from the cosmic microwave background to the current day, there still remains the unfortunate fact that we do not have a clear idea what dark matter and dark energy really is, despite knowing, for dark matter at least, in great detail how it *behaves*. In practice, of course, astronomers working in the fields of galaxy and structure evolution have a tendency to avert their gaze from such uncomfortable facts, but still they remain. Our current hope is that a combination of immensely expensive particle accelerator experiments, direct searches (in which one looks for the ghostly hand of a dark matter particle triggering the recoil of a nucleus in a detector) and perhaps galaxy surveys such as the proposed EUCLID space mission in which a large fraction of the visible universe will resolve this, one of science's ultimate mysteries.

Chapter 6

Teaching activities

6.1 Student and postdoc supervision and associated publications

6.1.1 Postdocs

- **Richard Bielby**, 2008-2010, postdoctoral researcher employed in the context of the ANR "DESIR" project. Publication: "The WIRCAM Deep Infrared Cluster Survey. I. Groups and clusters at $z > 1.1$", Bielby, Finoguenov, Tanaka, McCracken et al., 2010 A&A, 523, 66B; a second paper detailing the WIRDS survey will be submitted shortly.

6.1.2 PhD students

- (w. O. Le Fèvre) **S. Foucaud**, 1999-2001. Publication: "The Canada-France deep fields survey-II: Lyman-break galaxies and galaxy clustering at $z \sim 3$", Foucaud, McCracken, Le Fèvre et al., A&A, 2003, 409, 835F;

- (w. Y. Mellier) **J. Coupon**, 2006-2009. Publication: "Photometric redshifts for the CFHTLS T0004 deep and wide fields", Coupon, Ilbert, Kilbinger, McCracken et al., A&A, 2009, 500, 981C

- (w. Y. Mellier) **J. Coupon**, 2006-2009. Publication: "Galaxy clustering in the CFHTLS-Wide: the changing relationship between galaxies and haloes since $z \sim 1.2$", Coupon, Kilbinger, McCracken, Ilbert et al., astro-ph 1107.0616

- (w. S. Colombi) **M. Wolk**, 2010–current. The subject of Melody's thesis the halo model and clustering of galaxies at intermediate redshifts.

6.1.3 Master students

- Two stagiares supervised at IAP (2007/2008).

6.2 Teaching

6.2.1 Observing schools

I have participated as a tutor and lecturer at several schools concerned with teaching observational methods:

- Tutor for the NEON observing schools at IUCCA, Pune, February 2007;
- Tutor for the NEON archive school, ESO, Munich, September 2008;
- Lecturer and tutor for the IPM, Tehran, Iran observing school, Iran, November 2008;
- Lecturer for the NEON observing school in Calar Alto, Spain, June 2010;
- Lecturer and tutor for the forthcoming observing school in Peking, China October 2011;

6.2.2 Courses taught

- Since October 2007 to present: lecturer and tutor M2, "Traitement et analyse de données de l'OHP, M2 Astronomie et Astrophysique" (with H. Dole, K. Benabed, M. Dennefeld). $\sim 100hr$. In this course, over five days, students formulate a small research program which can be carried out on telescopes at the Observatoire d'Haute provence, then carry out the observations, reduce and analyse their data, write up a report, and finally present their results.
- Tutor for the "post-master" M2 course module (where PhD students must complete a small research problem, outside their thesis speciality, over the course of several days), Paris (December 2008, 2009)

6.3 Other outreach activities

- Public speaker, Irish Astronomical Society, Belfast (March 2009);
- Participation in the "Chercheurs d'art et de science": three visits to a school at the Porte de Pantin, Paris (Febuary-April 2009);
- Regular participation at public events at the Observatoire de Paris and the IAP.

Bibliography

Abraham, R. G., et al. 2004, AJ, 127, 2455

Adelberger, K. L., Steidel, C. C., Pettini, M., Shapley, A. E., Reddy, N. A., & Erb, D. K. 2005, ApJ, 619, 697

Arnouts, S., et al. 2002, MNRAS, 329, 355+

Arnouts, S., et al. 2007, A&A, 476, 137

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

Baum, W. A. 1962, Problems of Extra-Galactic Research, 15, 390

Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379

Bell, E. F., et al. 2004, ApJ, 608, 752

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

Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587

Blanc, G. A., et al. 2008, ApJ, 681, 1099

Boulade, O., et al. 2000, in Proc. SPIE Vol. 4008, p. 657-668, Optical and IR Telescope Instrumentation and Detectors, Masanori Iye; Alan F. Moorwood; Eds., 657–668

Broadhurst, T. J., Ellis, R. S., Koo, D. C., & Szalay, A. S. 1990, Nature (ISSN 0028-0836), 343, 726

Brown, M. J. I., Dey, A., Jannuzi, B. T., Lauer, T. R., Tiede, G. P., & Mikles, V. J. 2003, ApJ, 597, 225

Brown, M. J. I., et al. 2008, ApJ, 682, 937

Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R., & Dekel, A. 2001, MNRAS, 321, 559

Capak, P., et al. 2007, ApJS, 172, 99

Cappé, O., Douc, R., Guillin, A., Marin, J.-M., & Robert, C. 2008, Statist. Comput., 18(4), 447

Cappé, O., Guillin, A., Marin, J.-M., & Robert, C. 2004, J. Comput. Graph. Statist., 13, 907

Carlberg, R. G., Sullivan, M., & Borgne, D. L. 2009, ApJ, 694, 1131

Cimatti, A., et al. 2002, A&A, 381, L68

—. 2008, A&A, 482, 21

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

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

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

Colless, M., et al. 2001, MNRAS, 328, 1039

Connolly, A. J., Szalay, A. S., Dickinson, M., Subbarao, M. U., & Brunner, R. J. 1997, Astrophysical Journal Letters v.486, 486, L11

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., & Wechsler, R. H. 2009, ApJ, 696, 620

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

Coupon, J., et al. 2009, A&A, 500, 981

—. 2011, arXiv, astro-ph.CO

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

Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, Astronomical Journal v.112, 112, 839

Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303

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

Daddi, E., et al. 2005, ApJ, 631, L13

—. 2007, ApJ, 670, 156

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

de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, ApJ, 302, L1

Dickinson, M., Papovich, C., Ferguson, H. C., & Budavári, T. 2003, ApJ, 587, 25

Drory, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., & Pannella, M. 2005, ApJ, 619, L131

Elston, R. J., et al. 2006, ApJ, 639, 816

Faber, S. M., et al. 2007, ApJ, 665, 265

Fevre, O. L., Hudon, D., Lilly, S. J., Crampton, D., Hammer, F., & Tresse, L. 1996, Astrophysical Journal v.461, 461, 534

Fèvre, O. L., et al. 2005a, Nature, 437, 519

—. 2005b, A&A, 439, 845

Fontana, A., et al. 2004, A&A, 424, 23

—. 2006, A&A, 459, 745

Foucaud, S., Conselice, C. J., Hartley, W. G., Lane, K. P., Bamford, S. P., Almaini, O., & Bundy, K. 2010, MNRAS, 406, 147

Franx, M., et al. 2003, ApJ, 587, L79

Garilli, B., et al. 2008, A&A, 486, 683

Glazebrook, K., et al. 2004, Nature, 430, 181

Grazian, A., et al. 2007, A&A, 465, 393

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

Guhathakurta, P., Tyson, J. A., & Majewski, S. R. 1990, ApJ, 357, L9

Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, MNRAS, 1

Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37

Hartley, W. G., et al. 2008, MNRAS, 391, 1301

Hu, E. M., & Ridgway, S. E. 1994, AJ, 107, 1303

Huang, J., Cowie, L. L., Gardner, J. P., Hu, E. M., Songaila, A., & Wainscoat, R. J. 1997, ApJ, 476, 12+

Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, Astrophysical Journal Supplement Series (ISSN 0067-0049), 52, 89

Ilbert, O., et al. 2005, A&A, 439, 863

Ilbert, O., et al. 2006, A&A, 457, 841

Ilbert, O., et al. 2006, A&A, 457, 841

Ilbert, O., et al. 2006, A&A, 453, 809

—. 2009, ApJ, 690, 1236

—. 2010, ApJ, 709, 644

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

jie Peng, Y., et al. 2010, ApJ, 721, 193

Juneau, S., et al. 2005, ApJ, 619, L135

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 307, 529

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

Kong, X., et al. 2006, ApJ, 638, 72

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., & Primack, J. R. 2004, ApJ, 609, 35

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

Lane, K. P., et al. 2007, MNRAS, 379, L25

Le Fèvre, O., et al. 2005, A&A, 439, 845

Leauthaud, A., et al. 2011, eprint arXiv, 1104, 928

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

Lilly, S. J., Fevre, O. L., Crampton, D., Hammer, F., & Tresse, L. 1995, Astrophysical Journal v.455, 455, 50

Lilly, S. J., Fevre, O. L., Hammer, F., & Crampton, D. 1996, Astrophysical Journal Letters v.460, 460, L1

Lilly, S. J., et al. 2007, ApJS, 172, 70

Limber, D. N. 1953, ApJ, 117, 134

Limber, D. N. 1953, ApJ, 117, 134

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

Lin, H., Yee, H. K. C., Carlberg, R. G., Morris, S. L., Sawicki, M., Patton, D. R., Wirth, G., & Shepherd, C. W. 1999, ApJ, 518, 533

Longhetti, M., & Saracco, P. 2009, MNRAS, 394, 774

Loveday, J., Tresse, L., & Maddox, S. 1999, MNRAS, 310, 281

Ma, C.-P., & Fry, J. N. 2000, ApJ, 543, 503

Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388

Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, MNRAS, 242, 43P

Madgwick, D. S., et al. 2003, MNRAS, 344, 847

McCracken, H. J., Fèvre, O. L., Brodwin, M., Foucaud, S., Lilly, S. J., Crampton, D., & Mellier, Y. 2001, A&A, 376, 756

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

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

McCracken, H. J., Metcalfe, N., Shanks, T., Campos, A., Gardner, J. P., & Fong, R. 2000, MNRAS, 311, 707

McCracken, H. J., et al. 2003, A&A, 410, 17

McCracken, H. J., et al. 2007, ApJS, 172, 314

McCracken, H. J., et al. 2010, ApJ, 708, 202

Meneux, B., et al. 2006, A&A, 452, 387

Meneux, B., et al. 2009, A&A, 505, 463

Metcalfe, N., Shanks, T., Campos, A., McCracken, H. J., & Fong, R. 2001, MNRAS, 323, 795

Mignoli, M., et al. 2009, A&A, 493, 39

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

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

Moster, B. P., Somerville, R. S., Maulbetsch, C., van den Bosch, F. C., Macciò, A. V., Naab, T., & Oser, L. 2010, ApJ, 710, 903

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, Astrophysical Journal v.490, 490, 493

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

Noeske, K. G., et al. 2007, ApJ, 660, L43

Norberg, P., et al. 2002, MNRAS, 332, 827

Ostriker, J. P., & Peebles, P. J. E. 1973, ApJ, 186, 467

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

Peacock, J. A., et al. 2001, Nature, 410, 169

Peebles, P. J. E. 1980, Research supported by the National Science Foundation. Princeton

Peng, Y., Lilly, S. J., Renzini, A., & Carollo, M. 2011, eprint arXiv, 1106, 2546

Pérez-González, P. G., et al. 2008, ApJ, 675, 234

Pozzetti, L., et al. 2007, A&A, 474, 443

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

Renzini, A., & Daddi, E. 2009, eprint arXiv, 0906, 4662

Sanders, D. B., et al. 2007, ApJS, 172, 86

Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20

Scoville, N., et al. 2007, ApJS, 172, 1

Seljak, U. 2000, MNRAS, 318, 203

Shane, C., & Wirtanen, C. 1967, Publications of the Lick Observatory, XXII, Pt. 1

Shane, C. D., & Wirtanen, C. A. 1954, AJ, 59, 285

Shectman, S. A., Landy, S. D., Oemler, A., Tucker, D. L., Lin, H., Kirshner, R. P., & Schechter, P. L. 1996, Astrophysical Journal v.470, 470, 172

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

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

Silverman, J. D., et al. 2009, ApJ, 696, 396

Smith, R. E., et al. 2003, MNRAS, 341, 1311

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, Astrophysical Journal Letters v.462, 462, L17

Swanson, M. E. C., Tegmark, M., Blanton, M., & Zehavi, I. 2007, astro-ph/0702584

Taylor, E. N., et al. 2009, ApJ, 694, 1171

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. 2008, ApJ, 688, 709

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41

Totsuji, H., & Kihara, T. 1969, Publications of the Astronomical Society of Japan, 21, 221

van der Wel, A., et al. 2007, ApJ, 670, 206

Van Waerbeke, L., et al. 2000, A&A, 358, 30

Wake, D. A., et al. 2010, eprint arXiv, 1012, 1317

Weinberg, D. H., Davé, R., Katz, N., & Hernquist, L. 2004, ApJ, 601, 1

White, S. D. M., & Rees, M. J. 1978, Royal Astronomical Society, 183, 341

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

Williams, R. E., et al. 1996, AJ, 112, 1335+

Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S., & Kleinheinrich, M. 2003, A&A, 401, 73

Wraith, D., Kilbinger, M., Benabed, K., Cappé, O., Cardoso, J.-F., Fort, G., Prunet, S., & Robert, C. P. 2009, Phys. Rev. D, 80, 23507

Wuyts, S., Labbé, I., Schreiber, N. M. F., Franx, M., Rudnick, G., Brammer, G. B., & van Dokkum, P. G. 2008, ApJ, 682, 985

York, D. G., et al. 2000, AJ, 120, 1579

Zehavi, I., et al. 2004, ApJ, 608, 16

Zehavi, I., et al. 2005, ApJ, 630, 1

Zehavi, I., et al. 2010, eprint arXiv, 1005, 2413

Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H. 2005, ApJ, 624, 505

Zheng, Z., et al. 2005, ApJ, 633, 791

Zucca, E., et al. 2006, A&A, 455, 879