The Nature of Emission-Line Galaxies in Hierarchical Cosmologies

ALVARO ANTONINO ORSI-MOYANO

The Nature of Emission-Line Galaxies in Hierarchical Cosmologies

Alvaro Orsi

Abstract

We use a galaxy formation model to study the nature and evolution of emission line galaxies. In particular, we focus on the properties of Lyα and Hα emitters, due to their many cosmological applications being considered for current and future observational studies.

By combining a semianalytical model with a large N-body simulation we predict the clustering of Lyα emitters. With increasing redshift, Lyα emitters are found to trace progressively rarer, higher density regions of the Universe. We measure the clustering of Lyα emitters by constructing mock catalogues of surveys finding a good agreement between the model and the observational measurements. Furthermore, we use the mock catalogues to study the sample variance of current and forthcoming Lyα surveys. Current surveys should be extended significantly in solid angle to allow a robust measurement of the clustering of Lyα emitters, particularly at $z > 8$.

On the other hand, future space-based galaxy surveys will map the galaxy distribution using Hα emitters or H-band selected galaxies at $0.5 < z < 2$ to constrain the nature of the dark energy by measuring the large-scale structure of the Universe. Therefore, we investigate the abundance and clustering of galaxies found using these two selections. Hα emitters are found to avoid massive dark matter haloes, whereas H-band selected galaxies are found in the highest mass haloes. By using mock catalogues, we predict the effectiveness of measuring the large scale structure of the Universe for a range of survey configurations using both galaxy selections.

Finally, we study the escape of Lyα photons from galaxies using a Monte Carlo radiative transfer code. We simulate galactic outflows in a semianalytical model to study the physical properties of Lyα emitters in a cosmological context. We find that the escape fraction of Lyα emitters can vary greatly depending on the properties of the galaxies, although our results depend on the outflow model used. Our results suggest the need to consider additional physical effects to understand the observed properties of Lyα emitters.

The Nature of Emission-Line Galaxies in Hierarchical Cosmologies

by Alvaro Orsi

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

Department of Physics
University of Durham
September 2010

Contents

List of Figures

List of Tables

Declaration

The work described in this thesis was undertaken between 2007 and 2010 while the author was a research student under the supervision of Dr. Cedric G. Lacey and Prof. Carlton M. Baugh in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or any other University.

Chapter 3 of this thesis has been published in the form of a paper

- A. Orsi, C.G. Lacey, C.M. Baugh, and L. Infante, "The Clustering of Lyα emitters in a ΛCDM Universe", MNRAS, 391:1589-1604, 2008

Chapter 4 of this thesis has also been published in the form of a paper

- A. Orsi, C.M. Baugh, C.G. Lacey, A. Cimatti, Y. Wang, and G. Zamorani, "Probing dark energy with future redshift surveys: a comparison of emission line and broad-band selection in the near-infrared", MNRAS, 405:1006-1024, 2010

All the work shown in Chapters 3 and 4 was performed entirely by the author of this thesis. In addition, Chapters 5 and 6 are being planned to be submitted in the form of a paper after submission of this thesis.

Acknowledgements

First of all I want to thank my two supervisors Cedric and Carlton, for their guidance, patience and for teaching me, with their two different styles, most of the ideas now shape my vision about scientific research. I am also grateful for the time they spent reading earlier versions of this thesis.

The past three years have been full of interesting moments, and this is mainly thanks to the people I got to meet in and around the wonderful city of Durham. From those I want to express my gratitude first to my friend and flatmate for over two years Nikos Fanidakis, for his friendship, support in difficult moments, inexhaustible knowledge about the wonders of Greece and for shaping my ideas about cooking and nutrition.

There are so many cool people I would like to mention here to let them know how much they contributed for making these three years living in Durham a great experience. Obviously, a thesis is not the right place to do it, but still I would like to mention my good friends Raul Angulo, Claudia Lagos, Juan E. Gonzalez, Nicolas Tejos, Elise Jennings, Nikos Nikoloudakis, Alex Merson, Violeta Gonzalez, Gabriel Altay, Dave Murphy, Daniel John, Julie Wardlow, Pratika Dayal, Milan Raičević and Tim Rawle, for the good times in and outside the department, the stimulating discussions about science, the movie and poker nights, the nights at Dun Cow and Jam-Jah, the trips we made together, and many more.

A very special thanks to Lydia Heck, who patiently assisted me whenever I created trouble in COSMA.

I am grateful to those who were lucky to participate in our postgraduate journal clubs, for contributing to create an atmosphere suitable for feeling encouraged to ask even the most silly questions, and yet feeling at the end of it that we all learned a lot.

A special mention goes also to my family, for the support and encouragement during these years. They managed to keep our connection alive and stronger than ever, which is something I am truly grateful for.

A big thanks to Claudia, Nikos and Raul, who also contributed with comments while reading different parts of this thesis.

The work on this thesis has been supported by a STFC-Gemini scholarship.

Chapter 1 | *Introduction*

Over the past few years our understanding of the cosmic history of the Universe has improved at a rapid pace. The COsmic Background Explorer (COBE) satellite measured for the first time the anisotropies in the cosmic microwave background (CMB) radiation (Smoot et al., 1992; Gorski et al., 1994; Bennett et al., 1996). These temperature fluctuations correspond to primordial ripples in the density of the Universe, and are regarded as the seeds of the matter density fluctuations in the Universe as observed today. Furthermore, the discovery of the accelerating cosmic expansion of the Universe in 1998 by two independent teams (Riess et al. 1998 and Perlmutter et al. 1999) using Type Ia supernovae as standard candles was a major breaktrough for cosmology, since it introduced a new, unknown and little understood, major component in the energy budget of the Universe, the so-called *dark energy*.

A conclusion along the same lines had already been suggested nearly 10 years before as a result of the analysis of the spatial distribution of galaxies in the APM (Efstathiou et al., 1990) and QDOT (Saunders et al., 1991) surveys. In the following years, galaxy surveys were developed to become robust probes of both cosmology and galaxy formation. The present-day Universe has been mapped with unprecedented detail by two very large surveys completed over the last decade, the Sloan Digital Sky Survey (SDSS) (York et al., 2000) and the 2-degree Field Galaxy Redshift Survey (2dFGRS) (Colless et al., 2003). These two surveys have provided a huge amount of data, allowing the study of galaxy properties with a level of detail and robustness never achieved before. As a result, theories of galaxy formation have been greatly developed, both in complexity and in the scope of their predictions in order to account for this huge wave of observational data.

Furthermore, galaxies have also been tracked back in time to epochs when the Universe was in its infancy, just a few hundreds of million years after the Big Bang. New telescopes are being exploited to their maximum potential to try to reveal the *cosmic dawn*, the epoch when

the first stars and galaxies formed.

To undertake such high redshift galaxy surveys, observational techniques have been developed to search for particular features in the spectral energy distributions of galaxies. In a way, this makes the task of interpreting the observations more complex, since different spectral signatures are generated from different physical processes inside galaxies, and so different survey designs could probe galaxy populations with specific characteristics. For example, *emission-line galaxies* are found by detecting a spectral line in emission against the stellar continuum. In general, this spectral signature is produced by the recombination of ionised gas in the interstellar medium of galaxies. The physics governing the production and the escape of these photons could be very complex, as we will study later in this thesis.

It is clear that a full picture of galaxy formation and evolution requires an understanding of the multi-wavelength properties of galaxies in order to link the properties of primeval galaxies in models, to what is observed in the present day Universe.

A fundamental clue to understanding the intrinsic link between galaxy evolution and the cosmic history of the Universe is the spatial distribution of galaxies. Half of this thesis is devoted to the modelling using numerical techniques of the large scale structure of galaxies and its evolution. Apart from its significance from a galaxy formation perspective, the clustering properties of the galaxy distribution can be used to infer and constrain some of the fundamental cosmological parameters.

By combining data from several cosmological probes such as supernovae distances, CMB anisotropies and the large scale structure of galaxies, a flat Λ cold dark matter (ΛCDM) cosmology is favoured over alternative cosmologies (e.g. Efstathiou et al., 2002). The energy budget most consistent with observational data consists of a balance of matter (baryonic plus dark) of $\Omega_m \sim 0.25$, and dark energy with $\Omega_\Lambda \sim 0.75$, where Ω_m and Ω_Λ denote the ratio of the energy density of dark matter and dark energy to the critical density of the Universe, respectively (see, for example, Tegmark et al., 2004; Percival et al., 2007; Sánchez et al., 2009, for a full description of the cosmological parameters that fit the data best).

In the standard ΛCDM Universe, the dark matter dominates the large scale gravitational forces. It is defined as *cold*, because these particles have non relativistic velocities when their

interaction rate drops below the expansion rate; *dissipationless*, since they cannot cool radiatively, as they do not have electromagnetic interactions and so cannot produce photons; and *collisionless*, since these particles only interact gravitationally.

To date it is still not clear what is the nature of the dark matter. Light neutrinos were the first proposed candidates for dark matter (e.g. Gershtein and Zel'Dovich, 1966; Zel'dovich and Khlopov, 1981; Harari, 1989), however the requirement that the dark matter must be cold ruled them out (see, for example, White et al., 1984). Nowadays, Weakly Interacting Massive Particles (WIMPs) and axions are the two most promising candidates for dark matter, both of which are potentially detectable (for a review on dark matter candiates, see Krauss, 2007).

In a standard ΛCDM model the dark energy equation of state parameter $w_{DE} = P/\rho$ is equal to -1 at all redshifts, which means that dark energy behaves like a cosmological constant. To date, no departure from $w_{DE} = -1$ has been measured. A discovery of evolution in w_{DE} would be an indication of a departure from the standard ΛCDM model, and thus remains one of the most active topics of interest in modern cosmology (Peacock et al., 2006). A detailed review of alternative cosmological theories can be found in Caldwell and Kamionkowski (2009).

1.1 The large scale structure of the Universe

Before the theory of *inflation* was postulated (Guth, 1981), the origin of the observed structure in the Universe was not properly understood. Galaxies trace the underlying dark matter distribution, which is expected to evolve mainly due to gravity. Nevertheless, the classical model of cosmology does not account for the *initial* seeds from which a density field could evolve into something resembling what is observed in the Universe today. The proposal of an inflationary period in the early Universe solved this and several other unresolved questions in cosmology at that time.

During the period of inflation, starting shortly after the Big Bang, the Universe is supposed to experience a period of exponential growth, driven by a slowly rolling scalar field. The

original idea was proposed to explain other issues of the classical model of cosmology, of which two important ones were known as the *flatness* and *horizon* problems.

The first one arises because observations already suggested that the geometry of the Universe was nearly flat, however such a geometry is an unstable solution to the cosmological evolution equations and requires a very constrained set of initial conditions. The second problem refers to the apparent causal connection between distant regions of the sky having essentially the same CMB temperature (the anisotropies are very small), since such regions, according to the particle horizon at the time of matter-radiation decoupling, should not have been causally connected, and hence should not have had time to reach thermal equilibrium.

In addition to the success of inflation in solving the above problems, a side product is a primordial perturbed density field with a nearly scale invariant power spectrum. The subsequent gravitational growth of those density perturbations to the present day resembles the observed large scale structure of galaxies, supporting also the idea that galaxies form inside dark matter structures, and their evolution is intrinsically linked to the hierarchical growth of the underlying density peaks of the dark matter distribution (White and Rees, 1978).

Observationally, the distribution of galaxies on the sky is, in principle, easy to measure and characterise, since it only requires us to count and measure distances between objects. The distribution of galaxies on the sky was already known to be irregular from observations dating back almost 80 years ago (see, for example, Hubble, 1934), even considering the limitations due to the quality of photographic plates at that time. Galaxies seemed also to be grouped into *clusters*, as shown years later by Abell (1958) in his early catalogue of ~ 4000 galaxy clusters from the National Geographic Society Palomar Obervatory Sky Survey. This suggested already that the environment where galaxies live could be an important piece in the puzzle to understand their nature.

The task of measuring the spatial distribution of galaxies is, however, more complicated considering that an image from a telescope is only the projection on a plane of a 3 dimensional space. In principle, the statistical properties of the spatial distribution of galaxies can be inferred from the measured *angular* distribution after deprojecting if the redshift distribution of sources is known (Limber 1953; but see also Simon 2007).

In a long series of papers, Peebles and collaborators developed extensive quantitative analysis of the clustering of galaxies from surveys. In order to measure the large scale structure in the galaxy distribution they estimated n-point correlation functions. For example, the two-point correlation function $\xi(r)$, and its analogous on a plane, $w(\theta)$, are defined as the excess probability of finding galaxy pairs at a given separation to what is expected from a random distribution of galaxies. This and higher order correlation function estimators were applied to the Zwicky and Lick galaxy catalogues in order to characterise the large scale distribution of galaxies (see Peebles and Hauser, 1974; Peebles and Groth, 1975; Fry and Peebles, 1978; Seldner and Peebles, 1978; Fry and Peebles, 1980, and references therein).

Nevertheless, a more accurate measurement of the spatial clustering of galaxies can be achieved when the redshift of every source is known. The *CfA Redshift Survey* (Davis et al., 1982) mapped the 3D distribution of galaxies for the first time, detecting $\sim$ 2400 galaxies grouped in a filamentary structure featuring clusters and superclusters. This survey confirmed that the distribution of galaxies was anything but random. In addition to the famous human-shaped supercluster named the *Stick Man*, the CfA survey also detected a great superstructure, the *Great Wall*, made out of clusters, filaments and other amorphous structures.

Despite the success of redshift surveys at mapping the spatial distribution of structure, in reality the redshift of galaxies is not enough to get a real 3D picture of their distribution, since peculiar velocities of galaxies along the radial direction will Doppler shift the measured redshift, thus distorting the apparent distance to the galaxy. The peculiar velocities of galaxies arise from the gravity of the clustered structure of the matter distribution itself, which causes the *redshift distortions* to have a characteristic clustering pattern: In simple terms, galaxies inside clusters belong to virialized structures, making their peculiar motions random, whereas on larger scales galaxies have coherent peculiar velocity motions which point towards nearby massive structures. The clustering pattern resulting from these redshift distortions is studied in more detail in Chapter 4 of this thesis.

The large scale structure extracted from galaxy surveys is known to depend on the properties of the galaxy population used. The clustering strength depends on the luminosity,

colour, morphology, star formation rate and other properties of galaxies, but also depends on the scale lengths and redshift considered (see, for example, The SDSS Collaboration et al., 2010). Furthermore, the large scale structure of a galaxy population depends intrinsically on the masses of the haloes which host the galaxies. Since galaxies reside in the peaks of the matter density distribution, they are *biased* tracers of the dark matter distribution. This bias can be simply quantified as the ratio of the clustering strength of galaxies to that of dark matter. The bias function is known to vary with halo mass, in the sense that more massive structures appear more clustered with respect to the underlying matter. Also, for a given halo mass, the bias is expected to increase with redshift, since massive haloes are less abundant and correspond to higher peaks in the density field, which is more homogeneous at higher redshifts.

The clustering signal provided by large scale structure measurements is imprinted with features from the early Universe, when matter and radiation were mixed in a hot, dense and ionized plasma. This original plasma was homogeneous except for small mass density fluctuations, which create pressure gradients in the medium. The resulting sound waves in baryons are known as *Baryonic Acoustic Oscillations* (BAO).

At the time of decoupling between matter and radiation, the Universe had expanded so that its density and temperature decreased sufficiently to make possible the recombination of ionized atoms. This allows photons to escape from the primordial plasma. The oscillation of the baryonic gas is stalled since there is no more radiation pressure from the medium. The density field of baryons is thus imprinted with a characteristic peak, at a scale of approximately the sound horizon at that epoch, $s \sim 150$Mpc (Eisenstein and Hu, 1998). The acoustic peak is also imprinted as a feature of the CMB temperature anisotropy map, since it shows radiation density anisotropies at the time of the decoupling (Peebles and Yu, 1970; Bond and Efstathiou, 1984). The baryonic feature is imprinted on the power spectrum of the matter distribution. Both features correspond to the same comoving length scale, but the BAO peak can be measured from the galaxy distribution at different epochs, where the expansion of the Universe shifts the peak position in physical scale length. This is why BAOs are regarded as a *cosmic ruler*, since, when measured, the peak corresponds to the same scale length measure in the CMB map, which has been accurately measured by the WMAP satellite (Dunkley et al.,

2009).

The cosmic expansion can be followed by measuring the length of this cosmic ruler at different epochs with potentially high accuracy, making the detection of BAOs one of the most important probes of the cosmological paradigm. The acoustic peak of the CMB can determine the angular diameter scale at $z = 1089$ (the redshift of decoupling) very accurately. If a flat cosmology is assumed, then this measurement depends only on Ω_m or Ω_Λ. If the Universe is allowed to have non-zero curvature, or w_{DE} is allowed to vary, then the CMB data alone show parameter degeneracies in the best-fit of the angular diameter scale of the acoustic peak. Therefore, by adding the measurement of the BAO peak from the galaxy distribution, it is possible to break these degeneracies.

Using the spectroscopic sample of Luminous Red Galaxies from the SDSS, Eisenstein et al. (2005) detected the BAO peak in the correlation function of galaxies. The detection was rather noisy though, since the scale that must be probed is very large. Nevetheless, it was shown that the Universe is flat to within 1% accuracy, assuming $w_{DE} = -1$.

This motivated current efforts to construct very large redshift surveys of galaxies at different redshifts with the aim of measuring the BAO signal with accuracies better than $< 1\%$ in order to constrain the nature of the energy components of the Universe, and in particular of the dark energy.

Nowadays, understanding the nature of dark energy is regarded as one of the most important challenges in science. Several redshift surveys have been planned to measure the BAO peak with different techniques spanning a large range of redshifts. Some of the current and forthcoming redshift surveys attempting to measure the BAO peak are: WiggleZ (Blake et al., 2009), HETDEX (Hill et al., 2008), Euclid and WFIRST (Cimatti et al., 2009), BOSS (Ross et al., 2010), Big-BOSS (Mostek et al., 2010), PanSTARRS (Kaiser, 2006) and WFMOS (Nichol, 2006) . The analysis of these surveys in the forthcoming years is expected to revolutionize the current paradigm of cosmology by either validating it or ruling it out at an unprecedented level of accuracy.

From the above list of surveys, WiggleZ, WFMOS, Euclid, HETDEX and WFIRST are designed to trace the dark matter distribution using emission-line galaxies. In order to fully understand and be able to interpret the forthcoming data from these surveys, it is imperative to have a detailed understanding of the properties of emission-line galaxies from a theoretical perspective. Surprisingly, there has been little effort to model, predict and interpret the properties of this particular population of galaxies. Therefore, one of the goals of this thesis to shed light on some of the properties of emission line galaxies.

1.2 Emission-line galaxies

As mentioned earlier, emission lines produced in galaxies are the result of recombinations of hydrogen in the interstellar medium (ISM) caused by ionizing Lyman continuum (LC) photons. These photons are mostly generated by young, massive stars, which makes observed emission-lines tracers of the star formation rate in a galaxy. It is important to notice that observed emission lines from galaxies can also be produced by sources other than stars, like AGNs (e.g. Kaspi et al., 2000), Supernova remnants (e.g. Milisavljevic et al., 2010), or cold accretion streams generating collisionally excited line radiation (e.g. Dijkstra and Loeb, 2009). The study of these alternative mechanisms of emission line production is beyond the scope of this thesis. Therefore, we will focus on hydrogen emission lines produced by stellar photoionising radiation in actively star formation galaxies.

Hydrogen recombination lines have been historically used as a probe of the star formation activity in galaxies. The most common lines used for this purpose are Hα, Hβ, Lyα, Pα, Pβ, Brα and Brβ. Forbidden lines like [OII] and [OIII] are also common star formation indicators, although these are formed by collisionally excited radiation and thus are not directly coupled to ionizing radiation (see Kennicutt, 1998b, for a review of the different star formation indicators).

Several physical processes play important roles in shaping the emission line that is observed from a galaxy. Although these hydrogen emission lines are originated from the same source, their observed strength correlates with the rate of production of LC photons (and,

thus, with the star formation rate) in a different way for every line. For example, the presence of dust will attenuate the intrinsic strength of the lines. Moreover, dust attenuation is a function of wavelength, with UV photons being more prone to being absorbed by dust than photons in optical or infrared bands.

In this thesis, we will focus our attention on two recombination lines: The Hα and Lyα emission lines. Both have been extensively used in the past for estimating star formation rates and also to search for primeval galaxies at high redshifts (see Schaerer, 2007, for a review). Also, forthcoming redshift surveys will use these two lines as tracers of galaxies to probe the nature of dark energy, as described above. In the following we will briefly describe the properties of each of these lines.

The Hα line, at a restframe of $\lambda = 6562$Å, is perhaps the most commonly used estimator of the star formation activity in galaxies. Surveys of Hα emitters have been used to infer the global star formation history over the range $0 < z < 2$ (e.g. Gallego et al., 1995; Fujita et al., 2003; Pascual et al., 2001; Hopkins et al., 2000; Hayes et al., 2010b; Geach et al., 2008, among others).

The advantages of the Hα emission as a tracer of the star formation rate lies on a combination of its high signal, since it is the strongest of the Balmer transitions, and its relatively low attenuation by dust. The extinction of the Hα line can be estimated by comparing the observed Hα fluxes to other lines such as Hβ, and comparing this to the expected value from recombination theory. The mean extinction found in this way is usually around $A(H\alpha) = 0.8 - 1.1$ mag for nearby galaxies (Kennicutt, 1983; Niklas et al., 1997; Kennicutt, 1998b). However, more recent studies have shown a large spread in this measurement, and the data also suggests a correlation between the extinction and the morphological type of galaxies (James et al., 2005).

Another important emission line is Lyα, at a restframe wavelength of $\lambda = 1216$Å. This emission arises from the transition between the states $n = 2$ to $n = 1$ in the hydrogen atom, and is the strongest hydrogen recombination line in absence of attenuation, which is why it was suggested by Partridge and Peebles (1967) that the Lyα line could be used as a tracer of

young, distant galaxies.

However, it took over 30 years to establish Lyα detections as a robust technique to search for high redshift objects. The attenuation and properties of Lyα emitters are more difficult to model than for Hα photons, since Lyα is a resonant line. This means that photons created from the initial recombination of ionized gas by LC photons into Lyα will be continuously scattered by hydrogen atoms, thus increasing the path length photons travel compared to a non-resonant photon, such as Hα. The complicated paths photons travel before escaping alter the measured line profile, and since the photon path length is larger than with continuum photons, this increases the probability for the photons to be absorbed by dust grains. Nevertheless, using the measured spectrophotometric properties of Lyα emitters, there is general agreement that these galaxies are mostly young, actively forming stars, low in mass and dust (Cowie and Hu, 1998; Kudritzki et al., 2000; Gawiser et al., 2007; Nilsson and Meisenheimer, 2009).

The Lyα technique has been particularly successful for making surveys of galaxies in the redshift range $2 < z < 7$ (e.g. Cowie and Hu, 1998; Hu et al., 2002; Gronwall et al., 2007; Ouchi et al., 2008; Nilsson et al., 2009; Shimasaku et al., 2006; Kashikawa et al., 2006; Hu et al., 2010), and recently also at $z \sim 0.2$ thanks to data from the GALEX satellite (Deharveng et al., 2008; Cowie et al., 2010). Candidates with Lyα emission are most commonly found through photometry with a narrow band filter centred at a redshifted Lyα wavelength. In general, candidates are subsequently confirmed with spectroscopy. Other techniques for detecting Lyα emission of galaxies include spectroscopy of lensed galaxies (Stark et al., 2007) and integral-field-unit (IFU) spectroscopy (Blanc et al., 2007; Hill et al., 2008).

One of the pending tasks for Lyα searches is to extend the method to very high redshifts ($z \geq 7$) to get a statistical sample of galaxies at epochs when the reionization of the Universe was not complete (Kashikawa et al., 2006; Iye et al., 2006; Stark et al., 2007; Hibon et al., 2010). A number of surveys are planned to aim at these redshifts, like the DAZLE survey, which aims to detect Lyα emitters at $z \sim 7.5$ (Horton et al., 2004; Venemans et al., 2009) and the ELVIS survey of Lyα emitters at $z \sim 8.8$ (Nilsson et al., 2007a). At the moment, only an upper limit has been set to the luminosity function of $z \sim 7.7$ Lyα emitters (Hibon et al.,

2010; Tilvi et al., 2010).

Nevertheless, current surveys of Lyα emitters have been able to measure the clustering properties of this galaxy population at redshifts $z \sim 6$. It is important, from a galaxy formation perspective, to be able to characterise the clustering properties of galaxies at these high redshifts, since the clustering strength is directly related to the dark matter halo population hosting these galaxies. Furthermore, it is also important to understand the reliability of such measurements. In Chapter 3 we study this problem in detail making use of models of galaxy formation.

1.3 Numerical models of galaxy formation

We aim to study emission line galaxies in a cosmological context using theoretical models of galaxy formation. Despite the great progress in galaxy formation theory, the underlying physics describing the cosmological framework is fundamentally nonlinear, which motivates the need for a numerical approach to study galaxy formation.

Several techniques have been developed to tackle different aspects of galaxy and structure formation. The most direct method is to model the evolution of the dark matter content of the Universe with the *N-body technique*. This follows the gravitational interactions of a set of particles, using an efficient algorithm, to trace the evolution of a large number of particles from an early epoch down to the present time.

N-body simulations of dark matter are widely used nowadays (see, e.g. Trenti and Hut, 2008, for a review of the method and current applications). The strategy is based on setting the initial conditions to define a set of particles distributed according to a Gaussian random density field with a given primordial power spectrum (consistent with the CMB), and then let that set of particles interact gravitationally through time until the present day.

Of particular relevance to this thesis is the *Millennium simulation* (Springel et al., 2005), which follows the gravitational interaction of a set of 10 billion particles on a perdiodic box

of 500[Mpc/h] a side, from $z = 127$ to the present day, $z = 0$. This N-body simulation, the state of the art at its time, is able to resolve dark matter haloes with masses down to $1.7 \times 10^{10}[M_{\odot}/h]$, which proves to be ideal for our purposes, as we will discuss in the next chapters of this thesis. We also make use of the *BASICC simulation* (Angulo et al., 2008a), which resolves dark matter haloes down to a factor 32 times more massive than the Millennium minimum halo mass, but on a box size of 1340[Mpc/h] a side, a scale suitable for BAO and redshift distortions studies.

Clearly, to study galaxy formation we also need a model for the baryonic component of the Universe. In a particle-based method known as *Smoothed Particle Hydrodinamics* (SPH) (e.g. Monaghan, 1992; Springel and Hernquist, 2003), three sets of particles are followed, one representing dark matter, another representing gas and the other stars. Dark matter particles are collisionless and only respond to gravitational forces, whereas gas particles can also feel pressure and dissipate energy through cooling. Despite the sophistication of the technique, many of the physical processes regulating galaxy formation occur on a scale well below the resolution of these simulations, named *subgrid* physics in this context, and thus they are treated in a phenomenological way. Another drawback of the SPH technique is the size of the simulations, which in general is not large enough to study large scale structure on the cosmological scales we are interested in at the same time as resolving the internal structure of galaxies.

An alternative to the SPH technique are the so-called *semi-analytical* models. This technique, which will be explained in detail in the next chapter, treats various physical processes associated with galaxy formation using approximate, analytical expressions. The degree of approximation varies considerably with the complexity of the physical problem being modelled.

The primary advantage of the semi-analytical technique is that is considerably less expensive in computational resources, allowing one to construct galaxy samples with a mass resolution usually orders of magnitude better than what is possible with SPH simulations of comparable cosmological volumes. Furthermore, it has been proved that the semi-analytical technique can achieve a very good agreement with SPH simulations in a number of properties

(Benson et al., 2001; Berlind et al., 2003; Helly et al., 2003). The semianalytical technique can also be coupled to N-body simulations of the dark matter in order to probe the large scale structure of galaxies on cosmological scales. Their most important drawback is, obviously, the large degree of approximation, which is sometimes difficult to assess.

1.4 Motivation

Throughout this thesis we study several aspects of the nature and evolution of emission line galaxies (in particular Lyα and Hα). The numerical method of choice is the semianalytical technique, due to its ability to probe the large scale structure of galaxies on cosmological scales, and also for its flexibility, making the task of comparing the output of models with different physical recipes to observations easier. Two main ideas motivate our study of emission line galaxies on a cosmological context. These are reviewed next.

First, the current and forthcoming exploration of the high redshift Universe with Lyα emitters is expanding the understanding of galaxy formation and evolution. Furthermore, the properties of very high redshift galaxies (perhaps the *first* galaxies in the Universe) are still yet to be revealed observationally. Galaxy formation models are only now beginning to predict the properties of Lyα emitters and other high-z galaxy populations. Therefore, these are exciting times to aim for a robust understanding of the nature of Lyα emitters.

Throughout this thesis, two fundamental aspects of this galaxy population are studied in detail:

- Their **clustering properties**, which depend on the dark matter halo masses which host these galaxies and also affect the statistical robustness of current and forthcoming observations (cosmic variance).

- The modelling of the **Lyα escape fraction** through the coupling of a radiative transfer model of Lyα emission to the predicted galaxy properties obtained with a semianalytical model, to interpret and link the observed Lyα luminosities with physical properties of galaxies.

Our second motivation comes from the next era of large redshift surveys, which will possibly revolutionize our current cosmological paradigm profoundly, by characterizing the nature and evolution of dark energy. A number of surveys, both ground-based and space missions, are being planned to search for emission line galaxies (particularly Hα, Lyα, [OII] and [OIII]) to trace the underlying matter distribution of the Universe at different redshifts.

Hence, we undertake the task of assessing the outcome of such surveys. In particular, we compare the performance of a slitless spectroscopic Hα- based survey to an alternative, photometrically selected, multi-slit H-band survey. This assessment, making use of semi-analytical models and state of the art N-body simulations, is a crucial input in the planning of dark energy missions such as the `Euclid` survey.

1.5 Outline

The outline of this thesis is as follows:

Chapter 2 reviews the basic physical processes in galaxy formation theory and describes the way these are implemented in the semi-analytical model `GALFORM`. This model is then used throughout the following chapters, so it represents the backbone of this thesis.

Chapter 3 describes the clustering properties of Lyα emitters as a result of combining `GALFORM` with the Millennium N-body simulation. Mock catalogues of Lyα surveys at different redshifts are constructed to test our model predictions against observations, and also to quantify the effect of cosmic variance on the clustering measurements of current and forthcoming surveys.

Chapter 4 studies the nature and evolution of Hα emitters and compares the performance of a future space-based near-IR redshift survey using either slitless spectroscopy of Hα emitters of multi-slit spectroscopy of H-band magnitude selected galaxies in measuring the BAO peak and the redshift distortions due to the peculiar velocities of galaxies.

Chapter 5 returns the focus to Lyα emitters and describes in detail the physics of Lyα radiative transfer. In addition, a Monte Carlo code is developed, and its accuracy and performance is tested against known solutions.

Chapter 6 combines the radiative transfer code with GALFORM using two different outflow models which allows us to study the relation between the Lyα luminosity and the predicted physical properties of these galaxies.

Finally, Chapter 7 summarizes the main results of this thesis and describe possible future directions of research.

Chapter 2

GALFORM: *A synthesis of galaxy formation theory*

2.1 Introduction

In this thesis we use the semianalytical model of galaxy formation GALFORM to study the properties of emission line galaxies in a cosmological setting. Given the large variety of physical processes incorporated in the model, here we briefly review the main aspects of galaxy formation theory and the way in which these are modelled in GALFORM.

The basic philosophy of the semianalytical technique is to use simplified calculations, or recipes, to account for all of the important processes which regulate the formation and evolution of galaxies. This allows the model to incorporate a large variety of phenomena from different areas of astrophysics, and also ensures the flexibility to develop and incorporate additional physical ingredients of varying complexity. In this way, semianalytical models can be regarded as a synthesis of many different techniques. Despite the simplifications made, semianalytical models are based on a well understood hierarchical clustering cosmology, in which dark matter haloes (gravitationally bound dark matter structures) host galaxies (White and Rees, 1978), and thus the formation and evolution of these structures determines the cosmic history of galaxies.

In short, the main physical ingredients that shape the formation and evolution of galaxies (and their observed properties) are: (i) The formation, evolution, merging histories and internal structure of dark matter haloes; (ii) The shock heating, radiative cooling and collapse of gas inside haloes leading to the formation of galactic disks; (iii) The formation of stars in the cold gas; (iv) The regulation of the star formation processes through different feedback mechanisms (e.g. supernovae, AGN); (v) The chemical enrichment of the ISM and the hot

gas; (vi) The merging history of galaxies; (vii) The formation of galactic spheroids and possible associated bursts of star formation; (viii) The spectrophotometric evolution of the stellar populations of galaxies; and (ix) The attenuation of the starlight due to the dust content of galaxies.

In the following we will describe the implementation of the above aspects of the galaxy formation process in GALFORM, with special focus on those that are of particular interest and relevance for this thesis.

2.2 Dark matter haloes

The formation and merging history of dark matter haloes is crucial in the process of galaxy formation. Of particular importance is the abundance of haloes and their growth history.

Press and Schechter (1974) developed a theory for predicting the abundance of dark matter haloes assuming these correspond to the overdense regions of a Gaussian random density field. Their famous expression for the number density of haloes in the mass range M to $M+\delta M$ is

$$\frac{\mathrm{d}n}{\mathrm{d}M} = \left(\frac{2}{\pi}\right)^{1/2} \frac{\rho_0}{M^2} \frac{\delta_c(t)}{\sigma(M)} \left|\frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln M}\right| \exp\left[-\frac{\delta_c^2(t)}{2\sigma^2(M)}\right], \tag{2.1}$$

where ρ_0 is the mean density of the Universe, $\delta_c(t)$ is the critical overdensity for a spherical top-hat fluctiation to collapse at time t, and the linear perturbation theory mass variance $\sigma(M)$ is determined from the power spectrum of density fluctuations,

$$\sigma^2(M) = \frac{1}{2\pi^2}\int_0^\infty P(k)\hat{W}_M^2(k)k^2\mathrm{d}k, \tag{2.2}$$

where $P(k)$ is the power spectrum, and $\hat{W}_M(k)$ is the Fourier transform of the real-space top-hat window function.

A statistical description of the merging of haloes was developed almost 20 years later (Bond et al., 1991; Bower, 1991; Lacey and Cole, 1993), extending the original Press-Schechter formulation. Lacey and Cole (1993) showed that the distribution of masses M_1 of the progenitors at redshift z_1 for a halo of mass M_2 at a redshift z_2 is given by

$$\frac{\mathrm{d}N}{\mathrm{d}M_1} = \left(\frac{2}{\pi}\right)^{1/2} \frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln M_1} M_2 \frac{\sigma_1^2}{M_1^2} \frac{\delta_{c1}-\delta_{c2}}{(\sigma_1^2-\sigma_2^2)^{3/2}} \times \exp\left[-\frac{(\delta_{c1}-\delta_{c2})^2}{(\sigma_1^2-\sigma_2^2)}\right], \tag{2.3}$$

where $\sigma_1 = \sigma(M_1), \sigma_2 = \sigma(M_2), \delta_{c1} = \delta_c(z_1), \delta_{c2} = \delta(z_2)$.

In semianalytical models, there are two main approaches to incorporate the statistical properties of dark matter haloes. In the first, rather direct approach, haloes and their merging histories can be extracted from an *N-body simulation* (e.g. Kauffmann et al., 1999; Hatton et al., 2003; Bower et al., 2006). N-body simulations follow the non-linear gravitational growth of structures to an accuracy limited principally by the mass resolution of the dark matter particles used. This could turn into a serious complication depending on the galaxy population of interest (or, more exactly, the typical mass of the haloes hosting those galaxies).

An alternative approach, developed prior to the incorporation of N-body merger trees into semianalytical models, consists of generating merger trees using a *Monte Carlo* algorithm, based on the distribution of progenitor halo masses given by Eq.(2.3). This generates a set of merger trees for haloes of a given mass at a given redshift, which can then be used to estimate the number density of galaxies being hosted by such haloes. The Monte Carlo approach has the main advantage of following the merger history with an arbitrary mass resolution, because the whole of the computer memory can be devoted to the history of one halo rather than to a population, as it could be the case with an N-body simulation. The mass resolution achieved is in general higher than the one used in N-body simulations. A comparison between the two methods shows that Monte Carlo merger trees reproduce the N-body results reasonably well (see, for example, Helly et al., 2003; Orsi et al., 2008).

GALFORM has the flexibility of using either method to incorporate the statistical properties of haloes.

2.2.1 Halo structure

The inner properties of dark matter haloes regulate quantities like the radiative cooling rate, the angular momentum of the gas that cools to form disks, and the sizes and rotation speeds of galaxies.

The angular momentum of a dark matter halo is usually quantified by the dimensionless spin parameter

$$\lambda_H = \frac{J_H |E_H|^{1/2}}{G M_H^{5/2}}, \tag{2.4}$$

where M_H, J_H and E_H are the total mass, angular momentum and energy of the halo, respectively. In GALFORM, each newly formed dark matter halo is randomly assigned a value of λ_H from a log-normal probability distribution, found to provide a good fit to the simulations of Cole and Lacey (1996). Haloes retain this property until they become part of a larger halo with mass $f_{\rm form}$ times its original mass, where its properties are computed afresh. The formation time of a halo is consistently computed to be the moment when a merger produces a halo with mass $f_{\rm form}$ times the formation mass of the largest progenitor involved in the merger. $f_{\rm form} = 2$ is usually assumed in GALFORM.

The mass distribution inside dark matter haloes has been found to follow a nearly universal form in CDM cosmologies. The standard choice in GALFORM is the so-called NFW profile (Navarro et al., 1997),

$$\rho(r) = \frac{\Delta_{\rm vir}\rho_{crit}}{f(a_{NFW})} \frac{1}{r/r_{\rm vir}(r/r_{\rm vir} + a_{NFW})^2}, \tag{2.5}$$

where

$$f(a_{NFW}) = \ln\left(1 + \frac{1}{a_{NFW}}\right) - \frac{1}{(1 + a_{NFW})}, \tag{2.6}$$

$$r_{\rm vir} = \left(\frac{3M}{4\pi\rho_{\rm crit}\Delta_{\rm vir}}\right)^{1/3}, \tag{2.7}$$

where $\rho_{\rm crit}$ is the critical density of the Universe. The profile is truncated at the virial radius $r_{\rm vir}$. The overdensity of a collapsed halo $\Delta_{\rm vir}$ in the spherical top-hat model has a value ≈ 200 for $\Omega = 1$, and depends on the cosmological parameters assumed (Lacey and Cole, 1993). The free parameter a_{NFW} (sometimes defined as the inverse of the concentration parameter) is a function of the halo mass and redshift, and its value is obtained by fitting equation (2.5) to N-body simulations (Navarro et al., 1996, 1997).

2.3 The cooling of gas inside haloes

From the evolution of the dark matter from small perturbations until it forms virialized structures, the baryonic gas follows a similar density distribution to the dark matter, since the gas is diffuse and cold so it cannot to undergo any radiative process. When structures become virialized, the gas inside the haloes will be shock heated to a temperature close to the virial temperature of the halo. This hot, ionized gas will begin to cool subsequently. As the gas loses energy through radiative processes, it will cool down and the removal of pressure support causes the gas to sink in the gravitational potential of the halo.

In GALFORM, the gas is assumed to be heated to the virial temperature of the halo, given by

$$T_{\rm vir} = \frac{1}{2}\frac{\mu m_H}{k} V_H^2, \tag{2.8}$$

where m_H is the mass of the hydrogen atom, μ is the mean molecular mass, V_H is the circular velocity of the halo, and k is the Boltzmann constant.

If we assume the gas inside the halo is isothermal and in hydrostatic equilibrium we can express the density profile of the gas as

$$\rho(r) \propto \exp\left(-\frac{\Phi}{c_T^2}\right), \tag{2.9}$$

$$c_T^2 \equiv \frac{kT}{\mu m_H}, \tag{2.10}$$

where $\Phi(r)$ is the gravitational potential. The above solution is, although simple, unphysical, since the density diverges at $r = 0$. When imposing boundary conditions at the centre the density profile obtained is found to be well approximated by a King profile,

$$\rho(r) \propto \frac{1}{[1+(r/r_0)^2]^{3/2}}, \tag{2.11}$$

where

$$r_0 = \frac{3c_T}{\sqrt{4\pi G\rho_0}}. \tag{2.12}$$

If both gas and dark matter are in static equilibrium within the same potential, and both the velocity dispersion of dark matter particles σ and the temperature of the gas T are independent of r, then it can be shown that

$$\rho_{\rm gas}(r) \propto [\rho_{\rm dm}(r)]^{\beta}, \tag{2.13}$$

where

$$\beta \equiv \frac{\mu m_H \sigma^2}{kT}. \tag{2.14}$$

Hence, unlike dark matter haloes, the gas is assumed to be distributed in a spherical distribution following a β-model density profile

$$\rho_{\rm gas}(r) \propto \frac{1}{(r^2 + r_{\rm core}^2)^{3\beta/2}}, \tag{2.15}$$

where $\beta = 2/3$ and $r_{\rm core} = r_{NFW}/3$. This density profile has been found to fit well the density profile of gas in hydrodynamical simulations of clusters (Navarro et al., 1995; Eke et al., 1998).

Once the halo has formed the hot gas inside will begin to cool radiatively. The cooling process is characterized by a *cooling time* defined as

$$\tau_{\rm cool} = \frac{E}{\dot{E}}, \tag{2.16}$$

$$= \frac{3kT_{\rm gas}}{2\mu m_H \rho_{\rm gas} \Lambda(T_{\rm gas}, Z_{\rm gas})}, \tag{2.17}$$

where E is the energy per unit mass of the gas, $\dot{E}$ is the rate at which the gas is radiating energy, $Z_{\rm gas}$ is the metallicity of the gas and $\Lambda(T_{\rm gas}, Z_{\rm gas})$ is the radiative cooling function tabulated by Sutherland and Dopita (1993). The amount of gas that has cooled in a time t after the halo has formed is given by defining a cooling radius $r_{\rm cool}(t)$ where $\tau_{\rm cool} = t$. This gas is assumed to be accreted on to a disk at the centre of the halo. The time it takes the cold gas to fall into the disk corresponds to the free-fall time $t_{\rm ff}$:

$$t_{\rm ff} = \int_0^r \left[\int_r^{r''} -\frac{2GM(r')}{r'^2} dr' \right]^{-1/2} dr''. \tag{2.18}$$

In the same way a free-fall radius $r_{\rm ff}$ is defined as the radius where, after a time t, the cold gas has time to fall to $r = 0$. Thus, only the cool gas within a radius given by $r_{\rm min} = \min[r_{\rm cool}, r_{ff}]$ will be added to the disk in a given time interval.

Recent work with hydrodynamical simulations has shown that a significant fraction of the gas in low mass haloes may never be shock-heated to the virial temperature of the halo (see, e.g. Fardal et al., 2001; Kereš et al., 2005). This results in the generation of filaments of cold gas flowing towards the centre of the halo where the gas could, eventually, be shock-heated.

Despite the apparent crude assumption made in semianalytical models that gas in all haloes is shock-heated to the virial temperature, it has been shown (Croton et al., 2006; Benson and Bower, 2010) that the details of this rapid mode of accretion in low mass haloes has little effect on the star formation properties of galaxies (even less when accounting for feedback mechanisms) since in this case the cooling is dominated by the free fall time rather than the cooling time assumed, making the detailed physics of the cooling of gas secondary in this case.

The kinematics of the hot halo gas is characterized by its effective rotational velocity $V_{\rm rot}$, defined as

$$V_{\rm rot} = A(a_{NFW})\lambda_H V_H, \tag{2.19}$$

where

$$V_H = \left(\frac{GM}{r_{\rm vir}}\right)^{1/2} \tag{2.20}$$

is the circular velocity of the halo at the virial radius, and $A(a_{NFW})$ is a dimensionless constant weakly dependent on a_{NFW}, with values of $A \approx 3.9$ for $a_{NFW} = 0.01$ to $A \approx 4.5$ for $a_{NFW} = 0.3$. In GALFORM, as the hot gas cools and collapses to a disk, it conserves its angular momentum. Since $V_{\rm rot}$ is constant, the specific angular momentum increases linearly with the radius of the halo.

2.4 Star formation and feedback processes

The cold gas that settles in a disk forms stars at a rate given by

$$\psi = \frac{M_{\rm cold}}{\tau_*}, \tag{2.21}$$

where the star formation time-scale is τ_*. GALFORM adopts a generic form for τ_*, broadly compatible with the data from Kennicutt (1998a),

$$\tau_* = \frac{1}{\epsilon_*}\left(\frac{V_d}{200{\rm kms}^{-1}}\right)^{\alpha_*}\tau_d, \tag{2.22}$$

where V_d is the disk circular velocity and τ_d the dynamical timescale of the disk ($\tau_d = \pi r_{\rm disk}/V_d$). Thus, the efficiency ϵ_* determines the fraction of gas turned into stars per dynamical time of a Milky Way like galaxy (i.e. one with $V_d \approx 200{\rm kms}^{-1}$). The two free parameters ϵ_* and α_* are constrained to reproduce a set of observational data. Alternatively,

the Baugh et al. (2005) model uses a slightly different form for the star formation time-scale, given by

$$\tau_* = \tau_{*0} \left(\frac{V_c}{200 \mathrm{kms}^{-1}} \right)^{\alpha_*}, \tag{2.23}$$

where $\tau_{*0} = 8$ Gyr and $\alpha_* = -3$. As discussed in Baugh et al. (2005), there is no strong theoretical argument to prefer one prescription over the other. Furthermore, both can reproduce the observed gas fraction-luminosity relation at $z = 0$ with the right choice of free parameters.

A longstanding issue in theories of galaxy formation is the so-called *overcooling* problem. In massive dark matter haloes, for example, gas can cool so efficiently that it would end up producing galaxies far more massive and luminous than any observed. In addition, the faint end of the observed luminosity function exhibits a shallower slope compared to the steeper halo mass function, suggesting a physical mechanism regulating the formation of dwarf galaxies. This motivates the idea of implementing mechanisms to heat the gas, preventing it from cooling so efficiently (see, for example, Benson et al., 2003).

In reality, the star formation process not only converts gas into stars but also affects the physical state of the surrounding gas. Supernovae, for example, inject energy into the interstellar medium, which reheats and ejects the cold gas slowing down the star formation process, hence acting as a *feedback* mechanism. The enrichment process will also affect the hot halo gas, decreasing the cooling times too by increasing the metallicity of the hot gas.

Other feedback mechanisms are also found to be important to regulate the star formation rate, such as the photo-ionization of the IGM and the energy release from Active Galactic Nuclei (AGN), which are powered by the accretion of mass onto a supermassive black hole. These are described below.

2.4.1 Photo-ionization heating

During the reionization epoch, photons produced by stars and AGNs will be emitted in a range of energies, some of which can to ionize abundant elements such as hydrogen and helium. Once reionization is mostly complete, this photoionizing radiation will reach the gas inside dark matter haloes and change the ionization balance, heating the gas and thus altering the rate at which gas can cool to be turned into stars.

In GALFORM, the cooling suppresion from photoionization is treated in a very simple way: Reionization is assumed to happen instantly at a given redshift $z_{\rm reio}$. The standard value is $z_{\rm reio} = 6$, although more recent papers of GALFORM assume $z_{\rm reio} = 10$ to match the results from the WMAP satellite (Kogut et al., 2003; Dunkley et al., 2009). After this redshift, the photoionization of the IGM completely suppresses the cooling and collapse of gas in haloes with circular velocities lower than $V_{\rm cut}$. In GALFORM this value is set to $V_{\rm cut} = 60{\rm kms}^{-1}$, to match closely the results of a detailed treatment of the effect of the IGM in the cooling of the gas in haloes by Benson et al. (2002). Other versions of GALFORM use $V_{\rm cut} = 30{\rm kms}^{-1}$ to match the result of more recent calculations (Hoeft et al., 2006; Okamoto et al., 2008).

2.4.2 Feedback from supernovae

The photoionization mode of feedback will affect mostly the abundance and properties of galaxies in low mass haloes. The heating of gas by supernovae affects a wider range of galaxies. The energy released by Type II supernovae $\dot{E}_{SN}$ generates an outflow of material at a rate $\dot{M}_{\rm ej}$ given by (e.g. Dekel and Silk, 1986)

$$\frac{1}{2}\dot{M}_{\rm ej}V_{\rm esc}^2 = \epsilon\int_0^t \psi(t')\dot{E}_{SN}(t-t')dt', \quad (2.24)$$

$$\approx \psi E_{SN}, \quad (2.25)$$

where we assume a fraction ϵ of this energy is released as kinetic energy in the form of an outflow to the ISM at the escape velocity. In the second step we have assumed that the supernova released the energy instantaneously (the energy release from Type II supernova is usually of the order of $10^{51}{\rm ergs}^{-1}$).

In GALFORM, the mass ejection rate from galaxies due to supernovae is parametrised with two components:

$$\dot{M}_{\rm ej} = \beta(V_c)\psi \quad (2.26)$$

$$= \left[\beta_{\rm reh}(V_c) + \beta_{SW}(V_c)\right]\psi, \quad (2.27)$$

where

$$\beta_{\rm reh} = \left(\frac{V_c}{V_{\rm hot}}\right)^{-\alpha_{\rm hot}}, \quad (2.28)$$

$$\beta_{SW} = f_{SW} \min\left[1, \left(\frac{V_c}{V_{SW}}\right)^{-2}\right]. \quad (2.29)$$

The factor $\beta_{\rm reh}$ is known as the *reheating term*, which quantifies how much gas is reheated and ejected into the galaxy halo, where is available to eventually cool down again and fall back into the disk. This term has a greater effect in low-mass galaxies, where $V_c < V_{\rm hot}$ (Benson et al., 2003).

The factor β_{SW} describes the *superwind* term, in which the gas is ejected out of the halo and it is not allowed to come back again (Baugh et al., 2005; Nagashima et al., 2005a). This mode of feedback affects high mass galaxies, where the ejection of gas increases the cooling time due to the decrease in gas density.

The parameters $\alpha_{hot}, V_{\rm hot}, f_{SW}$ and V_{SW} are chosen to match the observed present day galaxy luminosity function in the optical and near-IR, as well as the metallicity-luminosity relation. In general, different versions of GALFORM have employed different values for these parameters. In particular, the version of GALFORM presented in Bower et al. (2006) does not include the superwind mode of feedback and instead invokes the feedback from AGNs, which is described next.

2.4.3 Feedback from AGN

The properties and evolution of the supermassive black hole (SMBH) hosted in the centres of galaxies are known from observations to be correlated with the host galaxy. For example, a fundamental link is suggested by the observed correlation between the mass of the galactic bulge and the SMBH in the centre of galaxies (Magorrian et al., 1998). Moreover, the energy release associated with the build up of SMBHs at the centres of galaxies can have a significant impact on the formation of the host galaxy. The mechanism by which the AGN, formed by the accretion of material onto the central supermassive black hole of a galaxy, heats the hot halo gas and thus contributes to the galaxy formation process like other feedback mechanisms is still not well understood. For low accretion rates, AGNs are believed to form a thick accretion

disk surrounding the black hole which can lead to the formation of powerful jets, if the supermassive black hole is rapidly spinning. These jets could, in turn, heat the cool gas in the halo, thus reducing the cooling rates. In addition, radiatively driven winds might also contribute as a mode of *mechanical* feedback on the galaxy. A combination of these mechanisms could be crucial during the whole galaxy formation process.

The growth of supermassive black holes in GALFORM is the result of gas accretion driven by galaxy mergers, disk instabilities and mergers with other black holes (Malbon et al., 2007; Fanidakis et al., 2009). AGN feedback is assumed to be effective only in haloes undergoing quasi-hydrostatic cooling, where the cooling time at the cooling radius is greater (or comparable) to the free-fall time at the same radius. In other words, if

$$t_{\rm cool}(r_{\rm cool}) > \alpha_{\rm cool}^{-1} t_{\rm ff}(r_{\rm cool}), \tag{2.30}$$

(where $\alpha_{\rm cool}$ is typically of order unity), then the central AGN can quench the cooling flow into the galaxy. This will happens if the AGN power is greater than the cooling luminosity of the flow. The AGN power is assumed to be a fraction ϵ_{SMBH} of the Eddington luminosity of the black hole (the luminosity at which the inward gravitational pull equals the outward radiation force due to Thompson scattering). Then, if

$$L_{\rm cool} < \epsilon_{SMBH} L_{\rm Edd} \tag{2.31}$$

the hot halo is prevented from cooling.

2.5 Additional star formation mechanisms: Bulge formation

The star formation recipe described above converts the cold gas in a galactic disk into stars with a star formation rate regulated by several feedback mechanisms. An additional mechanism to generate stars in a galaxy is switched on when a galaxy merger occurs.

A natural consequence of a hierarchical structure formation model is the merger of dark matter haloes. When such an event occurs, the galaxies hosted by the individual haloes will become members of the new remnant halo and may eventually merge through the decay of their orbits throughout dynamical friction. This process is thought to be the responsible for the formation of elliptical galaxies when a *major merger* occurs (i.e. when two galaxies of

similar mass collide). During a *minor merger* (i.e. when one of the galaxies is significantly less massive than the other) the merger is expected to leave the disk of the larger galaxy in place (although somewhat thickened), while adding material to the bulge of the galaxy.

Mergers are important since they can trigger bursts of star formation (e.g. Mihos and Hernquist, 1994). GALFORM assumes that all major mergers will create such a starburst, by turning all the cold gas from both galaxies into stars in the bulge of the newly formed galaxy. The condition that defines a major merger in GALFORM is the ratio of the masses of the two galaxies. If the ratio of smaller to larger exceeds the critical value $f_{\rm ellip}$ (which is a model parameter), then the merger is considered to be major, and is minor otherwise.

When a minor merger occurs, the disk component of the larger galaxy is left unchanged, and the stellar disk of the small galaxy is added to the stellar spheroid of the primary. Some minor mergers can still produce a starburst, depending on whether the galaxy mass ratio exceeds the parameter $f_{\rm burst}$ (where, obviously $f_{\rm burst} < f_{\rm ellip}$), and the gas fraction in the large galaxy exceeds $f_{\rm gas,crit}$, since a large gas fraction makes the disk dynamically unstable.

Galaxy mergers are not the only way to form a spheroid. Internal processes driven by *disk instabilities* also lead to the formation of a spheroid. The basic idea is that when the disk becomes sufficiently massive that its self-gravity is dominant, it will become unstable to small perturbations by minor satellites or dark matter substructures, leading to the formation of bars. These can redistribute the mass and angular momentum of the disk and become, after a few dynamical times, dense central mass concentrations, thus forming a spheroid. Efstathiou et al. (1982) found a criterion to judge when a disk becomes unstable, using numerical simulations of exponential stellar disks. If

$$\epsilon_m \equiv v_{\rm max}\left(\frac{GM_{\rm disk}}{R_{\rm disk}}\right)^{-1/2} < 1.1, \tag{2.32}$$

then the disk becomes unstable. GALFORM computes the quantity ϵ_m, and if the criterion is met, then the disk is considered to be unstable. The mass of the disk is transfered to the bulge, and any gas present undergoes a starburst. Also, in the version of GALFORM presented in Bower et al. (2006), a fraction $F_{\rm bh}$ of the cold gas goes into feeding the black hole.

2.6 Galaxy sizes

GALFORM determines the size of a galactic disk by assuming conservation of angular momentum and centrifugal equilibrium. The bulge size, on the other hand, is computed by assuming virial equilibrium and energy conservation of the merging galaxies and the remnant. The calculation is complicated due to the gravitational interaction between the galaxy disk, spheroid and the surrounding dark matter halo. To overcome this, adiabatically invariant quantities are used to estimate the response of the halo to the forming galaxy, a formalism developed by Blumenthal et al. (1986).

In GALFORM, the specific angular momentum, $rV_c(r)$, is assumed to be invariant. A better invariant is obtained by taking into account the eccentricity of the orbit of dark matter particles in haloes. Gnedin et al. (2004) showed that a more accurate invariant in eccentric orbits is obtained by the combination $M(\bar{r})r$, where $\bar{r}$ is the orbit-averaged radius and $M(\bar{r})$ is the mass enclosed at this radius.

To compute the contraction of the halo due to the baryons condensing in the galaxy, the mass distribution is treated as spherical. The circular velocity of the system is, then

$$V_c^2(r) = G\frac{M_H(r) + M_D(r) + M_B(r)}{r}, \tag{2.33}$$

where $M_H(r), M_D(r)$ and $M_B(r)$ are the final mass halo profie, disk and bulge mass profiles interior to radius r respectively, and G is the gravitational constant. In addition, the final halo mass inside radius r is related to the original mass by

$$M_H(r) = f_H M_{H0}(r_0), \tag{2.34}$$

where $M_{H0}(r_0)$ is the mass of a shell at radius r_0 before the contraction process, r is the radius of the same shell after the contraction of the halo. f_H is the fraction of original hot gas mass that remained in the hot halo instead of forming part of the condensed gas in the centre. Combining Eqs. 2.33 and 2.34 we find

$$r_0 M_{H0}(r_0) = r[f_H M_{H0}(r_0) + M_D(r) + M_B(r)]. \tag{2.35}$$

To compute the final halo mass profile GALFORM assumes that disks are described by an exponential surface density profile, whereas bulges are described by a projected *de Vaucouleurs* $r^{1/4}$ law.

If the specific angular momentum of the disk is given by

$$j_D = k_D r_d V_{cD}(r_D), \tag{2.36}$$

(where $k_D = 1.19$ for an exponential disk), then the radius of the disk is related to the angular momentum by

$$\begin{aligned} j_D^2 &= k_D^2 r_D^2 V_{cD}^2(r_D) & (2.37)\\ &= k_D^2 G r_D \left[f_H M_{H0}(r_{D0}) + \frac{1}{2} k_h M_D + M_B(r_D) \right], & (2.38) \end{aligned}$$

where $k_h = 1.25$ is a factor which arises from the disk geometry (in the spherical approximation to compute the halo contraction $k_h = 1$). Finally, the disk half-mass radius must also satisfy Eq.(2.35), i.e.

$$r_{D0} M_{H0}(r_{D0}) = r_D \left[f_H M_{H0}(r_{D0}) + \frac{1}{2} M_D + M_B(r_D) \right]. \tag{2.39}$$

The last two coupled equations can be solved to find r_D. A similar procedure for the bulge, using the de Vaucouleurs profile, leads to another pair of coupled equations which are solved to find r_B.

When galaxies undergo a merger, the size of the newly formed bulge is computed in a different way. GALFORM assumes that the two merging components spiral together experiencing dynamical friction until their separation equals the sum of their half-mass radii. Applying energy conservation and the virial theorem leads to

$$\frac{(M_1 + M_2)^2}{r_{\rm new}} = \frac{M_1^2}{r_1} + \frac{M_2^2}{r_2} + \frac{f_{\rm orbit}}{c} \frac{M_1 M_2}{r_1 + r_2}, \tag{2.40}$$

where $f_{\rm orbit}$ and c are constants related to the mutual orbital and self-binding energy respectively, $f_{\rm orbit} = 1.0$ and $c = 0.5$. When the bulge is formed via disk instabilities, then the resulting spheroid size is computed using the same virial equilibrium and energy conservation arguments as before, leading to

$$\frac{cB(M_{\rm disk} + M_{\rm bulge})^2}{r_{\rm new}} = \frac{c_B M_{\rm bulge}^2}{r_{\rm bulge}} + \frac{c_D M_{\rm disk}^2}{r_{\rm disk}} + f_{\rm int} \frac{M_{\rm bulge} M_{\rm disk}}{r_{\rm bulge} + r_{\rm disk}}, \tag{2.41}$$

where $c_D = 0.49$, $c_B = 0.45$ are the form factors appropppiate for an exponential disk and $r^{1/4}$-law spheroid respectively, and $f_{\rm int} = 2.0$.

2.7 Chemical evolution

The first stars in the Universe (the so-called Population III) were formed from primordial gas, which is supposed to be practically metal free. Stellar evolution and the subsequent transfer of material from the interstellar medium to the intergalactic medium has a significant impact on the later generation of galaxies. Heavy elements will affect the cooling function, increasing the cooling rate, and will also lead to the formation of dust, which attenuates radiation, mostly in the UV but also in the optical part of the spectrum, re-emitting a fraction of it at IR wavelengths. It is clear then that an understanding of the chemical evolution is essential, both for the galaxy formation process and also to understand the observed properties of galaxies.

The fraction of material returned to the ISM by a stellar population as a function of time is given by

$$R(t) = \int_{M(t,Z)}^{M_u} [M - M_r(M,Z)]\phi(M)\frac{dM}{M}, \tag{2.42}$$

where $\phi(M)$ is the initial mass function (IMF), which quantifies the number of stars generated with a given mass per unit stellar mass, and $M_r(M)$ is the remnant mass of a star of initial mass M. Similarly, the *yield* of an element i is

$$p_i(t) = \int_{M(t,Z)}^{M_u} M_i(M_0,Z)\phi(M_0)\frac{dM_0}{M_0}, \tag{2.43}$$

where $M_i(M_0,Z)$ is the mass of metals produced by stars of initial mass M_0.

GALFORM simplifies the above equations adopting the *instantaneous recycling* approximation, in which mass and metals are returned to the interstellar medium instantaneously after the stars are produced, thus assuming that stellar evolution happens instantaneously. This makes $R(t) \equiv R$ and $p(t) \equiv p$. The approximation is reasonable when the ages of typical stellar populations in galaxies is of a few Gyrs. The values for R and p will depend on the choice of the IMF made (see Nagashima et al., 2005a,b, for further details).

The stellar initial mass function is commonly modelled as a power law with index x, i.e.

$$\frac{dN}{d\ln m} \equiv \phi(M) \propto m^{-x}, \tag{2.44}$$

Two choices for the IMF are commonly used in GALFORM, depending on the model variant. A common choice for quiescent star formation is the Kennicutt (1983) IMF, where the index x is

$$x = \begin{cases} 0.4, & \text{for } 0.15 < m/M_\odot < 1 \\ 1.5, & \text{for } 1 < m/M_\odot < 125 \end{cases} \tag{2.45}$$

Another option of GALFORM is to use a top-heavy IMF, in which $x = 0$ over the whole mass range. The top-heavy IMF is only used in the Baugh et al. (2005) model when a starburst occurs, and the Kennicutt (1983) IMF is used otherwise.

The instantaneous recycling approximation makes the calculation of the process of transfer of material between the hot gas, cold gas and stars in galaxies much simpler, since all of these are modelled simultaneously in GALFORM. This rather complex transfer of material is computed by solving the following set of differential equations:

$$\dot{M}_* = (1-R)\psi \tag{2.46}$$

$$\dot{M}_{\rm hot} = -\dot{M}_{\rm cold} + \beta\psi \tag{2.47}$$

$$\dot{M}_{\rm cold} = \dot{M}_{\rm cool} - (1-R+\beta)\psi \tag{2.48}$$

$$\dot{M}_*^Z = (1-R)Z_{\rm cold}\psi \tag{2.49}$$

$$\dot{M}_{\rm hot}^Z = -\dot{M}_{\rm cool}Z_{\rm hot} + (pe + \beta Z_{\rm cold})\psi \tag{2.50}$$

$$\dot{M}_{\rm cold}^Z = \dot{M}_{\rm cool}Z_{\rm hot} + [p(1-e) - (1+\beta-R)Z_{\rm cold}]\psi, \tag{2.51}$$

where $\dot{M}_*$ is the net formation rate of stars, $\dot{M}_{\rm hot}$ is the production rate of hot gas, $\dot{M}_{\rm cold}$ is the production rate of cold gas, $\dot{M}_*^Z$ is the production rate of mass in metals in stars, $\dot{M}_{\rm hot}^Z$ is the production rate of mass in metals in the hot gas, $\dot{M}_{\rm cold}^Z$ is the production rate of mass in metals in the cold gas, $Z_{\rm cold} = M_{\rm cold}^Z/M_{\rm cold}$ is the metallicity of the cold gas, $Z_{\rm hot} = M_{\rm hot}^Z/M_{\rm hot}$ is the metallicity of the hot gas, and e is the fraction of newly produced metals ejected directly from the stellar disk to the hot gas phase.

The solutions to the above set of coupled differential equations can be found in Cole et al. (2000). GALFORM computes the chemical evolution (and the resulting abundances of material in each mass phase, hot, cold and stellar) over each time step, assuming that the cooling rate $\dot{M}_{\rm cool}$ and the metallicity of the hot gas $Z_{\rm hot}$ can be taken to be constant.

2.8 The observed luminosities from galaxies

The calculation of the radiation emitted from galaxies (or the lack of it) is crucial to infer their physical properties. A theoretical model must also be able to reproduce a set of observational data in order to predict other, potentially observable properties. The calculation of the observed radiation can be thought of as a two step process. First, the intrinsic luminosity must be calculated. Then, one must compute how much of this radiation, after interacting with gas and dust from the ISM and IGM, will reach the observer. Both ideas are briefly explained next.

2.8.1 Stellar population synthesis

The spectral energy distribution (SED) of a galaxy is the luminosity emitted as a function of frequency (or wavelength). In the absence of attenuation, it can be thought of as a sum over the individual SEDs of all the stars in the galaxy. The luminosity $L_\nu^{\rm galaxy}$ of a galaxy at frequency ν can then be written as the following convolution

$$L_\nu^{\rm galaxy} = \int_0^t \int_0^\infty \dot{M}_*(t',Z') L_\nu^{\rm SSP}(t-t',Z',\phi) {\rm d}t' {\rm d}Z' \tag{2.52}$$

Since we already know $\dot{M}_*(t,Z)$, the problem is to determine $L_\nu^{\rm SSP}$, the SED of each stellar population in the galaxy. This depends on each individual star, and it can be written as

$$L_\nu^{\rm SSP}(t,Z,\phi) = \int_{M_{\rm min}}^{M_{\rm max}} \phi(M') L_\nu^{\rm star}(t,Z) {\rm d}M', \tag{2.53}$$

where $M_{\rm min}$ and $M_{\rm max}$ are the minimum and maximum mass of stars, respectively, and $L_\nu^{\rm star}$ is the SED of a single star. Several authors have computed Eq. (2.53) generating libraries for different stellar ages, metallicities and initial mass functions (e.g. Bruzual and Charlot, 2003; Maraston, 2005; Dotter et al., 2007; Lee et al., 2009). These are based on theoretical models for stellar evolution, observations of stars with known ages and metallicity, and theoretical models of stellar atmospheres.

GALFORM has the flexibility to use different stellar population synthesis models. When coupled to the star formation history of a galaxy predicted by GALFORM, the output is the SED of each simulated galaxy. This can be convolved with a filter transmission curve to obtain the amount of light passing through any desired band, thus obtaining magnitudes and colours for a sample of galaxies.

2.8.2 The role of dust

Simply put, the presence of dust in a galaxy attenuates its stellar luminosity. Dust grains absorb radiation, typically at short wavelengths, and as a result get heated by the absorbed energy, thus re-emiting energy at long wavelengths, mostly at infrared and submillimetre wavelengths. Models of galaxy formation should ideally model the effect of dust on the observed luminosity of a galaxy.

In the original version of GALFORM (Cole et al., 2000), the modelling of dust is made assuming that the mass of dust is proportional to the metallicity and mass of the ISM. Then, after giving to each galaxy a random inclination an extinction as a function of wavelength is computed using the results from a Monte Carlo radiative transfer calculation carried out by Ferrara et al. (1999).

A considerable improvement was made when combining GALFORM with the spectrophotometric code GRASIL (Silva et al., 1998), which is designed to compute the radiative transfer of star light through an idealised galactic geometry consisting of a disk and bulge, each of which may contain both diffuse and clumpy gas and dust. The code computes the absorption and the emission of the warm dust at long wavelengths, so it is ideal to study the properties of galaxies in the submillimetre and infrared part of the spectrum. The only drawback of the coupling of the two codes is the time performance, since GRASIL can take up to a few minutes to compute the dust properties of each galaxy, making computationally infeasible to run it over the whole population of galaxies predicted by GALFORM. Thus, a sub-sample of galaxies is made covering the whole range of properties of interest (Granato et al., 2000; Baugh et al., 2005; Lacey et al., 2008). In regions where the SED is smooth enough, a *shortcut* was designed to fit the output of GRASIL for the dust emission so to avoid running the code to get luminosities at those wavelengths. Unless specified, throughout this thesis we use the Cole

et al. (2000) approach to compute dust attenuation.

2.8.3 Emission-line modelling

Essential to this thesis is the modelling of emission lines. GALFORM computes the emission lines produced as the result of the recombination of ionized gas in the ISM, where young, massive stars, are the sources of ionizing photons. Thus, the intensity of the emission lines will be strongly dependent on the star formation rate of the galaxy. Other sources of ionizing radiation, such as AGNs, are not modelled.

After computing the SED of each galaxy, the production rate of Lyman continuum photons (those photons able to ionize a hydrogen atom) can be calculated by integrating the UV part of the spectrum,

$$\dot{n}_{Lyc} = \int_{\nu_0}^{\infty} \frac{L_\nu^{\rm galaxy}}{h\nu} {\rm d}\nu, \tag{2.54}$$

where $\nu_0 = 2.73 \times 10^{21}$Hz is the so-called Lyman limit frequency, i.e. the energy needed to ionize an hydrogen atom.

GALFORM assumes that all these photons are absorbed somewhere in the ISM of the galaxies by hydrogen atoms. Once the atoms absorb these photons, they will be ionized, but then rapidly recombined, experiencing a cascade of downward radiative transitions, each of which will emit photons as the atom returns to the ground state. The outcome of the process can be computed by solving the equation of statistical equilibrium inside an HI cloud,

$$n_p n_e \alpha_{nL}(T) + \sum_{n'>n}^{\infty} \sum_{L'} n_{n'L'} A_{n'L',nL} = n_{nL} \sum_{n''=1}^{n-1} \sum_{L''} A_{nL,n''L''}, \tag{2.55}$$

where n_p and n_e are the number densities of protons and electrons respectively, $\alpha_{nL}(T)$ is the recombination coefficient for the energy level nL at temperature T, n_{nL} is the number density of atoms in the energy level nL and $A_{n'L',nL}$ is the radiative transition probability between the energy levels $n'L'$ to nL. Furthermore, if b_{nL} accounts for the deviation from thermodynamic equilibrium, we can make use of the Saha and Boltzmann equations to write an expression for n_{nL}:

$$n_{nL} = b_{nL}(2L+1) \left(\frac{h^2}{2\pi m k T} \right)^{3/2} \exp(X_n/kT) n_p n_e, \tag{2.56}$$

where k is the Boltzmann constant, and X_n is the ionization potential of the level n. This equation can be substituted into Eq. (2.55) to find the values of each n_{nL} and b_{nL} in a downward iterative process, as detailed in Osterbrock (1989). Once these values are known, we can write the emission coefficient for each line as

$$j_{nn'} = \frac{h\nu_{nn'}}{4\pi} \sum_{L=0}^{n-1} \sum_{L'=L\pm1} n_{nL} A_{nL,n'L'}. \tag{2.57}$$

In general, the above situation is called *Case A* recombination, which is valid when the medium is optically thin for all HI resonance lines. However, for most of the observed astrophysical media this is not the case. When the medium is optically thick the transitions to the ground level are omitted, since whenever this happens, the resulting photon will be inmediately absorbed by an adjacent atom, and thus does not contributes to the energy balance. This is known as *Case B* recombination (Osterbrock, 1989).

Instead of going through the above formalism, GALFORM uses the values tabulated in Stasińska (1990) to get the number of photons of different emission lines per LC photon generated assuming case B recombination. This can then be used to calculate the intrinsic luminosity. Whenever it is relevant, the attenuation of the emission line by dust can be estimated as the attenuation of the continuum at the wavelength of the emission line. However, as we will discuss in the next chapter, resonance line radiation (such as Lyα emission) interacts with the gas in a much more complicated way due to the many scattering events these photons undergo, so the attenuation of Lyα inferred by the standard approach is not reliable. Other non-resonant lines, such as Hα, do not suffer from this problem, so their extinction can be calculated as the extinction suffered by the continuum at the wavelength of Hα.

2.9 Two versions of GALFORM

Throughout most of this thesis, the version of GALFORM we choose to use is the one described in Baugh et al. (2005), (see also Lacey et al., 2008), mainly because we will focus our predictions in the high redshift Universe, where this model has been shown to account correctly for the abundances and luminosity functions of submillimetre and Lyman break galaxies.

However, in Chapter 4 we focus our predictions on redshifts $z < 2$, so in this case we compare the model predictions of the Baugh et al. (2005) and Bower et al. (2006) models.

This motivates the need to summarise the main features and differences between both models (see a list of parameter values used in both models in table 2.1). These versions of GALFORM are based on the code developed by Cole et al. (2000), and then subsequently in Benson et al. (2003).

2.9.1 Key features of the Baugh et al. (2005) model

As explained before, the Baugh et al. (2005) model was motivated by the study of the high redshift galaxy population. Earlier versions of the code (Cole et al., 2000; Granato et al., 2000; Benson et al., 2003) could not account for the number counts and luminosity function of submillimetre and Lyman break galaxies, even when the dust attenuation (and re-emission) is computed using the radiative transfer code GRASIL (Silva et al., 1998).

The main feature of the model is the introduction of a top-heavy IMF for stars generated during a burst. This change has a large impact on the predicted number counts of submillimetre sources, since the production of UV photons is increased a factor 6 compared to the result of using a standard Kennicutt (1983) IMF. In addition, the boost in the generation of massive stars results in an increase of the yield of metals from Type II supernovae, which means that more dust is produced, thus increasing the infrared/submillimetre radiation as well.

Another important feature of the model is the modelling of bursts by minor galaxy mergers. Hernquist and Mihos (1995) showed that mergers between gas-rich disks and small satellites could trigger a burst of star formation. Thus, in the Baugh et al. (2005) model, for mergers where the ratio between the mass of the galaxies is smaller than $f_{\rm ellip} = 0.3$, but greater than $f_{\rm burst} = 0.05$, and the gas fraction of the bigger galaxy exceeds $f_{\rm gas,crit} = 0.75$ a minor merger occurs leading to the formation of a burst.

This model implements both supernova reheating and superwind modes of feedback, as opposed to the Bower et al. (2006) model (see below). Dark matter halo growth is taken from Monte Carlo merger trees to follow the merger histories, upon which galaxies are then formed. The cosmological parameters adopted correspond to the *concordance* cosmology, i.e.

Parameter	Baugh et al. (2005)	Bower et al. (2006)
Ω_m	0.30	0.25
Ω_Λ	0.70	0.75
Ω_b	0.040	0.045
h	0.70	0.73
σ_8	0.93	0.90
$V_{\rm cut}$[km s^{-1}]	60	50
$z_{\rm reio}$	6	6
τ_{*0}[Gyr]	8	-
ϵ_*[Gyr^{-1}]	-	0.028
α_*	-3.0	-1.5
$\alpha_{\rm hot}$	2.0	3.2
$V_{\rm hot,disk}$[km s^{-1}]	300	485
$V_{\rm hot,burst}$[km s^{-1}]	300	485
$f_{\rm SW,disk}$	2	0
$f_{\rm SW,burst}$	2	0
$V_{\rm SW,disk}$[km s^{-1}]	200	0
$V_{\rm SW,burst}$[km s^{-1}]	200	0
ϵ_{SMBH}	0	0.039
$\alpha_{\rm cool}$	0	0.58
$f_{\rm ellip}$	0.3	0.3
$f_{\rm burst}$	0.05	0.10
$f_{\rm gas,burst}$	0.75	0.10
R	0.40	0.40
p	0.02	0.02
$R_{\rm burst}$	0.91	0.40
$p_{\rm burst}$	0.15	0.02

Table 2.1: Summary of the main parameter values used in the Baugh et al. (2005) and Bower et al. (2006) models.

$\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.04$, $\sigma_8 = 0.93$ and $h = 0.7$.

2.9.2 Key features of the Bower et al. (2006) model

The subsequent published version of GALFORM, the Bower et al. (2006) model, introduced feedback due to AGNs to obtain the correct abundance of bright galaxies seen in the present day luminosity function.

The Bower et al. (2006) model thus incorporates a model for the formation and growth of black holes, as described in detail in Malbon et al. (2007) (see also Fanidakis et al., 2009). The black holes then power an AGN which can stop the cooling flow thus quenching the star formation following the process described in the previous sections.

Another feature of this model is the incorporation of disk instabilities (which are not included in the Baugh et al. 2005 model). Also, the superwind mode of feedback is not included in this model, since AGN feedback is already altering the cooling inside massive haloes. In addition, compared to the Baugh et al. (2005) model, only the IMF from Kennicutt (1983) is used for both quiescent and starburts. The implementation of an improved cooling algorithm results in a faster return of re-heated gas to the cold phase, as compared to the Baugh et al. (2005) model.

The merger trees used in the Bower et al. (2006) model were constructed from the Millennium simulation (Springel et al., 2005), using a technique similar to the one described in Helly et al. (2003) (see also Harker et al., 2007). For this reason, the cosmological parameters in this model correspond to those used in the Millennium simulation, which are slightly different from those used in the Baugh et al. (2005) model: $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.045$, $\sigma_8 = 0.9$ and $h = 0.73$.

Chapter 3

The Clustering of Lyα *emitters*

3.1 Introduction

The study of galaxies at high redshifts opens an important window on the process of galaxy formation and conditions in the early universe. The detection of populations of galaxies at high redshifts is one of the great challenges in observational cosmology. Currently three main observational techniques are used to discover high redshift, star-forming galaxies: (i) The Lyman-break drop-out technique, in which a galaxy is imaged in a combination of three or more optical or near-IR bands. The longer wavelength filters detect emission in the rest-frame ultraviolet from ongoing star formation, whereas the shorter wavelength filters sample the Lyman-break feature. Hence, a Lyman-break galaxy appears blue in one colour and red in the other (Steidel et al., 1996, 1999). By shifting the whole filter set to longer wavelengths, the Lyman-break feature can be isolated at higher redshifts; (ii) Sub-millimetre emission, due to dust being heated when it absorbs starlight (Smail et al., 1997; Hughes et al., 1998). The bulk of the energy absorbed by the dust comes from the rest-frame ultra-violet and so the dust emission is sensitive to the instantaneous star formation rate; (iii) Ly-α line emission from star forming galaxies, typically identified using either narrowband imaging (Hu et al., 1998; Kudritzki et al., 2000; Gawiser et al., 2007; Ouchi et al., 2008) or long-slit spectroscopy of gravitationally lensed regions (Ellis et al., 2001; Santos et al., 2004; Stark et al., 2007). The Ly-α emission is driven by the production of Lyman-continuum photons and so is dependent on the current star formation rate.

The Lyman-break drop-out and sub-millimetre detection methods are more established than Lyα emission as a means of identifying substantial populations of high redshift galaxies. Nevertheless, in the last few years there have been a number of Lyα surveys which have

successfully found high redshift galaxies e.g. (Hu et al., 1998; Kudritzki et al., 2000). The observational samples have grown in size such that statistical studies of the properties of Lyα emitters have now become possible: for example, the SXDS Survey (Ouchi et al., 2005, 2008) has allowed estimates of the luminosity function (LFs) and clustering of Lyα emitters in the redshift range $3 < z < 6$, and the MUSYC survey (Gronwall et al., 2007; Gawiser et al., 2007) has also produced clustering measurements at $z \sim 3$. Furthermore, the highest redshift galaxy ($z = 6.96$) robustly detected to date was found using the Lyα technique (Iye et al., 2006). Taking advantage of the magnification of faint sources by gravitational lensing, Stark et al. (2007) reported 6 candidates for Lyα emitters in the redshift range $8.7 < z < 10.2$, but these have yet to be confirmed. The DAzLE Project (Horton et al., 2004) is designed to find Lyα emitters at $z = 7.73$ and $z = 8.78$. However, the small field of view of the instrument (6.83‘ $\times$ 6.83‘) makes it difficult to use to study large scale structure (LSS) at such redshifts. On the other hand, the ELVIS Survey (Nilsson et al., 2007b,a) would appear to offer a more promising route to study the LSS of very high redshift galaxies ($z = 8.8$).

Despite these observational breakthroughs, predictions of the properties of star-forming Lyα emitting galaxies are still in the relatively early stages of development. Often these calculations employ crude assumptions about the galaxy formation process to derive a star formation rate and hence a Lyα luminosity, or use hydrodynamical simulations, which, due to the high computational overhead, study relatively small cosmological volumes. Haiman and Spaans (1999) made predictions for the escape fraction of Lyα emission and the abundance of Lyα emitters using the Press-Schechter formalism and a prescription for the dust distribution in galaxies. Radiative transfer calculations of the escape fraction have been made by Zheng and Miralda-Escudé (2002), Ahn (2004) and Verhamme et al. (2006) for idealized geometries, while Tasitsiomi (2006) and Laursen and Sommer-Larsen (2007) applied these calculations to galaxies taken from cosmological hydrodynamical simulations. Barton et al. (2004) and Furlanetto et al. (2005) calculated the number density of Ly-α emitters using hydrodynamical simulations of galaxy formation. Nagamine et al. (2006, 2008) used hydrodynamical simulations to predict the abundance and clustering of Lyα emitters. The typical computational boxes used in these calculations are very small ($\sim 10 - 30h^{-1}$Mpc), which makes it impossible to evolve the simulation accurately to $z = 0$. Hence, it is difficult to test if the galaxy formation model adopted produces a reasonable description of present day galaxies. Furthermore, the small box size means that reliable clustering predictions can only

be obtained on scales smaller than the typical correlation length of the galaxy sample. As we will show in this chapter, small volumes are subject to significant fluctuations in clustering amplitude.

The semi-analytical approach to modelling galaxy formation allows us to make substantial improvements over previous calculations of the properties of Lyα emitters. The speed of this technique means that large populations of galaxies can be followed. The range of predictions which can be made using semi-analytical models is, in general, broader than that produced from most hydrodynamical simulations, so that the model predictions can be compared more directly with observational results. A key advantage is that the models can be readily evolved to the present-day, giving us more faith in the ingredients used; i.e. we can be reassured that the physics underpinning the predictions presented for a high-redshift population of galaxies would not result in too many bright/massive galaxies at the present day.

The first semi-analytical calculation of the properties of Lyα emitters based on a hierarchical model of galaxy formation was carried out by Le Delliou et al. (2005). This is the model used throughout this work, which has been shown to be successful in predicting the properties of Lyα emitters over a wide range of redshifts. The semi-analytical model allows us to connect Lyα emission to other galaxy properties. Le Delliou et al. (2006) showed that this model succesfully predicts the observed Lyα LFs and equivalent widths (EWs), along with some fundamental physical properties, such as star formation rates (SFRs), gas metallicities, and stellar and halo masses. In Nilsson et al. (2007b), we used the model to make further predictions for the LF of very high redshift Lyα emitters and to study the feasibility of current and forthcoming surveys which aim to detect such high redshift galaxies. Kobayashi et al. (2007) developed an independent semi-analytical model to derive the luminosity functions of Lyα emitters.

The focus of this chapter is to use the model introduced by Le Delliou et al. (2005) to study the clustering of high-redshift Lyα emitting galaxies and to extend the comparison of model predictions with current observational data. Le Delliou et al. (2006) already gave an indirect prediction of the clustering of Lyα emitters by studying galaxy bias as a function of Lyα luminosity. However, these results depend on an analytical model for the halo bias (Sheth et al., 2001), and furthermore the linear bias assumption breaks down on small scales. Here we will present an explicit calculation of the clustering of galaxies by implementing the semi-analytical model on top of a large N-body simulation of the hierarchical clustering of the dark

matter distribution. This allows us to predict the spatial distribution of Lyα emitting galaxies, and to create realistic maps of Lyα emitters at different redshifts. These maps can be analysed with simple statistical tools to quantify the spatial distribution and clustering of galaxies at high redshifts. The N-body simulation used in this work is the *Millennium Simulation*, carried out by the Virgo Consortium (Springel et al., 2005). The simulation of the spatial distribution of Lyα emitters is tested by creating mock catalogues for different surveys of Lyα emitting galaxies in the range $3 < z < 9$. The clustering of Lyα emitters in our model is analysed with correlation functions and halo occupation distributions. Taking advantage of the large volume of the Millennium simulation, we also compute the errors expected on correlation function measurements from various surveys due to cosmic variance.

The outline of this chapter is as follows: Section 3.2 gives a brief description of the semi-analytical galaxy formation model and describes how it is combined with the N-body simulation. In Section 3.3 we establish the range of validity of our simulated galaxy samples by studying the completeness fractions in the model Lyα luminosity functions. Section 3.4 gives our predictions for the clustering of Lyα emitters in the range $0 < z < 9$. In Section 3.5 we compare our simulation with recent observational data and we also make predictions for future measurements (clustering and number counts) expected from the ELVIS Survey. Finally, Section 3.6 gives our conclusions.

3.2 The Model

We use the semi-analytical model of galaxy formation, `GALFORM`, to predict the properties of the Lyα emission of galaxies and their abundance as a function of redshift. The `GALFORM` model is fully described in Cole et al. (2000) (see also the review by Baugh, 2006) and the variant used here was introduced by Baugh et al. (2005) (see also Lacey et al., 2008, for a more detailed description). The model computes star formation histories for the whole galaxy population, following the hierarchical evolution of the host dark matter haloes.

As reviewed in Chapter 2, a critical assumption of the Baugh et al. model is that stars formed in starbursts have a top-heavy initial mass function (IMF), where the IMF is given by $\mathrm{dN/d\,ln(m)} \propto \mathrm{m}^{-\mathrm{x}}$ and $x = 0$. Stars formed quiescently in discs have a solar neighbourhood IMF, with the form proposed by Kennicutt (1983): $x = 0.4$ for $m < 1M_{\odot}$ and $x = 1.5$ for $m > 1M_{\odot}$. Both IMFs cover the mass range $0.15M_{\odot} < m < 125M_{\odot}$. Within the framework of

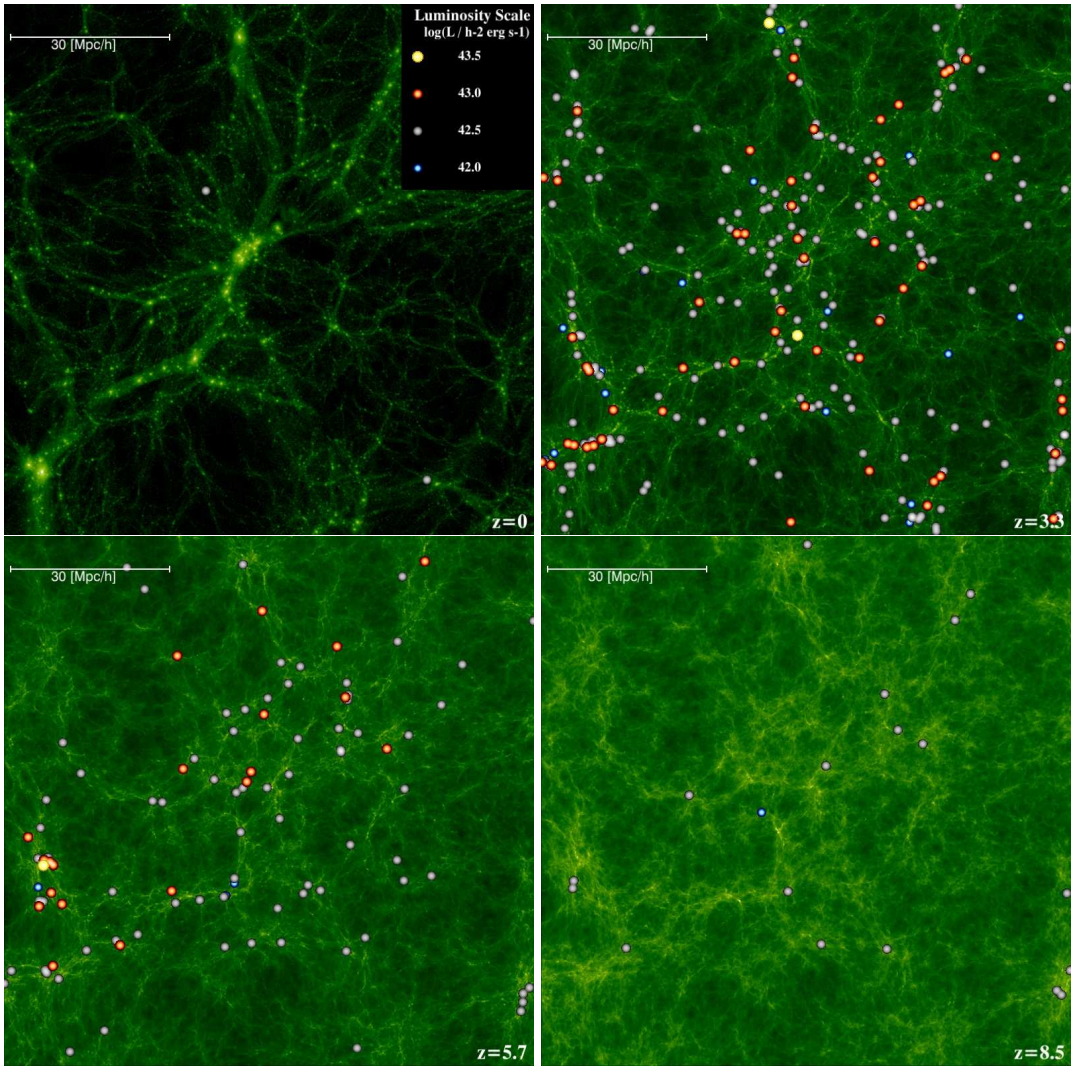

Figure 3.1: The spatial distribution of Lyα emitting galaxies (coloured circles) in a slice from the Millennium simulation, with the dark matter distribution in green. The four panels are for redshifts in the range $0 < z < 8.5$, as indicated in each panel. The colour of the circles changes with the Lyα luminosity of the galaxies, as shown in the key in the upper-right corner of the first panel. Only galaxies brighter than $\log(L_{\mathrm{Ly}\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) = 42.2$ are plotted. Each image covers a square region $100 \times 100 h^{-1}\mathrm{Mpc}^2$ across and having a depth of $10 h^{-1}\mathrm{Mpc}$, which is less than one thousandth the volume of the full simulation box.

ΛCDM, Baugh et al. argued that the top-heavy IMF is essential to match the counts and redshift distribution of galaxies detected through their sub-millimetre emission, whilst retaining the match to galaxy properties in the local Universe, such as the optical and far-IR luminosity functions and galaxy gas fractions and metallicities. Nagashima et al. (2005a,b) showed that such a top-heavy IMF also results in predictions for the metal abundances in the intra-cluster medium and in elliptical galaxies in much better agreement with observations. Lacey et al. (2008) showed that the same model predicts galaxy evolution in the IR in good agreement with observations from *Spitzer*, and also discussed independent observational evidence for a top-heavy IMF.

The model used to predict the luminosities and equivalent widths of the Lyα galaxies is identical to that described in Le Delliou et al. (2005, 2006). The Lyα emission is computed by the following procedure: (i) The integrated stellar spectrum of the galaxy is calculated, based on its star formation history, including the effects of the distribution of stellar metallicities, and taking into account the IMFs adopted for different modes of star formation. (ii) The rate of production of Lyman continuum (Lyc) photons is computed by integrating over the stellar spectrum, and assuming that all of these ionizing photons are absorbed by neutral hydrogen within the galaxy. We calculate the fraction of Lyα photons produced by these Lyc photons, assuming Case B recombination (Osterbrock, 1989).

(iii) The observed Lyα flux depends on the fraction of Lyα photons which escape from the galaxy ($f_{\rm esc}$), which is assumed to be constant and independent of galaxy properties.

Calculating the Lyα escape fraction from first principles by following the radiative transfer of the Lyα photons is very demanding computationally. A more complete calculation of the escape fraction would have into account the structure and kinematic properties of the intestellar medium (ISM) (Zheng and Miralda-Escudé, 2002; Ahn, 2004; Verhamme et al., 2006). In this model, we adopt the simplest possible approach, which is to fix the escape fraction, $f_{\rm esc}$, to be the same for each galaxy, without taking into account its dust properties. This results in a surprisingly good agreement between the predicted number counts and luminosity functions of emitters and the available observations at $3 \lesssim z \lesssim 7$ (Le Delliou et al., 2005, 2006). Le Delliou et al. (2005) chose $f_{\rm esc} = 0.02$ to match the number counts at $z \approx 3$ at a flux $f \approx 2 \times 10^{-17}$ergcm^{-2}s^{-1}. The same value is used in this work. This value for the Lyα escape fraction seems very small, but is consistent with direct observational estimates for low redshift galaxies: Atek et al. (2008) derive escape fractions for a sample of nearby star-

forming galaxies by combining measurements of Lyα, Hα and Hβ, and find that most have escape fractions of 3% or less. Le Delliou et al. (2006) also showed that if a standard solar neighbourhhod IMF is adopted for all modes of star formation, then a substantially larger escape fraction would be required to match the observed counts of Lyα emitters, and even then the overall match would not be as quite good as it is when the top-heavy IMF is used in bursts.

Once we obtain the galaxy properties from the semi-analytical model, we plant these galaxies into a N-body simulation, in order to add information about their positions and velocities. The simulation used here is the *Millennium Simulation* (Springel et al., 2005). This simulation adopts concordance values for the parameters of a flat ΛCDM model, $\Omega_{\rm m} = 0.25$ and $\Omega_{\rm b} = 0.045$ for the densities of matter and baryons at $z = 0$, $\rm h = 0.73$ for the present-day value of the dimensionless Hubble constant, $\sigma_8 = 0.9$ for the *rms* linear mass fluctuations in a sphere of radius $8\rm h^{-1}Mpc$ at $z = 0$ and $n = 1$ for the slope of the primordial fluctuation spectrum. The simulation follows 2160^3 dark matter particles from $z = 127$ to $z = 0$ within a cubic region of comoving length $500\rm h^{-1}Mpc$. The individual particle mass is $8.6\times10^8\rm h^{-1}M_\odot$, so the smallest dark halo which can be resolved has a mass of $2 \times 10^{10}\rm h^{-1}M_\odot$.

Dark matter haloes are identified using a *Friends-Of-Friends* (FOF) algorithm. To populate the simulation with galaxies from the semi-analytical model, we use the same approach as in Benson et al. (2000). First, the position and velocity of the centre of mass of each halo is recorded, along with the positions and velocities of a set of randomly selected dark matter particles from each halo. Second, the list of halo masses is fed into the semi-analytical model in order to produce a population of galaxies associated with each halo. Each galaxy is assigned a position and velocity within the halo. Since the semi-analytical model distinguishes between central and satellite galaxies, the central galaxy is placed at the centre of mass of the halo, and any satellite galaxy is placed on one of the randomly selected halo particles. Once galaxies have been generated, and positions and velocities have been assigned, it is a simple process to produce catalogues of galaxies with spatial information and any desired selection criteria.

The combination of the semi-analytical model with the N-body simulation is essential to study the detailed clustering of a desired galaxy population, although the clustering amplitude on large scales can also be estimated analytically (Le Delliou et al., 2006). An example of the output of the simulation is shown in the four images of Fig. 3.1 which show redshifts

$z = 0$, $z = 3.3$, $z = 5.7$ and $z = 8.5$. The dark matter distribution (shown in green) becomes smoother as we go to higher redshifts, due to the gravitational growth of structures. As shown in Fig. 3.1, for this particular luminosity cut, the number density of Lyα emitters varies at different redshifts. As we will show in the next section, these catalogues at high redshift are not complete at faint luminosities, so we have to restrict our predictions to brighter luminosities as we go to higher redshifts.

3.3 Luminosity Functions

The model presented by Le Delliou et al. (2005, 2006) differs in two main ways from the one presented in this work: (i) there is a slight difference in the values of the cosmological parameters used, and (ii) the earlier work used a grid of halo masses together with an analytical halo mass function, rather than the set of haloes from an N-body simulation. In §3.3.1, we investigate the impact of the different choice of cosmological parameters on the luminosity function of Lyα emitters, to see if the very good agreement with observational data obtained by Le Delliou et al. (2005, 2006) is retained on adopting the Millennium cosmology. In §3.3.2, we assess the completeness of our samples of Lyα emitters due to the finite mass resolution of the Millennium simulation.

3.3.1 Comparison of model predictions with observed luminosity functions

In this section, we investigate the impact on the model predictions of the choice of cosmological parameters by re-running the model of Le Delliou et al. (2005, 2006), keeping the galaxy formation parameters the same but changing the cosmological parameters to match those used in the Millennium simulation. To recap, the original Le Delliou et al. (2006) model used $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_{\rm b} = 0.04$, $\sigma_8 = 0.93$ and $h = 0.7$. In Fig. 3.2, we compare the cumulative luminosity functions obtained with `GALFORM` for the two sets of cosmological parameters with current observational data in the redshift range $3 < z < 7$. The observational data are taken from: Kudritzki et al. (2000) (crosses), Cowie and Hu (1998) (asterisks), Gawiser et al. (2007) (diamonds), Ouchi et al. (2008) (triangles and squares) in the $z = 3.3$ panel; Ajiki et al. (2003) (pluses), Maier et al. (2003) (asterisks), Hu et al. (2004) (diamonds), Rhoads et al. (2003) (triangles), Shimasaku et al. (2006) (squares) and Ouchi et al. (2008) (crosses) in the $z = 5.7$ panel; and Taniguchi et al. (2005) (crosses) and Kashikawa et al. (2006) (as-

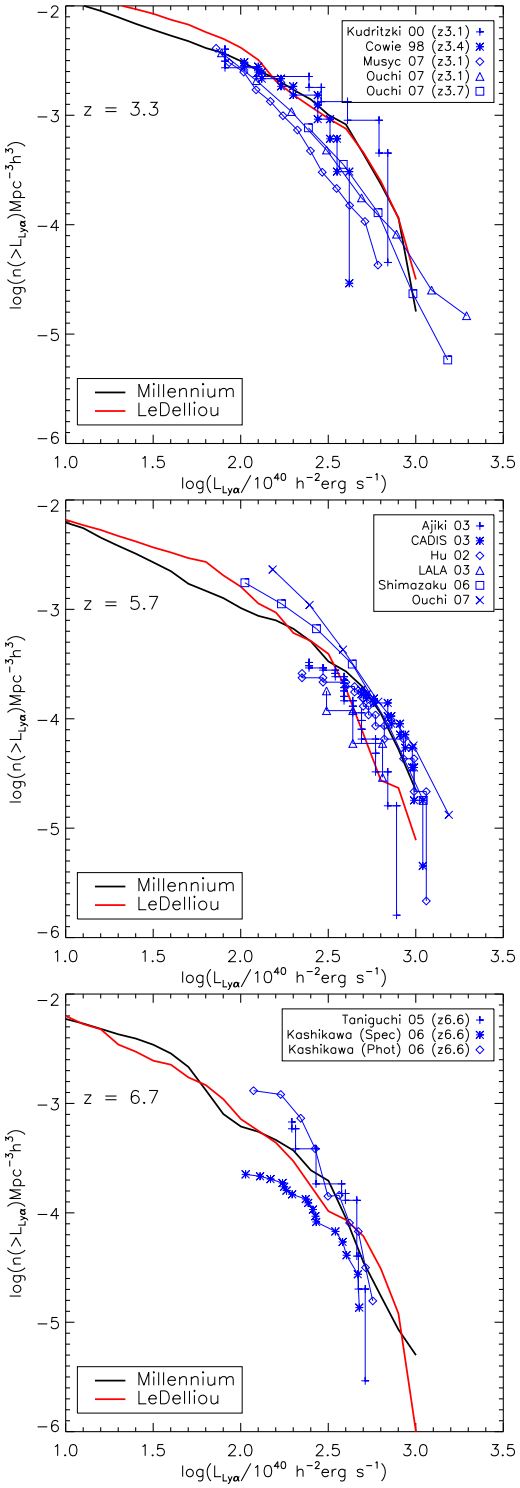

Figure 3.2: The cumulative luminosity functions of Lyα emitters at redshifts $z = 3.3$ (Top), $z = 5.7$ (Center) and $z = 6.7$ (Bottom). The blue points correspond to observational data (as indicated by the key with full references in the text). The black and red curves correspond, respectively, to the GALFORM predictions using the cosmological parameters of the Millennium Simulation and those adopted in Le Delliou et al.

terisks and diamonds) in the $z = 6.7$ panel. At $z = 3.3$, the two model curves agree very well, and are consistent with the observational data shown. At $z = 5.72$, the two models do not match as well as in the previous case, but both are still consistent with the observational data. Finally, at $z = 6.7$ the differences are small and both curves are consistent with observational data. The conclusion from Fig. 3.2 is that there is not a significant change in the model predictions on using these slightly different values of the cosmological parameters. Furthermore, the observational data are not yet sufficiently accurate to distinguish between the two models or to motivate the introduction of further modifications to improve the level of agreement, such as using a different Lyα escape fraction.

3.3.2 The completeness of the Millennium galaxy catalogues

The Millennium simulation has a halo mass resolution limit of $1.72 \times 10^{10} h^{-1} M_{\odot}$. In a standard GALFORM run, a grid of haloes which extends to lower mass haloes than the Millennium resolution is typically used, with $M_{\rm res} = 5 \times 10^{9} h^{-1} M_{\odot}$ at $z = 0$. A fixed dynamic range in halo mass is adopted in these runs, but with the mass resolution shifting to smaller masses with increasing redshift: for our standard setup, we have $M_{\rm res} = 7.8 \times 10^{7} h^{-1} M_{\odot}$ and $1.4 \times 10^{7} h^{-1} M_{\odot}$ at $z = 3$ and 6 respectively. Therefore, when putting GALFORM galaxies into the Millennium, our sample does not contain galaxies which formed in haloes with masses below the resolution limit of the Millennium. This introduces an incompleteness into our catalogues when compared to the original GALFORM prediction. The incompleteness of the galaxy catalogues is more severe for low luminosity galaxies because they tend to be hosted by low mass haloes, as will be shown in the next section. Hereafter, we will use *N-body sample* to refer to the GALFORM galaxies planted in the Millennium haloes, to distinguish them from the *pure* GALFORM catalogues generated using a grid of halos masses.

In order to quantify the incompleteness of the N-body sample as a function of luminosity, we define the completeness fraction as the ratio of the cumulative luminosity function for the N-body sample to that obtained for a pure GALFORM calculation, and look for the luminosity at which the completeness fraction deviates from unity. The panels of Fig. 3.3 give different views of the completeness of the N-body samples. The top panel shows the luminosity above which a catalogue can be considered as complete: we define the completeness limit as the luminosity at which the completeness fraction first drops to 0.85. The figure clearly shows how the luminosity corresponding to this completeness limit becomes progressively brighter

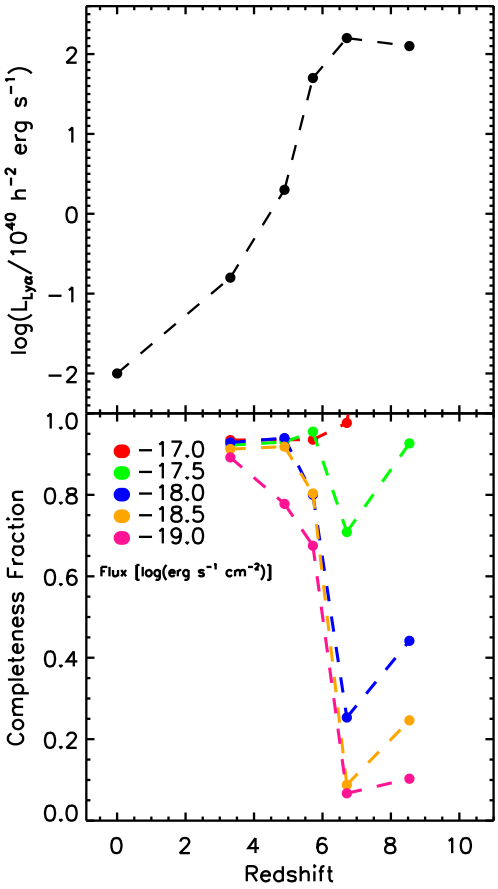

Figure 3.3: Completeness of the Millennium galaxy catalogues with respect to Lyα luminosity or flux. (Top): The minimum luminosity down to which the catalogues are 85% complete. (Bottom): The completeness fraction as a function of redshift for a range of fluxes $-19 < log(F_{\mathrm{Ly}\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) < -17$, as indicated by the key.

as we move to higher redshifts. For $z > 9$ the N-body sample is incomplete at all luminosities plotted.

The bottom panel of Fig. 3.3 shows how the sample becomes more incomplete at any redshift as we consider fainter fluxes. A sample with galaxies brighter than $\log(F_{\mathrm{Ly}\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -19$) is less than 70% complete at all redshifts $z > 5$, while a sample with galaxies brighter than $\log(F_{\mathrm{Ly}\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -17$ is always over 90% complete for $z < 9$. The completeness fraction monotonically decreases with increasing redshift until $z \sim 6$ for very faint fluxes. For $z > 6$ the completeness rises again: the shape of the bright end of the luminosity function at this redshift is sensitive to the choice of the redshift of reionization. Here we use $z_{\mathrm{reio}} = 10$.

In summary, the requirement that our samples be at least 80% complete restricts the

range of validity of the predictions from the Millennium simulation to redshifts below 9, and fluxes brighter than $\log(F_{\mathrm{Ly}\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -17.5$.

3.4 Clustering Predictions

In this section we present clustering predictions using Lyα emitters in the full Millennium volume. To study the clustering of galaxies we calculate the two-point correlation function, $\xi(r)$, of the galaxy distribution. In order to quantify the evolution of the clustering of galaxies, we measure the correlation function over the redshift interval $0 < z < 9$.

To calculate $\xi(r)$ in the simulaation, we use the standard estimator (e.g. Peebles 1980):

$$1 + \xi(r) = \frac{\langle DD \rangle}{\frac{1}{2} N_{\mathrm{gal}} n \Delta V(r)}, \tag{3.1}$$

where $\langle DD \rangle$ stands for the number of distinct data pairs with separations in the range r to $r + \Delta r$, n is the mean number density of galaxies, N_{gal} is the total number of galaxies in the simulation volume and $\Delta V(r)$ is the volume of a spherical shell of radius r and thickness Δr. This estimator is applicable in the case of periodic boundary conditions. In the correlation function analysis, we consider two parameters which help us to understand the clustering behaviour of Lyα galaxies: the correlation length, r_0, and the galaxy bias, b, both of which are discussed below.

3.4.1 Correlation Length evolution

A common way to characterize the clustering of galaxies is to fit a power-law to the correlation function:

$$\xi(r) = \left(\frac{r}{r_0} \right)^{-\gamma}, \tag{3.2}$$

where r_0 is the correlation length and $\gamma = 1.8$ gives a good fit to the slope of the observed correlation function over a restricted range of pair separations around r_0 at $z = 0$ (e.g. Davis and Peebles (1983)). The correlation length can also be defined as the scale where $\xi = 1$, and quantifies the amplitude of the correlation function when the slope γ is fixed.

Fig. 3.4 shows the correlation function of Lyα emitting galaxies, ξ_{gal} (solid black curves) of the full catalogues down to the completeness limits at each redshift, calculated using Eq. (3.1). The red curve shows ξ_{dm}, the correlation function of the dark matter. At $z = 0$,

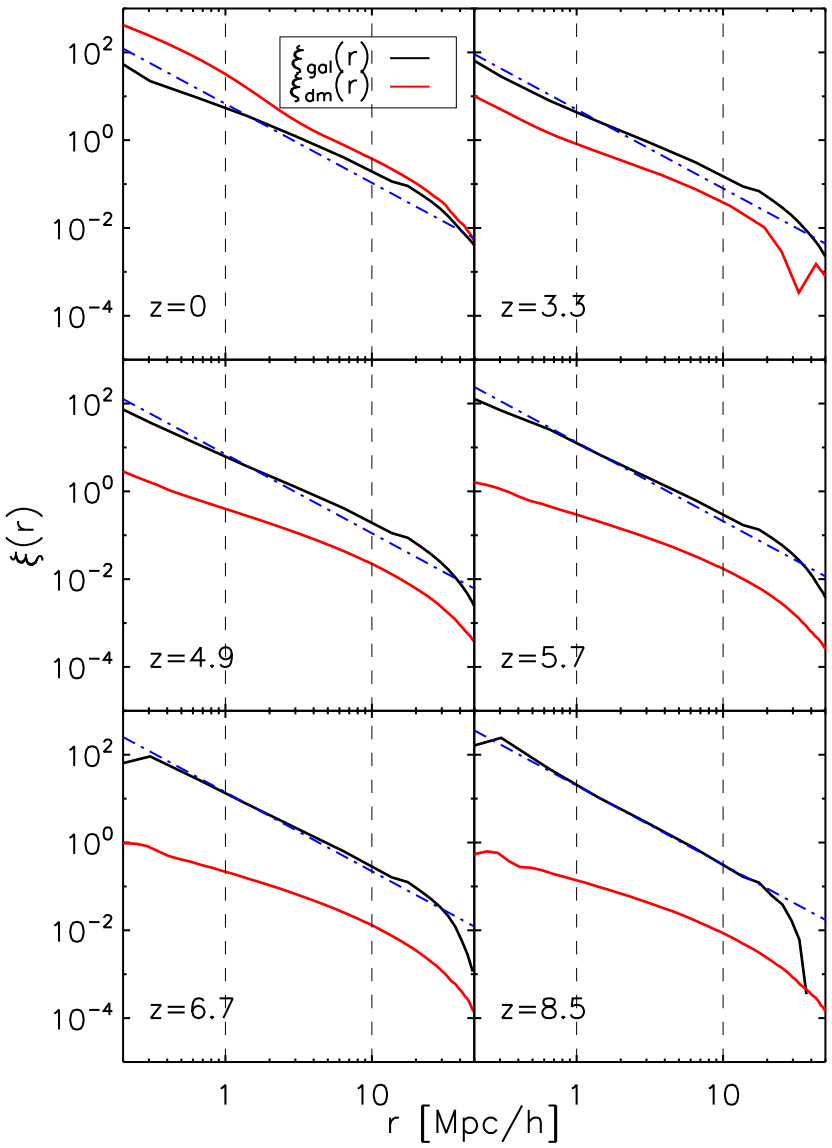

Figure 3.4: The correlation function predicted for Lyα emitters (black solid curve) for a range of redshifts, as indicated in each panel. Lyα emitters are included down to the completeness limit at each redshift shown in Fig 3.3. The solid red curve shows the correlation function of the dark matter at the same epochs. The blue dashed line shows the power law fit of Eq. (3.2), evaluated in the range $1 < r\,[\mathrm{Mpc/h}] < 10$, as delineated by the vertical dashed lines.

$\xi_{\rm dm}$ is larger than $\xi_{\rm gal}$, but for $z > 0$ $\xi_{\rm dm}$ is increasingly below $\xi_{\rm gal}$. We will study in detail the comparison of the dark matter and Lyα galaxy correlation functions in §4.2.

Another notable feature of Fig. 3.4 is that $\xi_{\rm gal}(r)$ differs considerably from a power law, particularly on scales greater than 10 h^{-1}Mpc. When fitting Eq. (3.2) to the correlation functions plotted in Fig. 3.4, we use only the measurements in the range [1,10] h^{-1}Mpc, where $\xi_{\rm gal}(r)$ behaves most like a power law. We fix the slope $\gamma = 1.8$ for all $\xi_{\rm gal}(r)$ to allow a comparison between different redshifts, although we note that for $z < 5$, the slope of $\xi_{\rm gal}(r)$ is closer to $\gamma = 1.6$. By using the power law fit we can compare the clustering amplitudes of different galaxy samples. To determine the clustering evolution of Lyα emitters, we split the catalogues of Lyα emitters into luminosity bins. For each of these sub-samples, we calculate the correlation function and then we obtain r_0 by fitting Eq. (3.2) as described. Fig. 3.5 (top) shows the dependence of r_0 on luminosity for different redshifts in the range $0 < z < 9$. The errors are shown by the area enclosed by the thin solid lines for each set of points, and are calculated as the 90% confidence interval of the χ^2 fit of the correlation functions to Eq. (3.2) (ignoring any covariance between pair separation bins). The range of luminosities plotted is set by the completeness limit of the simulation described in the previous section. We also discard galaxy samples with fewer than 500 galaxies, as in such cases, the errors are extremely large and the correlation functions are poorly defined. The clustering in high redshift surveys of Lyα emitters is sensitive to the flux limit that they are able to reach, as shown by Fig. 3.5.

The model predictions show modest evolution of r_0 with redshift for most of the luminosity range studied. Over this redshift interval, on the other hand, the correlation length of the dark matter changes dramatically, as shown by Fig. 3.4. Typically, at a given redshift, we find that r_0 shows little dependence on luminosity until a luminosity of $L_{\rm Ly\alpha} \sim 10^{42}$[erg s^{-1} h^{-2}] is reached, brightwards of which there is a strong increase in clustering strength with luminosity. This trend is even more pronounced at higher redshifts. Galaxies at $z = 0$ are less clustered than galaxies in the range $3 < z < 7$, except at luminosities close to $L_{\rm Ly\alpha} \sim 10^{40}$[erg s^{-1} h^{-2}]. At $z = 8.5$, r_0 increases from $r_0 \sim 5$ h^{-1}Mpc at $L_{\rm Ly\alpha} \sim 10^{42}$[erg s^{-1} h^{-2}] to $r_0 \sim 12$ h^{-1}Mpc at $L_{\rm Ly\alpha} > 10^{42.5}$[erg s^{-1} h^{-2}].

The growth of r_0 with limiting luminosity is related to the masses of the haloes which host Lyα galaxies. As shown in the bottom panel of Fig. 3.5, there is not a simple relation between the median mass of the host halo and the luminosity of Lyα emitters. For a given

luminosity, Lyα galaxies tend to be hosted by haloes of smaller masses as we go to higher redshifts. In addition, for all redshifts but $z = 0$, there is a trend of more luminous Lyα emitters being found in more massive haloes. The key to explaining the trends in clustering strength is to compare how the effective mass of the haloes which host Lyα emitting galaxies is evolving compared to the typical or characteristic mass in the halo distribution (M_*) (Mo and White, 1996); if Lyα emitters tend to be found in haloes more massive than M_*, then they will be more strongly clustered than the dark matter. This difference between the clustering amplitude of galaxies and mass is explored more in the next section. In a hierarchical model for the growth of structures, haloes more massive than M_* are more clustered, and thus we expect a strong connection between the evolution of r_0 and the masses of the halos. Fig. 3.5 shows that the dependence of r_0 (and host halo mass) on luminosity becomes stronger at higher redshifts.

3.4.2 The bias factor of Lyα emitters

The galaxy bias, b, quantifies the strength of the clustering of galaxies compared to the clustering of the dark matter. One way to calculate the bias is by taking the ratio of $\xi_{\rm gal}$ and $\xi_{\rm dm}$, $\xi_{\rm gal} = b^2 \xi_{\rm dm}$. Both correlation functions are estimated using Eq. (3.1). Since the simulation contains ten billion dark matter particles, a direct pair-count calculation of $\xi_{\rm dm}$ would demand a prohibitively large amount of computer time, so we extract dilute samples of the dark matter particles, selecting randomly $\sim 10^7$ particles. In this way we only enlarge the pair-count errors on $\xi_{\rm dm}$ (which nevertheless are still much smaller than for $\xi_{\rm gal}$) but obtain the correct amplitude of the correlation function itself.

To obtain the bias parameter of Lyα emitters as a function of luminosity, we split the full catalogue of galaxies at each redshift into luminosity bins. For each of these bins we calculate $\xi_{\rm gal}$ and divide by $\xi_{\rm dm}$ to get the square of the bias. Due to non-linearities, the ratio of $\xi_{\rm gal}$ and $\xi_{\rm dm}$ is not constant on all scales. As a reasonable estimation of the bias we chose the mean value over the range 6 h^{-1}Mpc $< r <$ 30 h^{-1}Mpc. Over these scales the bias does seem to be constant and independent of scale. This range is quite similar to the one used by Gao et al. (2005) to measure the bias parameter of dark matter haloes in the Millennium Simulation.

The bias parameter can also be calculated approximately using various analytical formalisms (Mo and White, 1996; Sheth et al., 2001; Mandelbaum et al., 2005). These proce-

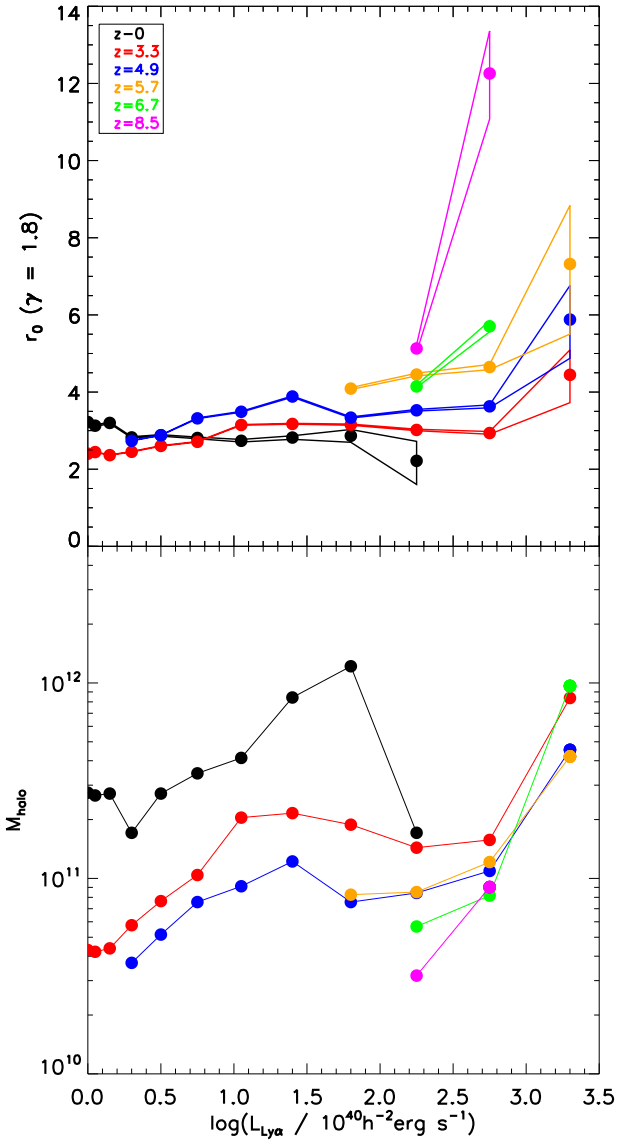

Figure 3.5: *(Top):* The evolution of the correlation length r_0 as a function of Lyα luminosity for several redshifts in the range $0 < z < 9$, as indicated by the key. The thin solid coloured lines shows the errors on the correlation length. *(Bottom):* The evolution of the median mass of halos which host Lyα emitting galaxies as a function of Lyα luminosity, for the same range of redshifts as above.

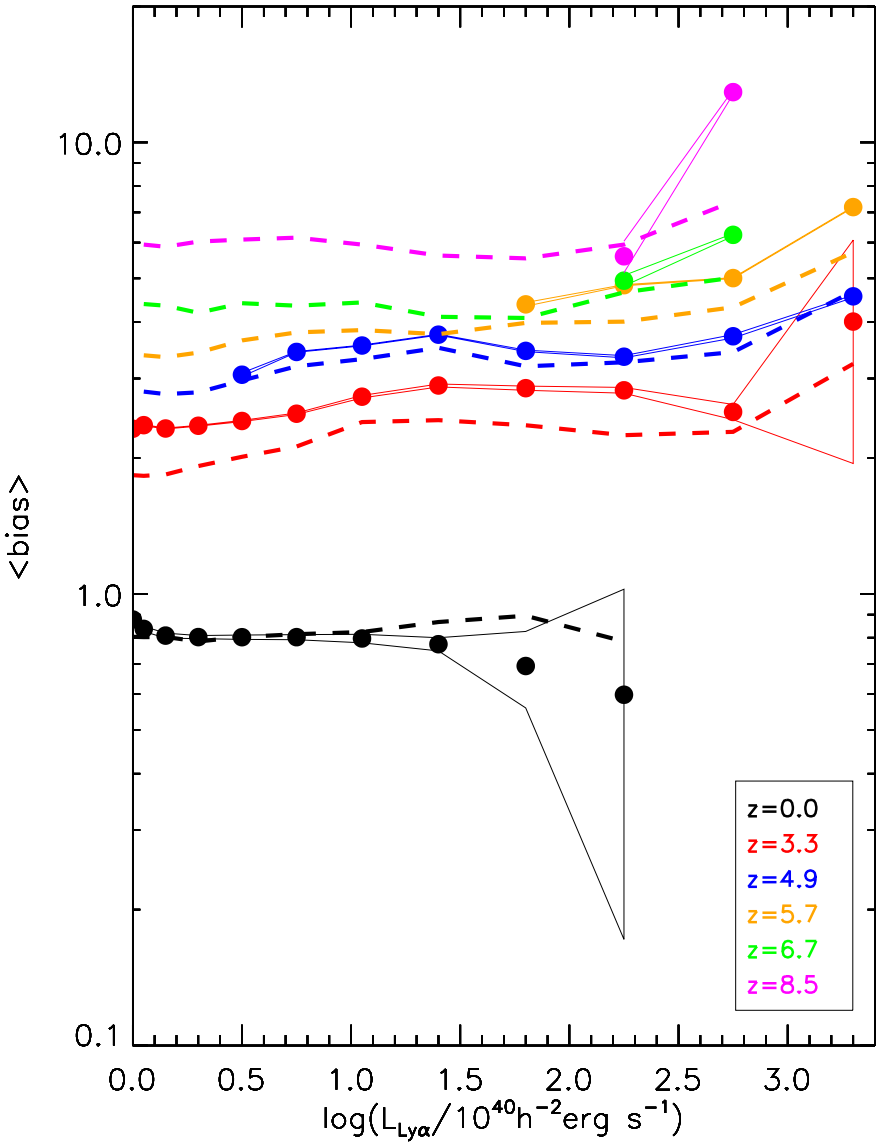

Figure 3.6: The galaxy bias as a function of Lyα luminosity at different redshifts, as indicated by the key. The solid lines show the results from the simulation and the dashed lines show the analytical expression of SMT. The area enclosed by the thin solid lines shows the error on the bias estimation for each redshift.

dures relate the halo bias to $\sigma(m,z)$, the *rms* linear mass fluctuation within a sphere which on average contains mass m. The bias factor for galaxies of a given luminosity is then obtained by averaging the halo bias over the halos hosting these galaxies. Le Delliou et al. (2006) used the analytical expression of Sheth et al. (2001) (hereafter SMT) to calculate the bias parameter for the semi-analytical galaxies. This gives a reasonable approximation to the large-scale halo bias measured in N-body simulations (e.g Angulo et al. (2008a)).

Fig. 3.6 shows the bias parameter as a function of luminosity for redshifts in the range $0 < z < 9$, and compares the direct calculation using the N-body simulation (solid lines) with the analytical estimation (dashed lines). In order to calculate the uncertainty in our value of the bias, we assume an error on $\xi_{\rm gal}(r)$ of the form $\Delta\xi_{\rm gal} = 2\sqrt{(1+\xi_{\rm gal})/DD}$ (Baugh et al., 1996), and assuming a negligible error in $\xi_{\rm dm}$ we get

$$\Delta b = \frac{1}{b\xi_{\rm dm}}\sqrt{\frac{1+b^2\xi_{\rm dm}}{DD}}, \tag{3.3}$$

for the error in the bias estimation. This error is shown in Fig. 3.6 as the range defined by the thin solid lines surrounding the bias measurement shown by the points.

The first noticeable feature of Fig. 3.6 is the strong evolution of bias with increasing redshift: From $z = 0$ to $z = 8.5$ the bias factor increases from $b(z = 0) \sim 0.8$ to $b(z = 8.5) \sim 12$, which means that the clustering amplitude of Lyα emitters at $z = 8.5$ is over 140 times the clustering amplitude of the dark matter at this redshift. Another interesting prediction is the dependence of bias on Lyα luminosity. For $z > 3$ there seems to be a strong increase of the bias with luminosity for bins where $L_{\rm Ly\alpha} > 10^{42}[\rm erg\ s^{-1}\ h^{-2}]$. The agreement between the analytic calculation of the bias and the simulation result is reasonable over the range $0 < z < 5$, but becomes less impressive as higher biases are reached. A similar discrepancy was also noticed by Gao et al. (2005), where they compared the halo bias extracted from the simulation with different analytic formulae (see also Angulo et al. (2008a)).

Another way to describe galaxy clustering is through the halo occupation distribution (HOD; Benson et al. (2000), Berlind et al. (2003), Cooray and Sheth (2002)). The HOD gives the mean number of galaxies which meet a particular observational selection as a function of halo mass. For flux-limited samples, the HOD can be broken down into the contribution from central galaxies and satellite galaxies. In a simple picture, the mean number of central galaxies is zero below some threshold halo mass, M_{min}, and unity for higher halo masses. With increasing halo mass, a second threshold is reached, M_{crit}, above which a halo can

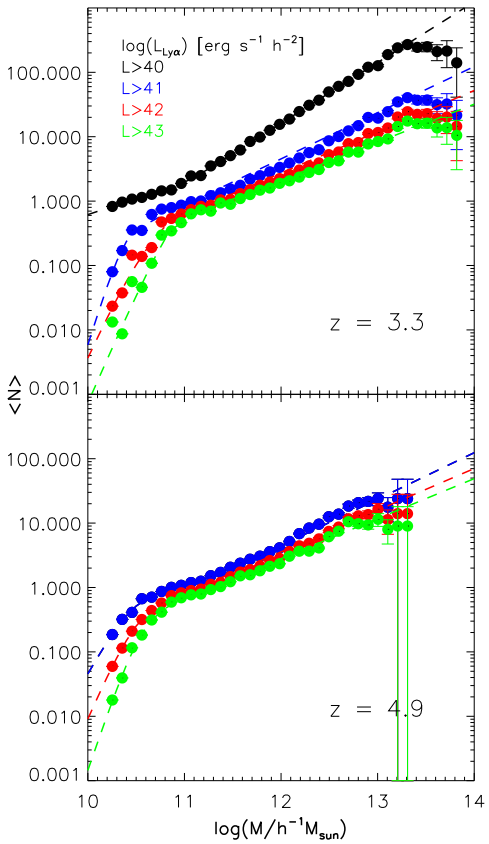

Figure 3.7: The HOD of Lyα emitters at $z = 3.3$ (*top*) and $z = 4.9$ (*bottom*). Each set of points represents a model sample with a different luminosity limit, as given by the key in the upper panel. The dashed line in each panel correspond to a "best" fit using the Berlind et al. (2003) parametrization.

also host a satellite galaxy. The number of satellites is usually described by a power-law of slope β. In the simplest case, three parameters are needed to describe the HOD (Berlind and Weinberg, 2002; Hamana et al., 2004); more detailed models have been proposed to describe the transition from 0 to 1 galaxy (Berlind et al., 2003).

We can compute the HOD directly from our model. The results are shown in Fig. 3.7, where we plot the HOD at two different redshifts for different luminosity limits. For comparison, we plot the HOD parametrization of Berlind et al. (2003) against our model predictions. In general, this HOD does a reasonable job of describing the model output, and is certainly preferred over a simple three parameter model. However, for the $z = 3.3$ case (top panel of Fig. 3.7), the shape of the model HOD for $\log(M/M_{\odot}) > 13$ is still more complicated than can be accommodated by the Berlind et al. parametrization, showing a flattening in the number of satellites as a function of increasing halo mass. There is less disagreement in the $z = 4.9$

case (bottom panel), but our model HOD becomes very noisy for large halo masses.

Table 3.1: Summary of survey properties and simulation results.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Survey	$z_{\rm survey}$	$z_{\rm simulation}$	Δz	Area [arcmin]^2	$EW_{\rm obs}$[Å]	$F_{{\rm Ly}\alpha}$[erg s^{-1} cm^{-2}]	$N_{\rm obs}$	$N_{\rm mock}^{\rm median}$	10-90%	C_ν
MUSYC	3.1	3.06	0.04	961	80	1.5×10^{-17}	162	142	89-207	0.41
SXDS	3.1	3.06	0.06	3538	328	1.1×10^{-17}	356	316	256-379	0.19
	3.7	3.58	0.06	3474	282	2.7×10^{-17}	101	80	60-110	0.31
	5.7	5.72	0.10	3722	335	7.4×10^{-18}	401	329	255-407	0.23
ELVIS	8.8	8.54	0.10	~ 3160	100	3.7×10^{-18}	–	20	14-29	0.37

Column (1) gives the name of the survey; (2) and (3) show the redshift of the observations and nearest output from the simulations, respectively; (4) shows the redshift width of the survey, based on the FWHM filter width; (5) shows the area covered by each survey; (6) and (7) show the equivalent width and Lyα flux limits of the samples, respectively; (8) shows the number of galaxies detected in each survey; (9) and (10) show the median of the number of galaxies and the 10-90 percentile range found in the mock catalogues for each survey. Finally, column (11) gives the fractional variation of the number of galaxies, defined in Eq. (3.13).

3.5 Mock Catalogues

In this section we make mock catalogues of Lyα emitters for a selection of surveys. In the previous section, we used the full simulation box to make clustering predictions, exploiting the periodic boundary conditions of the computational volume. The simulation is so large that it can accommodate many volumes equivalent to those sampled by current Lyα surveys, allowing us to examine the fluctuations in the number of emitters and their clustering. The characteristics of the surveys we replicate are listed in Table 3.1.

The procedure to build the mock catalogues is the following:

1. We extract a catalogue of galaxies from an output of the Millennium Simulation that matches (as closely as possible) the redshift of a given survey. The simulation output contains 64 snapshots spaced roughly logarithmically in the redshift range $[127, 0]$.

2. We choose one of the axes (say, the z-axis) as the line-of-sight, and we convert it to *redshift space*, to match what is observed in real surveys. To do this we replace r_z (the comoving space coordinate) with

$$s_z = r_z + \frac{v_z}{a\mathrm{H(z)}} \quad [h^{-1}\mathrm{Mpc}], \tag{3.4}$$

   where v_z is the peculiar velocity along the z-axis, $a = 1/(1+z)$ and H(z) is the Hubble parameter at redshift z.

3. We then apply the flux limit of the particular survey, to mimic the selection of galaxies. Table 3.1 shows the flux limits of the surveys considered.

4. Then we extract many mock catalogues using the same geometry as the real survey. We extract slices of a particular depth Δz (different for each survey), and within each slice we extract as many mock catalogues as possible using the same angular geometry as the real sample. Δz is determined using the transmission curves of the narrow-band filters used in each survey. To derive the angular sizes we use:

$$D_t(\theta, z) = d_c(z)\Delta\theta, \tag{3.5}$$

$$d_c(z) = \frac{c}{H_0}\int_0^z \frac{\mathrm{d}z'}{\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}}, \tag{3.6}$$

   where D_t is the transverse comoving size in h^{-1}Mpc, d_c is the comoving radial distance, c and H_0 are the speed of light and the Hubble constant respectively, Ω_m and Ω_Λ are

the density parameters of matter and the cosmological constant respectively. Eq. (3.5) is valid for $\Delta\theta \ll 1$[radians], which is the case for the surveys we analyse in this work. We assume a flat cosmology.

5. From the line-of-sight axis we invert Eq. (3.6) to obtain the redshift distribution of Lyα galaxies within each mock catalogue, converting the galaxy position (say, the z-coordinate) to redshift. This is then converted into an observed Lyα wavelength to take into account the shape of the filter transmission curve for each survey, which controls the minimum flux and equivalent width as a function of redshift. The value given in Table 3.1 corresponds to the minimum flux and $EW_{\rm obs}$ at the peak of the filter transmission curve. For redshifts at which the transmission is smaller (the tails of the curve) the minimum flux and $EW_{\rm obs}$ required for a Lyα emitter to be included are proportionally bigger.

6. Finally, we allow for incompleteness in the detection of Lyα emitters at a given flux due to noise in the observed images (where this information is available). To do this, we randomly select a fraction of galaxies in a given Lyα flux bin to match the completeness fraction reported for the survey at that flux.

Real surveys of Lyα emitters usually lack detailed information about the position of galaxies along the line-of-sight. Hence, instead of measuring the spatial correlation function defined in Eq. (3.1), it is only possible to estimate the angular correlation function, $w(\theta)$, which is the projection on the sky of $\xi(r)$.

We estimate $w(\theta)$ from mock catalogues using the following procedure, which closely matches that used in real surveys. To compute the angular correlation function we use the estimator (Landy and Szalay, 1993):

$$w_{LS}(\theta) = \frac{\langle DD(\theta)\rangle - 2\langle DR(\theta)\rangle + \langle RR(\theta)\rangle}{\langle RR(\theta)\rangle}, \tag{3.7}$$

where $\langle DR\rangle$ stands for data-random pairs, $\langle RR\rangle$ indicates the number of random-random pairs and all of the pair counts have been appropriately normalized. In the case of a finite volume survey, this estimator is more robust than the one defined in Eq. (3.1) because it is less sensitive to errors in the mean density of galaxies, such as could arise from boundary effects. In practice, the measured angular correlation function can be approximated by a

power law:

$$w(\theta) = A_w \left(\frac{\theta}{1^\circ} \right)^{-\delta}, \tag{3.8}$$

where A_w is the dimensionless amplitude of the correlation function, and δ is related to slope of the spatial correlation function, γ, from Eq. (3.2) by $\delta = \gamma - 1$. A relation between r_0 and A_w can be obtained using a generalization of Limber's equation (Simon, 2007).

Surveys of Lyα emitters typically cover relatively small areas of sky and can display significant clustering even on the scale of the survey. As a result, the mean galaxy number density within the survey area will typically differ from the cosmic mean value. If the number of galaxies within the survey is used to estimate the mean density, used in Eq. (3.7), rather than the unknown true underlying density, this leads to a bias in the estimated correlation function. This effect is known as the integral constraint (IC) bias. Landy and Szalay (1993) show that when their estimator is used, the expected value of the estimated correlation function $w_{LS}(\theta)$ is related to the true correlation function $w(\theta)$ by

$$\langle w_{LS}(\theta) \rangle = \frac{w(\theta) - w_\Omega}{1 + w_\Omega}, \tag{3.9}$$

where the integral constraint term w_Ω is defined as

$$w_\Omega \equiv \frac{1}{\Omega^2} \int d\Omega_1 d\Omega_2 w(\theta_{12}), \tag{3.10}$$

integrating over the survey area, and is equal to the fractional variance in number density over that area.

When the clustering is weak Eq. (3.9) simplifies to $\langle w_{LS}(\theta) \rangle \simeq w(\theta) - w_\Omega$. This motivates the additive IC correction which is customarily used in practice:

$$w_{corr}(\theta) = w_{LS}(\theta) + w_\Omega. \tag{3.11}$$

We use this to correct the angular correlation functions from our mock catalogues. In order to estimate the term w_Ω, we approximate the true correlation function as a power law, as in Eq. (3.8), and use

$$w_\Omega \simeq A_w \frac{\sum_i \langle RR_i \rangle \theta_i^{-\delta}}{\sum \langle RR_i \rangle}, \tag{3.12}$$

(Daddi et al., 2000), where $\langle RR \rangle$ are the same random pairs as used in the estimate of $w_{LS}(\theta)$.

To quantify the sample variance expected for a particular survey, we use the mock catalogues to calculate a *coefficient of variance* (C_v), which is a measure of the fractional variation

in the number of galaxies found in the mocks

$$C_v = \frac{N_{90} - N_{10}}{2N_{\rm med}}, \tag{3.13}$$

where N_{10} and N_{90} are the 10 and 90 percentiles of the distribution of the number of galaxies in the mocks, respectively, and $N_{\rm med}$ is the median. The value of C_v allows us to compare the sampling variance between different surveys in a quantitative way.

To analyse the clustering in the mock catalogues, we measured the angular correlation function of each mock catalogue using the procedure explained above. Then we fit Eq. (3.8) to each of the mock $w(\theta)$ and we choose the median value of A_w as the representative power law fit. We fix the slope of $w(\theta)$ to $\delta = 0.8$ for all surveys, except for ELVIS, where we found that a steeper slope, $\delta = 1.2$, agreed much better with the simulated data. To express the variation in the correlation function amplitude found in the mocks, we calculate the 10 and 90 percentiles of the distribution of A_w for each set of mock surveys. We also calculate $w(\theta)$ using the full transverse extent of the simulation, with the same selection of galaxies as for the real survey. This estimate of $w(\theta)$, which we call the *Model* $w(\theta)$, represents an ideal measurement of the correlation function without boundary effects (so there is no need for the integral constraint correction).

The surveys we mimic are the following: the MUSYC Survey (Gronwall et al., 2007; Gawiser et al., 2007), which is a large sample of Lyα emitting galaxies at $z = 3.1$; the SXDS Survey (Ouchi et al., 2005, 2008), which covers three redshifts: $z = 3.1$, $z = 3.7$ and $z = 5.7$, and finally, we make predictions for the forthcoming ELVIS survey (Nilsson et al., 2007b,a), which is designed to find Lyα emitting galaxies at $z = 8.8$. We now describe the properties of the mock catalogues for each of these surveys in turn.

3.5.1 The MUSYC Survey

The Multi-wavelength Survey by Yale-Chile (MUSYC) (Quadri et al., 2007; Gawiser et al., 2006, 2007; Gronwall et al., 2007) is composed of four fields covering a total solid angle of one square degree, each one imaged from the ground in the optical and near-infrared. Here we use data from a single MUSYC field consisting of narrow-band observations of Lyα emitters made with the CTIO 4-m telescope in the Extended Chandra Deep Field South (ECDFS) (Gronwall et al., 2007). The MUSYC field, centred on redshift $z = 3.1$, contains 162 Lyα emitters in a redshift range of $\Delta z \sim 0.04$ over a rectangular area of $31' \times 31'$ with flux and

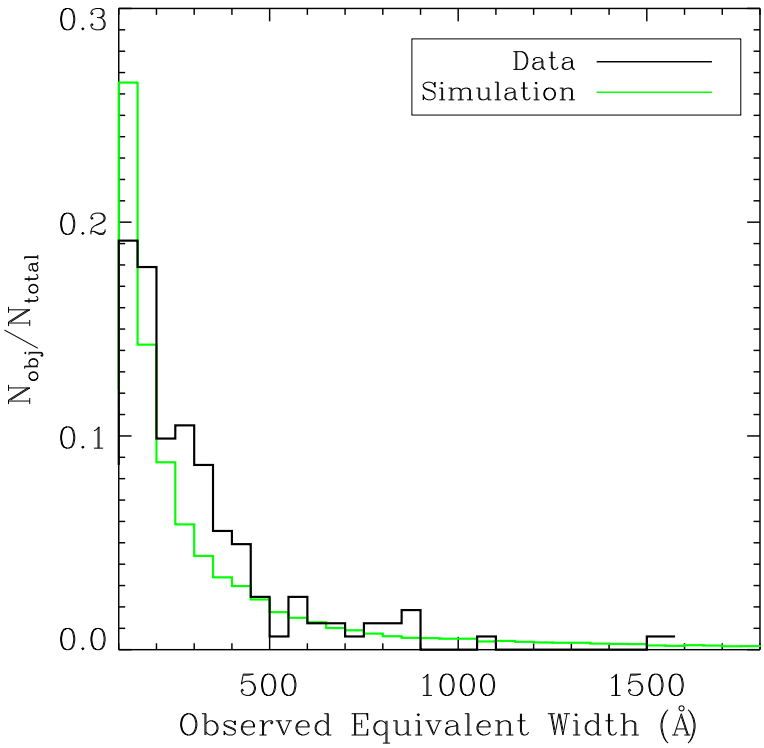

Figure 3.8: The observed EW_{obs} distribution of the MUSYC survey at $z = 3.1$ (solid black line) and the simulation (solid green line).

$EW_{\rm obs}$ limits described in Table 3.1.

To test how well the model reproduces the Lyα emitters seen in the MUSYC survey, we first compare the predicted (green) and measured (black) distributions of Lyα equivalent widths in Fig. 3.8. Here the predicted distribution comes from the full simulation volume. Overall, the simulation shows remarkably good agreement with the real data, with a slight underestimation in the range $200 < EW_{\rm obs}[\text{Å}] < 400$. For $EW_{\rm obs}[\text{Å}] > 400$ both distributions seem to agree well, although the number of detected Lyα emitters in the tail of the distribution is small.

For the MUSYC survey we built 252 mock catalogues from the Millennium simulation volume using the procedure outlined above. Fig. 3.9 shows an example of one of these mock catalogues. Many of the Lyα emitters are found in high dark matter density regions, and thus they are biased tracers of the dark matter. Fig. 3.10 shows the distribution of the number of galaxies in the ensemble of mocks. The green line shows the number detected in the real survey (162), which falls within the 10-90 percentile range of the mock distribution and is close to the median (142). The 10-90 percentile range spans an interval of $89 < N_{\rm gal} < 207$,

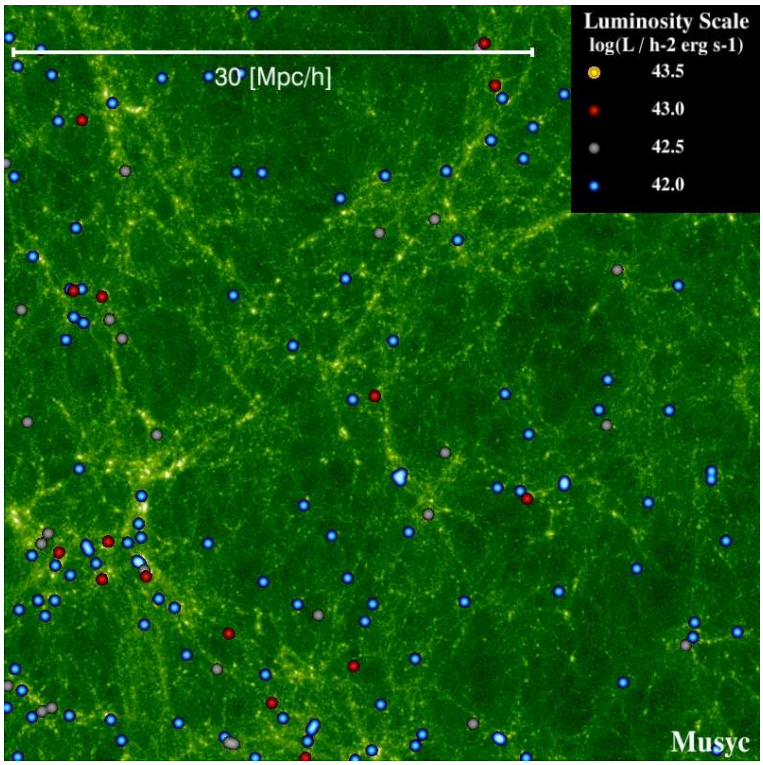

Figure 3.9: An image of a mock catalogue of the MUSYC Survey of Lyα emitters at $z = 3.1$. The colour format and legend are the same as used in Fig. 3.1. The angular size of the image is 31‘ × 31‘.

indicating a large cosmic variance for this survey configuration, with $C_v = 0.41$.

The next step is to compare the clustering in the simulations with the real data. Fig. 3.11 plots the correlation functions from the mock catalogues alongside that measured in the real survey (Gawiser et al., 2007). There is reasonable agreement between the mock catalogue results and the observed data. The median $w(\theta)$ from the mocks is slightly higher than the observed values, but the observed $w(\theta)$ is within the range containing 95% of the mock $w(\theta)$ values (i.e. between the 2.5% and 97.5% percentiles, shown by the light grey shaded region). We quantified this difference by fitting the power law of Eq. (3.8) to both real and mock data. The power-law fits were made over the angular range 1-10 arcmins. We find the value of A_w (Eq. 3.8) for each of the mock catalogue correlation functions by χ^2-fitting (using the same expression as in §3.4.2 for the error on each model datapoint) and then we plot the power law corresponding to the median value of A_w. We find $A_w = 0.53^{+1.01}_{-0.33}$ for the mocks, where the central value is the median, and the range between the error bars contains 95% of the values from the mocks. For the real data, we find the best-fit A_w and the 95% confidence

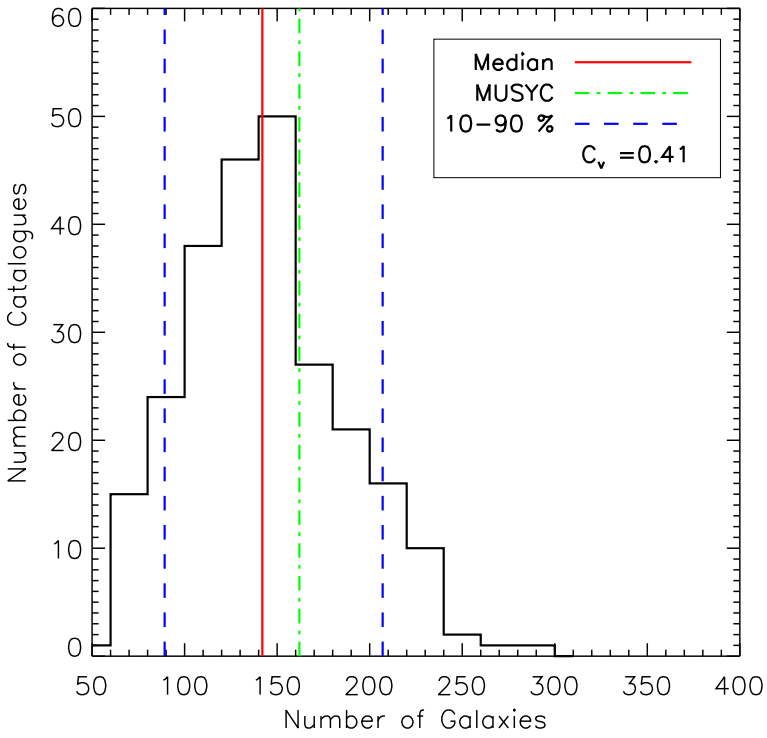

Figure 3.10: Histogram of the number of Lyα emitters found in the mock MUSYC catalogues. The red line shows the median, the dashed blue lines show the 10-90 percentile range, and the green line shows the number of galaxies detected in the real survey.

interval around it by χ^2-fitting, using the error bars on the individual datapoints reported by Gawiser et al.. This gives $A_w = 0.29 \pm 0.17$ for the real data. We again see that the observed value is within the 95% range of the mocks, and is thus statistically consistent with the model prediction. We also see that the 95% confidence error bar on the observed A_w is much smaller than the error bar we find from our mocks. This latter discrepancy arises from the small errors quoted on $w(\theta)$ by Gawiser et al. (2007), which are based on modified Poisson pair count errors, but neglect variations between different sample volumes (i.e. cosmic variance). On the contrary, using our mocks, we are able to take cosmic variance fully into account. This underlines the importance of including the cosmic variance in the error bars on observational data, to avoid rejecting models by mistake.

The red open circles in Fig. 3.11 show the correlation function obtained using the full angular size attainable with the Millennium simulation but keeping the same flux, EW and redshift limits as in the MUSYC survey (averaging 7 different slices), and so this measurement has a smaller sample variance. The area used here is $\sim$ 120 times bigger than the MUSYC

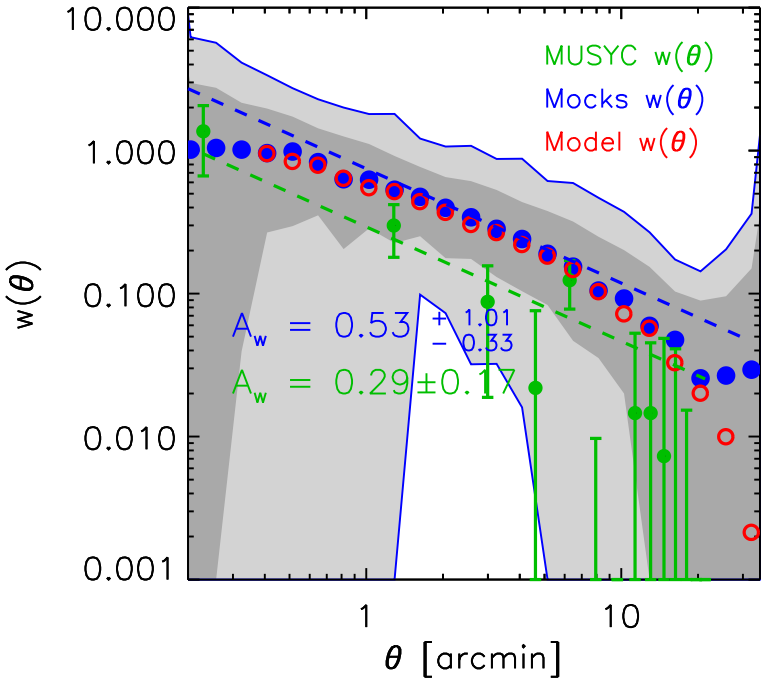

Figure 3.11: Angular clustering for the MUSYC Survey. Green circles show $w(\theta)$ calculated from the observed catalogue (Gawiser et al., 2007). The blue circles show the median $w(\theta)$ from all mock catalogues, corrected for the integral constraint effect. The dark and light grey shaded regions respectively show the 68% and 95% ranges of the distribution of $w(\theta)$ measured in the mock catalogues. The red open circles show the *Model* correlation function, obtained using the width of the entire simulation box (and the same EW, flux and redshift limits). The dashed lines show the power-law fit to the observed $w(\theta)$ (green) and the median fit to $w(\theta)$ from the mock catalogues (blue). The amplitudes A_w of these fits are also given in the figure.

area, so IC effects are negligible on the scales studied here. We refer to this as the *Model* prediction for $w(\theta)$.

The median of the mock correlation functions (including the IC correction, blue circles) is seen to agree reasonably well with the *Model* correlation function (red open circles) for $\theta < 20$[arcmin]. This shows that for this survey it is possible to obtain an observational estimate of the correlation function which is unbiased over a range of scales, by applying the integral constraint correction. However, on large scales the median $w(\theta)$ of the mocks (with IC correction included) lies above the Model $w(\theta)$, which shows that the IC correction is not

perfect, even on average. Presumably this failure is due (at least in part) to the fact that the IC correction procedure assumes that $w(\theta)$is a power law, while the true $w(\theta)$ departs from a power law on large scales. It is also important to note that these statements only apply to the median $w(\theta)$ derived from the mock samples - the individual mocks show a large scatter around the true $w(\theta)$ (as shown by the grey shading), and the IC correction does not remove this. This scatter rapidly increases at both small and large angular scales, so the best constraints on $w(\theta)$ from this survey are for intermediate scales, $1 \lesssim \theta \lesssim 5$[arcmin].

3.5.2 The SXDS Surveys

The Subaru/*XMM-Newton* Deep Survey (SXDS) (Ouchi et al., 2005, 2008; Kashikawa et al., 2006) is a multi-wavelength survey covering ~ 1.3 square degrees of the sky. The survey is a combination of deep, wide area imaging in the X-ray with *XMM-Newton* and in the optical with the Subaru Suprime-Cam. Here we are interested in the narrow-band observations at three different redshifts: $3.1, 3.6$ and 5.7 (Ouchi et al., 2008).

We build mock SXDS catalogues following the same procedure as outlined above. Fig. 3.12 shows examples of our mock catalogues for each redshift. As in the previous case, we see that Lyα emitters on average trace the higher density regions of the dark matter distribution. The real surveys have a well defined angular size. However, the area sampled is slightly different at each redshift. In order to keep the cross-like shape in our mock catalogues and be consistent with the exact area surveyed, we scaled the cross-like shape to cover the same angular area as the real survey at each redshift.

Fig. 3.13 shows the distribution of the number of galaxies in the mock catalogues for the three redshifts surveyed. The median number of galaxies in the mocks at $z = 3.1$ is 316, which is remarkably similar to the observed number, 356. The 10-90 percentile range of the mocks covers 256–379 galaxies. The coefficient of variation is $C_\nu = 0.19$, less than half the value found for the MUSYC mock catalogues, $C_\nu = 0.41$. This reduction is due mainly to the larger area sampled by the SXDS survey. In the second slice ($z = 3.6$), the redshift is only slightly higher than in the previous case, but the number of galaxies is much lower. Looking at the top panel of Fig. 3.2 we see that the observed LFs are basically the same for these two redshifts. The difference between the two samples is explained mostly by the different Lyα flux limits (1.2×10^{-17}[erg s^{-1} cm^{-2}] for $z = 3.1$ and 2.6×10^{-17}[erg s^{-1} cm^{-2}] for $z = 3.6$). For the $z = 3.6$ mocks, we find a median number of 80 and 10-90% range 60–110,

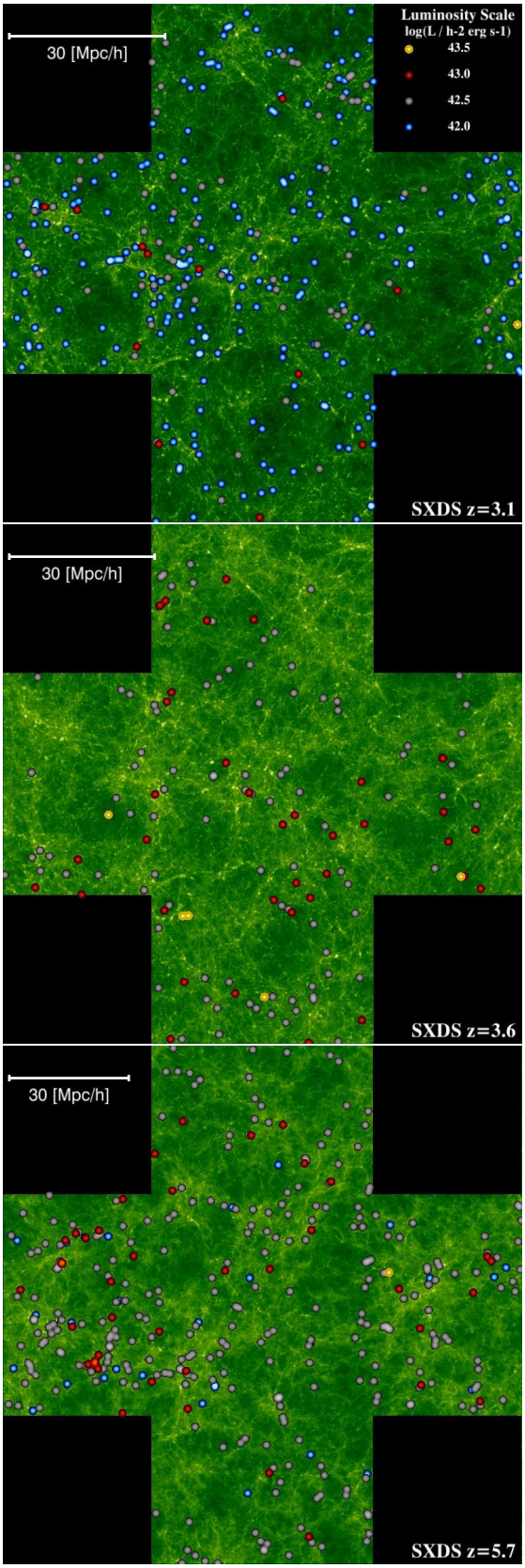

Figure 3.12: Mock catalogues of the SXDS survey for redshifts 3.1 (*top*), 3.6 (*centre*) and 5.7 (*bottom*). The colour scheme and legend are the same as used previously. The angular size of the image is $1.4° \times 1.4°$.

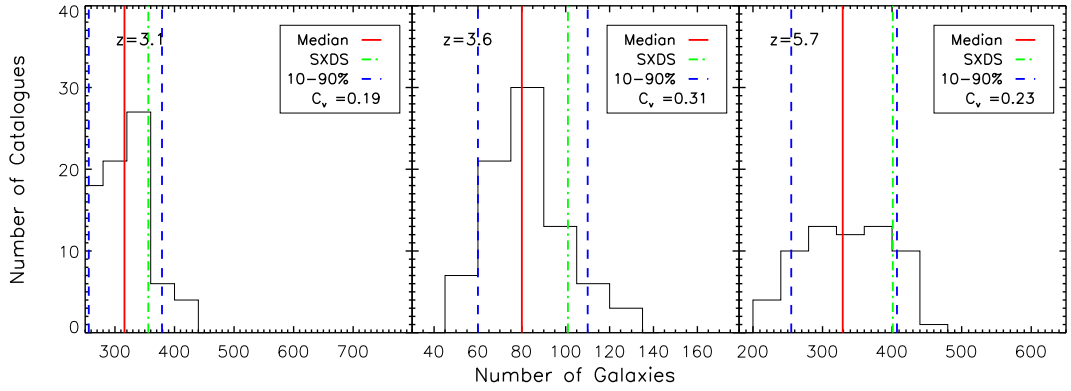

Figure 3.13: The distribution of the number of galaxies in mock SXDS catalogues, for $z = 3.1$ *(left)*, $z = 3.6$ *(centre)* and $z = 5.7$ *(right)*. The red line shows the median of the number of galaxies inside the mock catalogues, the blue lines show the 10-90 percentiles of the distribution, and the green line shows the number observed in the SXDS.

in reasonable agreement with the observed number of galaxies, 101. The fractional variation in the number of galaxies, quantified by $C_v = 0.31$, is larger than in the previous case, due to the smaller number of galaxies. The $z = 5.7$ case is similar to the lower redshifts. The median number in the mocks is 329, with 10-90% range 255–407, again consistent with the observed number, 401. The coefficient of variation for this survey is $C_v = 0.23$, so the sampling variance is intermediate between that for the $z = 3.1$ and $z = 3.6$ surveys.

The angular correlation functions of the mock catalogues are compared with the real data in Fig. 3.14. The observational data shown are preliminary angular correlation function measurements in the three SXDS fields, with errorbars based on bootstrap resampling (M. Ouchi, private communication). As in our comparison with the MUSYC survey, we plot the median correlation function measured from the mocks, after applying the IC correction, as a representative $w(\theta)$. As before, we also perform a χ^2 fit of a power law to the $w(\theta)$ measured in each mock, and to the observed values, to determine the amplitude A_w. The fit is performed over the range $1 < \theta < 10$[arcmin] as before.

The left panel of Fig. 3.14 shows the correlation functions at $z = 3.1$. According to both the error bars on the observational data, and the scatter in $w(\theta)$ in the mocks (shown by the grey shading), this survey provides useful constraints on the clustering for $1 \lesssim \theta \lesssim 10$[arcmin], but not for smaller or larger angles, where the scatter becomes very large. The fitted amplitude A_w for the observed correlation function is $A_w = (0.32 \pm 0.22)$ (95% con-

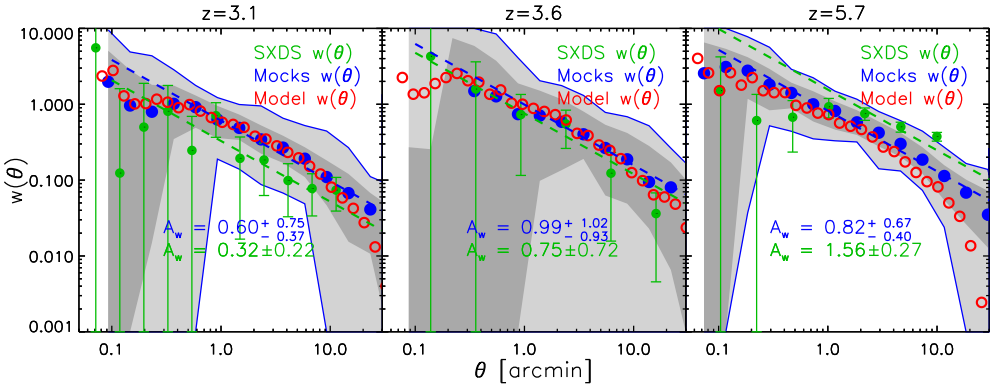

Figure 3.14: Angular correlation functions for the mock SXDS catalogues at $z = 3.1$ *(Left)*, $z = 3.6$ *(Center)* and $z = 5.7$ *(Right)*. The blue circles show the median $w(\theta)$ from the mock catalogues (after applying the IC correction). The dark and light grey shaded regions respectively show the 68% and 95% ranges of the distribution of $w(\theta)$ measured in the mock catalogues. The red open circles are the Model $w(\theta)$ calculated using the full simulation width, averaged over many slices. The green circles shows the observational data from Ouch et al. The dashed lines show the power-law fit to the observed $w(\theta)$ (green) and the median fit to $w(\theta)$ from the mock catalogues (blue). The amplitudes A_w of these fits are also given in the figure.

fidence, using the error bars reported by Ouchi et al.), somewhat below the median value found in the mocks, $A_w = 0.60$, but within the 95% range for the mocks ($A_w = 0.23 - 1.35$). Based on the mocks, the model correlation function is consistent with the SXDS data at this redshift.

Comparing these results with those we found for the MUSYC survey (which has a very similar redshift and flux limit to SXDS at $z = 3.1$), we see that the results seem very consistent. The MUSYC survey has a larger sample variance than SXDS, but the measured clustering amplitude is very similar in the two surveys.

The middle panel of Fig. 3.14 shows the correlation function for the $z = 3.6$ survey. In this case, the error bars on the observational data and the scatter in the mocks are both larger, due to the lower surface density of galaxies in this sample. For the observed correlation amplitude, we obtain $A_w = 0.75 \pm 0.72$, while for the mocks we find a median $A_w = 0.99$, with 95% range 0.06–2.01, entirely consistent with the observational data.

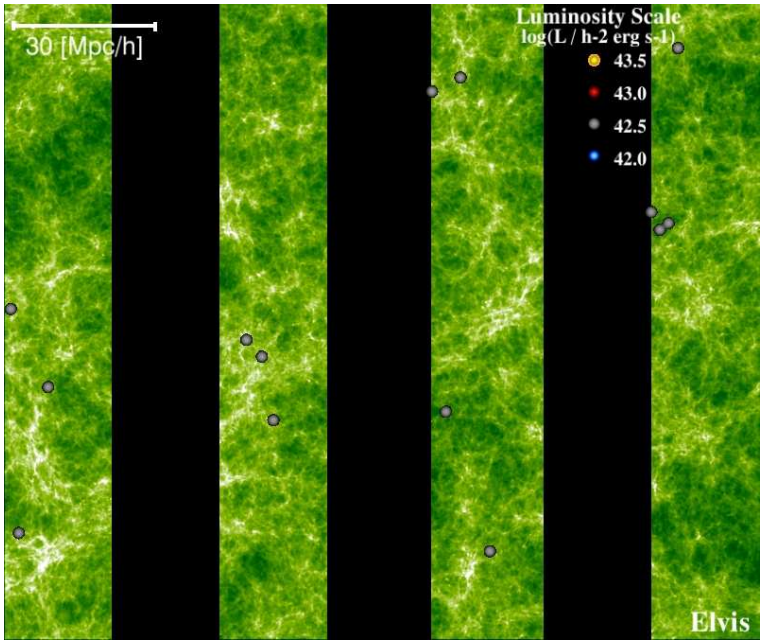

Figure 3.15: An example of a mock catalogue for the ELVIS Survey. The image shows the observed field of view (four strips). The legend and colour format are the same as in Figs.3.9 and 3.12.

The right panel of Fig. 3.14 shows the correlation function predictions for $z = 5.7$. According to the spread of mock catalogue results, the $w(\theta)$ measured here is the most accurate of the three surveys, due to the large number of galaxies. For the mocks, we find a median correlation amplitude $A_w = 0.82$, with 95% range 0.42–1.49. For the observations, we find $A_w = 1.56 \pm 0.27$, if we assume a slope $\delta = 0.8$. The average correlation function in the mocks agrees well with this slope over the range fitted, but the observational data for $w(\theta)$ at this redshift prefer a flatter slope. The model is however still marginally consistent with the observational data at 95% confidence. Similarly flat shapes were also found in some previous surveys (Shimasaku et al., 2004; Hayashino et al., 2004) in the same field, but at redshifts 3.1 and 4.9 respectively. However, these surveys are much smaller in terms of area surveyed and number of galaxies (this is particularly so in Shimasaku et al. (2004)). This behaviour in $w(\theta)$ might be produced by the high density regions associated with protoclusters in the SXDS fields (M. Ouchi, private communication), but still this behaviour of $w(\theta)$ must be confirmed to prove that it is a real feature of the correlation function.

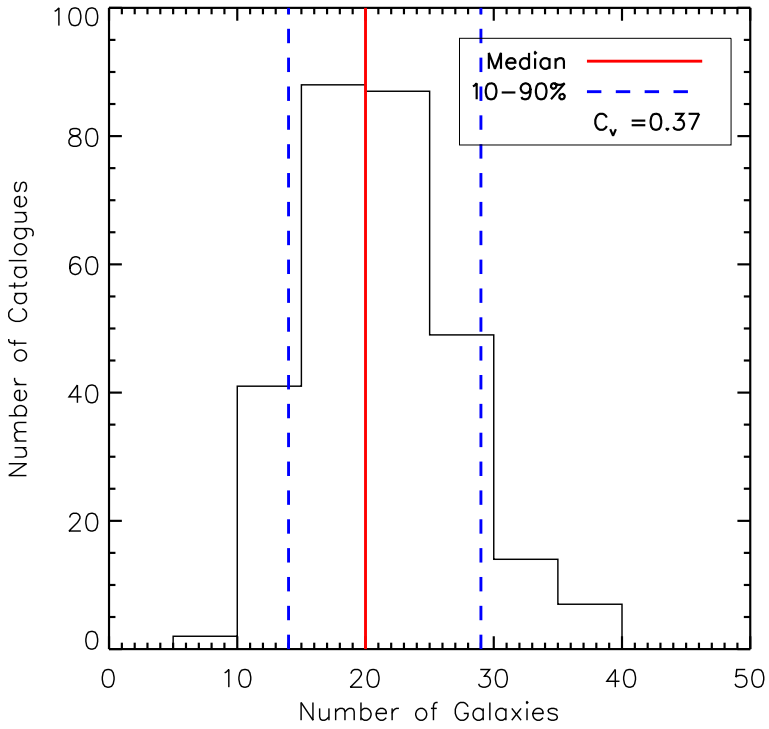

Figure 3.16: Histogram of the number of galaxies in mock catalogues expected for the ELVIS Survey. The red line shows the median of the distribution, and the blue dashed lines the 10-90 percentiles of the distribution.

3.5.3 ELVIS Survey

One of the main goals of the public surveys planned for the Visible and Infrared Survey Telescope for Astronomy (VISTA) is to find a significant sample of very high redshift ($z \sim 8.8$) Lyα emitters. This program is called the Emission-Line galaxies with VISTA survey (ELVIS) (Nilsson et al., 2007b,a). The plan for ELVIS is to image four strips of $11.6' \times 68.27'$, covering a total area of $0.878\mathrm{deg}^2$, as shown in Fig. 3.15. This configuration is dictated by the layout of the VISTA IR camera array. The only current detections of Lyα emitters at $z > 8$ are those of Stark et al. (2007), which have not yet been independently confirmed. Lyα emitting galaxies at such redshifts will provide us with valuable insights into the reionization epoch of the Universe, as well as galaxy formation and evolution.

For our mock ELVIS catalogues, we select galaxies with a minimum flux of $F_{\mathrm{Ly}\alpha} = 3.7 \times 10^{-18}[\mathrm{erg\ s^{-1}\ cm^{-2}}]$ and $EW_{\mathrm{obs}} > 100$ Å, as listed in Table 3.1. (The EW_{obs} limit is just a rough estimate, although our predictions should not be sensitive to the exact value chosen.) Fig. 3.15 shows the expected spatial distribution of $z = 8.5$ galaxies in one of the ELVIS mock

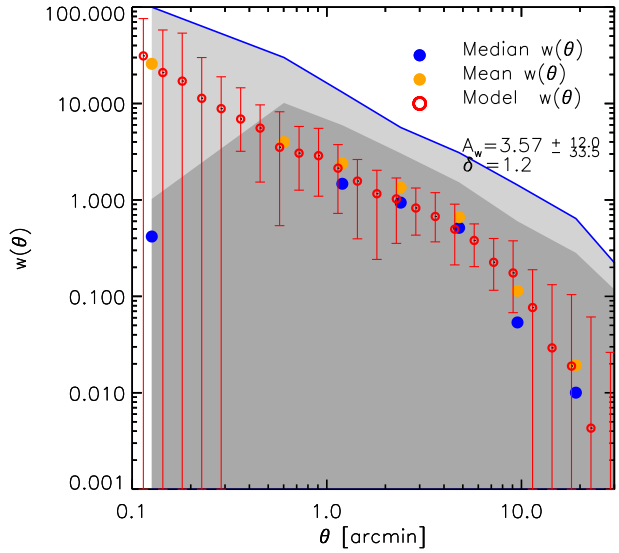

Figure 3.17: Angular Correlation Functions in the mock catalogues of the ELVIS Survey. The blue circles shows the median $w(\theta)$ from the mock catalogues, while the orange circles shows the mean. The dark and light grey shaded regions respectively show the 68% and 95% ranges of the distribution of $w(\theta)$ in the mocks. The red open circles show the Model $w(\theta)$ obtained using the full width of the simulation box. The amplitude and slope of the median power-law fit to the mocks are also given.

catalogues. Each mock catalogue has four strips, matching the configuration planned for the real survey. The median number of galaxies within the mock catalogues is 20, with a 10-90 percentile spread of 14 to 29 galaxies, as shown in Fig. 3.16. The fractional variation in number between different mocks is $C_\nu = 0.37$, which is quite large, but no worse than for the MUSYC survey at $z = 3.1$, even though that survey has 10 times as many galaxies.

The angular correlation functions of the mock ELVIS catalogues were calculated in the same way as before (including the integral constraint correction). Fig. 3.17 shows the median of the $w(\theta)$ values measured from each mock catalogue (blue circles), and also the mean (orange circles). In this case, the distribution of $w(\theta)$ values within each angle bin is very skewed, due to the small number of galaxies in the mock catalogues, and so the mean and median can differ significantly. The dark and light grey shaded regions show the ranges containing 68% and 95% of the $w(\theta)$ values from the mocks, from which it can be seen that the cosmic variance for this survey is very large. We also show the Model $w(\theta)$ (red circles), which provides our best estimate of the true correlation function based on the Millennium

simulation, and was calculated by averaging over 10 slices of the simulation, using the full width of the simulation box. Even here, the error bars on $w(\theta)$ are quite large, due to the very low number density of galaxies predicted. We see that the mean and median $w(\theta)$ in the mocks lie close to the Model values for $2[\mathrm{arcmin}] < \theta < 20[\mathrm{arcmin}]$, so in this sense they provide an unbiased estimate.

The most noticeable feature of Fig. 3.17 is the large area covered by both the 68% and 95% ranges of the distribution of $w(\theta)$ in the mocks, which extend down to $w(\theta) = 0$. This indicates that the ELVIS survey will only be able to put a weak upper limit on the clustering amplitude of $z \sim 8.8$ Lyα emitters, if our model is correct. As before, we can quantify this by fitting a power law to $w(\theta)$ in our mocks. We notice that the Model $w(\theta)$ for this sample has a significantly steeper slope, $\delta = 1.2$, than the canonical value $\delta = 0.8$, and so we do our fits to the mocks using $\delta = 1.2$. We find a median amplitude in the mocks $A_w = 3.57^{+12.0}_{-33.5}$, where the error bars give the 95% range.

3.6 Summary and conclusions

We have combined a semi-analytical model of galaxy formation with a high resolution, large volume N-body simulation to make predictions for the spatial distribution of Lyα emitters in a ΛCDM universe.

Our model for Lyα emitters is appealingly simple. Using the star formation history predicted for each galaxy from the semi-analytical model to compute the production of Lyman continuum photons, we find that on adopting a constant escape fraction of Lyα photons the observed number of Lyα emitters can be reproduced amazingly well over a range of redshifts (Le Delliou et al., 2006). Our modelling of Lyα emission may appear overly simplistic on first comparison to other calculations in the literature. For example, Nagamine et al. (2008) predicted the clustering of Lyα emitters in a gas-dynamic simulation, modelling the Lyα emission through a Lyα escape fraction or a duty cycle scenario. However, the fraction of active emitters in the duty cycle scenario needs to be tuned at each redshift, for which there is no physical justification. Since our predictions for Lyα emission are derived from a full model of galaxy formation, it is straightforward to extract other properties of the emitters, such as their stellar mass or the mass of their host dark matter halo (Le Delliou et al., 2006). In this work we have extended this work to include explicit predictions for the spatial and angular

clustering of Lyα emitters.

We have studied how the clustering strength of Lyα emitters depends upon their luminosity as a function of redshift. Generally, we find that Lyα emitters show a weak dependence of clustering strength on luminosity, until the brightest luminosities we consider are reached. At the present day, Lyα emitters display weaker clustering than the dark matter. This changes dramatically at higher redshifts ($z > 3$), with currently observable Lyα emitters predicted to be much more strongly clustered than the dark matter, with the size of the bias increasing with redshift. We compared the simulation results with analytical estimates of the bias. Whilst the analytical results show the same trends as the simulation results, they do not match well in detail, and this supports the use of an N-body simulation to study the clustering.

A key advantage of using semi-analytical modelling is that the evolution of the galaxy population can be readily traced to the present day. This gives us some confidence in the star formation histories predicted by the model. The semi-analytical model passes tests on the predicted distribution of star formation rates at high redshift (the number counts and redshifts of galaxies detected by their sub-millimetre emission and the luminosity function of Lyman-break galaxies), whilst also giving a reasonable match to the present day galaxy luminosity function (Baugh et al. 2005), and also matching the observed evolution of galaxies in the infrared (Lacey et al., 2008). Gas dynamic simulations as a whole struggle to reproduce the present-day galaxy population, due to a combination of a limited simulation volume (set by the need to attain a particular mass resolution) and a tendency to overproduce massive galaxies. The small box size typically employed in gas dynamic simulations means that fluctuations on the scale of the box become nonlinear at low redshifts, and their evolution can no longer be accurately modelled. A further consequence of the small box size is that predictions for clustering are limited to small pair separations (e.g. Nagamine et al. (2008)) use a box of side 33 h^{-1}Mpc, limiting their clustering predictions to scales $r \lesssim 3h^{-1}$Mpc). By using a simulation with a much larger volume than that of any existing Lyα survey, we can subdivide the simulation box to make many mock catalogues to assess the impact of sampling fluctuations (including cosmic variance) on current and future measurements of the clustering of Lyα emitters.

We made mock catalogues of Lyα emitters to compare with the MUSYC ($z = 3$) and SXDS ($z = 3 - 6$) surveys, and to make predictions for the forthcoming ELVIS survey at $z \sim 9$. In the case of MUSYC and SXDS, we found that the observed number of galaxies

lies within the 10-90 percentile interval of the number of Lyα emitters found in the mocks. We find that high-redshift clustering surveys underestimate their uncertainties significantly if they fail to account for cosmic variance in their error budget. Overall, the measured angular correlation functions are consistent with the model predictions. The clustering results in our mock catalogues span a wide range of amplitudes due to the small volumes sampled by the surveys, which results in a large cosmic variance. ELVIS will survey Lyα emitters at very high redshift ($z = 8.8$). Our predictions show that a single pointing will be strongly affected by sample variance, due to the small volume surveyed and the strong intrinsic clustering of the Lyα emitters which will be detected at this redshift. Many ELVIS pointings will be required to get a robust clustering measurement.

We have shown that surveys of Lyα emitters can open up a new window on the high redshift universe, tracing sites of active star formation. With increasing redshift, the environments where Lyα emitters are found in current and planned surveys become increasingly unusual, sampling the galaxy formation process in regions that are likely to be proto-clusters and the progenitors of the largest dark matter structures today. Our calculations show that with such strong clustering, surveys of Lyα emitters covering much larger cosmological volumes are needed in order to minimize cosmic variance effects.

Chapter 4

Probing dark energy with Hα emitters

4.1 Introduction

A number of approaches have been proposed to uncover the nature of the accelerating expansion of the Universe which involve measuring the large scale distribution of galaxies (e.g Albrecht et al., 2006; Peacock et al., 2006). The ability of galaxy surveys to discriminate between competing models depends on their volume. Once the solid angle of a survey has been set, the useful volume can be maximised by choosing a tracer of the large-scale structure of the Universe which can effectively probe the geometrical volume. This depends on how the abundance of tracers drops with increasing redshift, and how much of this decline is offset by an increase in the clustering amplitude of the objects.

Several wide-angle surveys have probed the redshift interval between $0 < z < 1$ (e.g Colless et al., 2003; York et al., 2000; Cannon et al., 2006; Blake et al., 2009). The next major step up in volume will be made when the range from $0.5 < z < 2$ is opened up with large near-infrared cameras and spectrographs which are mounted on telescopes able to map solid angles running into thousands of square degrees. From the ground, this part of the electromagnetic spectrum is heavily absorbed by water vapour in the Earth's atmosphere and affected by the strong atmospheric OH emission lines. A space mission to construct an all-sky map of galaxies in the redshift range $0.5 < z < 2$ would have a significant advantage over a ground based survey in that the sky background in the near-infrared (NIR) is around 500 times weaker in space than it is on the ground.

An important issue yet to be resolved for a galaxy survey extending to $z \sim 2$ is the construction of the sample and the method by which the redshifts will be measured. One option is to use slitless spectroscopy and target the Hα emission line. Hα is located at a restframe

wavelength of $\lambda = 6563$Å, which, for galaxies at $z > 0.5$, falls into the near-infrared part of the electromagnetic spectrum (Thompson et al., 1996; McCarthy et al., 1999; Hopkins et al., 2000; Shim et al., 2009). Hα emission is powered by UV ionizing photons from massive young stars. The only source of attenuation is dust, which is less important at the wavelength of Hα than it is for shorter wavelength lines. This makes Hα a more direct tracer of galaxies which are actively forming stars than other lines such as Lyα, OII, OIII, Hβ or Hγ, which suffer from one or more sources of attenuation (i.e. dust, stellar absorption, resonant scattering) and which are more sensitive to the metallicity and ionisation state of the gas. The second option is to use some form of multi-slit spectrograph to carry out a redshift survey of a magnitude limited sample. The use of a slit means that unwanted background is reduced, allowing fainter galaxies to be targetted. Also, it is easier to identify which spectrum belongs to which galaxy with a slit than it is with slitless spectroscopy. Targets could be selected in the H-band at an effective wavelength of just over 1 micron, which is around the centre of the near infrared wavelength part of the spectrum. The slitless option has the advantage of not needing an initial target selection and relies on a technique that is has already been used in space and is potentially cheaper than the multi-slit solution.

Space missions designed to carry out redshift surveys like the ones outlined above are currently being planned and assessed on both sides of the Atlantic. At the time of writing, the European Space Agency is conducting a Phase A study of a mission proposal called Euclid [1], one component of which is a galaxy redshift survey. Both of the selection techniques mentioned above are being evaluated as possible spectroscopic solutions. The slit solution for Euclid is based on a novel application of digital micromirror devices (DMDs) to both image the galaxies to build a parent catalogue in the H-band and to measure their redshifts (see Cimatti et al. 2009 for further details about the Euclid redshift survey). A Hα mission is also being discussed in the USA [2]. At this stage, the sensitivity of these missions is uncertain and subject to change. For this reason we consider a range of Hα flux limits and H-band magnitudes when assessing the performance of the surveys. The specifications and performance currently being discussed for these missions have motivated the range of fluxes that we consider.

A simple first impression of the relative merits of different selections methods can be

[1] http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=43226

[2] http://jdem.gsfc.nasa.gov/

gained by calculating the effective volume of the resulting survey. This requires knowledge of the survey geometry and redshift coverage, along with the redshift evolution of the number density of sources and their clustering strength. In this chapter we use published galaxy formation models to predict the abundance and clustering of different samples of galaxies in order to compute the effective volumes of a range of Hα and H-band surveys. Observationally, relatively little is known about the galaxy population selected by Hα emission or H-band magnitude at $0.5 < z < 2$. Empirically it is possible to estimate the number density of sources from the available luminosity function data and, on adopting a suitable model, to use the limited clustering measurements currently available to infer the evolution of the number density and bias (Shioya et al., 2008; Morioka et al., 2008; Geach et al., 2008). Geach et al. (2009), in a complementary study to this one, make an empirical estimate of the number density of Hα emitters, and combine this with the predictions of the clustering of these galaxies presented in this work to estimate the efficiency with which Hα emitters can measure the large scale structure of the Universe. We remind the reader that the effective volume is just one aspect that needs to be taken into account when choosing between different spectroscopic solutions and we do not address here issues of cost or survey feasibility.

The outline of this chapter is as follows: in Section 4.2 we give a brief overview of the models. Some general properties of Hα emitters in the models, such as luminosity functions (LF), equivalent widths (EW) and clustering bias are presented in Section 4.3. In Section 4.4 we show how our models can be used to build mock survey catalogues. We analyse the differences in the clustering of Hα emitters and H-band selected galaxies and present an indication of the efficiency with which different surveys trace large-scale structure (LSS). Finally, we give our conclusions in Section 4.5.

4.2 The Models

In this work we present predictions for the clustering of galaxy samples selected in the near-infrared using two published versions of the semi-analytic model `GALFORM`. An overview of the semi-analytical approach to modelling galaxy formation has already been given in Chapter 2.

The two models considered in this work are explained fully in the original papers, Baugh et al. (2005) (hereafter the `Bau05` model) and Bower et al. (2006) (hereafter the `Bow06` model).

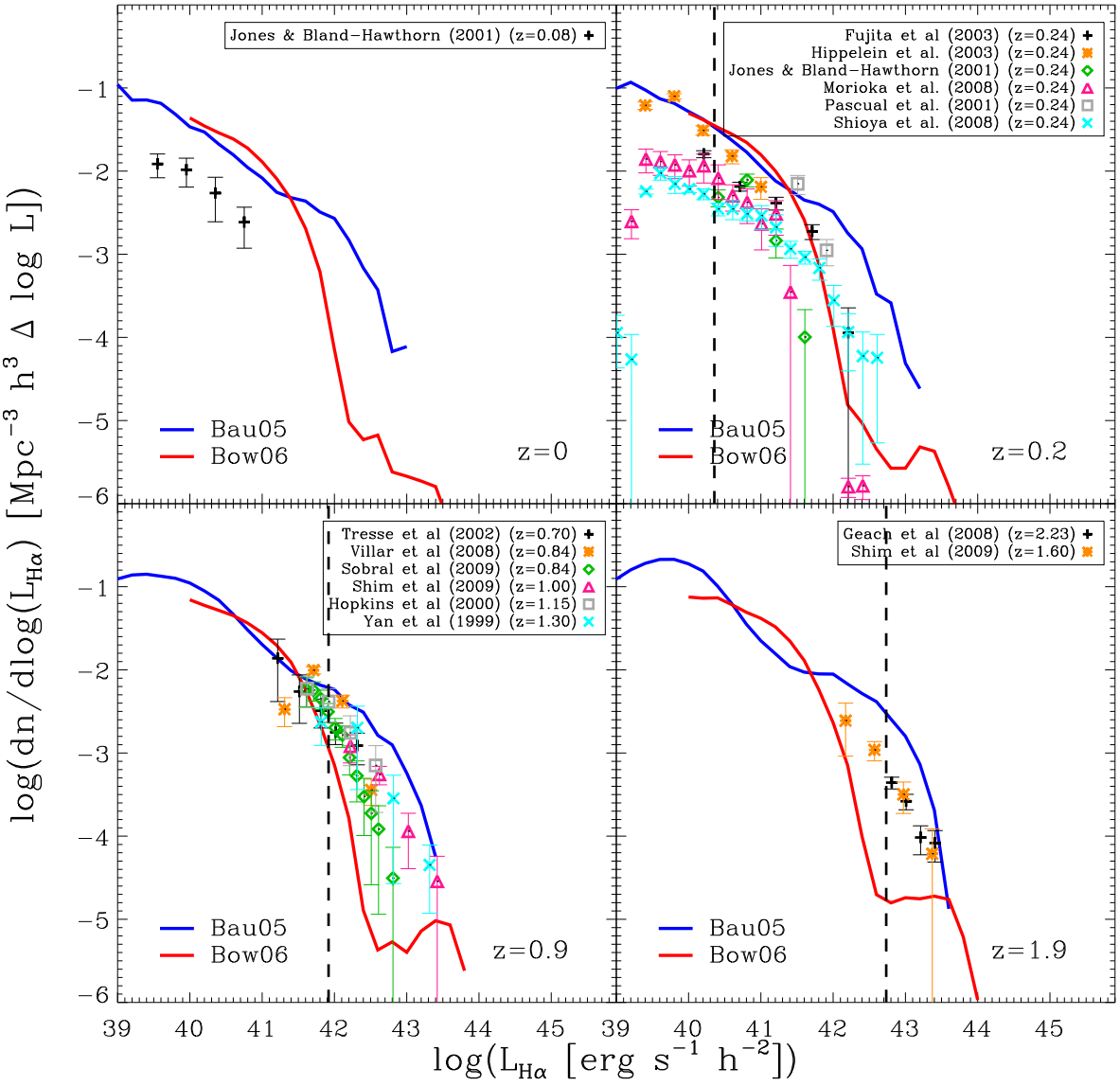

Figure 4.1: The Hα luminosity function, including attenuation by dust, at different redshifts. The blue curves show the predictions of the `Bau05` model, whereas red curves show the `Bow06` model. The observational estimates are represented by the symbols (see text for details). The redshift displayed in the bottom-right corner of each panel gives the redshift at which the `GALFORM` models were run. The vertical black dashed line shows the Hα luminosity corresponding to the flux $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -15.4$ for $z > 0$, displayed to show the expected luminosity limit of current planned space missions.

The calculation of H-band flux and Hα line emission is the same in both models. The model predicts the star formation history of each galaxy, recording the star formation rate and the metallicity with which stars are made in each of the galaxy's progenitors. This allows a composite stellar population and spectral energy distribution to be built up. The model predicts the scale size of the galaxy and, through a chemical evolution model, the metal content of the disk and bulge. The H-band magnitude is computed by convolving the model galaxy spectral energy distribution with an H-band filter, appropriately shifted in wavelength if the galaxy is observed at $z > 0$. The effect of dust extinction is taken into account by assuming that the dust and disk stars are mixed together (Cole et al. 2000). The spectral energy distribution also gives the rate of production of Lyman continuum photons. Then, all of the ionizing photons are assumed to be absorbed by the neutral gas in the galaxy, and, by adopting case B recombination (Osterbrock 1989), the emissivity of the Hα line (and other emission lines) is computed. Here we assume that the attenuation of the Hα emission is the same as that experienced by the continuum at the wavelength of Hα . To predict the equivalent width (EW) of the Hα emission, we simply divide the luminosity of the line by the luminosity of the continuum around the Hα line.

4.3 Properties of Hα emitters

We first concentrate on the nature of Hα emitters in the models, which have not been discussed elsewhere for GALFORM , before examining the clustering of Hα and H-band selected samples in more detail in the next section. In this section we present the basic predictions for the abundance, equivalent width distributions and clustering of Hα emitters. Note that all the results presented here include the attenuation of the Hα emission by dust in the ISM at the same level experienced by the continuum at the wavelength of Hα .

4.3.1 The Hα luminosity function

A basic prediction of the models is the evolution of the Hα luminosity function (LF). Fig. 4.1 shows the Hα LFs predicted by the two versions of GALFORM compared with observational data, over the redshift interval $0.2 < z < 2$. At each redshift plotted, the Bau05 model predicts a higher number density of Hα emitters than the Bow06 model for luminosities brighter than $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) \simeq 42$. This reflects two processes: the relative efficiency of the

feedback mechanisms used in the two models to suppress the formation of bright galaxies, and the top-heavy IMF adopted in starbursts in the Bau05 model, which, for a galaxy with a given star formation rate, boosts the Hα flux emitted. The bright end of the Hα LF is dominated by bursting galaxies.

At faint luminosities, Fig. 4.1 shows that the predicted model LFs are more similar. At these luminosities, the star formation in both models predominantly takes place in galactic disks and produces stars with a standard IMF. For luminosities fainter than $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) \simeq 40$, the Bow06 model suffers from the limited mass resolution of Millennium Simulation halo merger trees (Springel et al. 2005) compared with that of the Monte Carlo trees used in the Bau05 model (Helly et al., 2003).

The observational data shown in Fig. 4.1 comes from Fujita et al. (2003),Hippelein et al. (2003) ,Jones and Bland-Hawthorn (2001), Morioka et al. (2008),Pascual et al. (2001),Shioya et al. (2008) for $z \sim 0.2$; Tresse et al. (2002), Villar et al. (2008),Sobral et al. (2009),Shim et al. (2009),Hopkins et al. (2000) for $z \sim 0.9$, Shim et al. (2009), Yan et al. (1999) for $z = 1.3$ and Geach et al. (2008),Shim et al. (2009), Hayes et al. (2010b) for $z = 1.9$. Most of this observational data have not been corrected by the authors for dust extinction, and hence it can be directly compared to the GALFORM predictions, which include dust attenuation. However, in some cases the data were originally presented after correction for an assumed constant attenuation. In such cases we have undone this "correction". Hence, our comparison concerns the actual observed number of Hα emitters, which is the relevant quantity for assessing the performance of a redshift survey.

In general both models overpredict the number of low luminosity Hα emitters at $z \leq 0.3$, as shown by Fig. 4.1. At $z = 0$, (upper-left panel in Fig. 4.1), the amplitude of the LF in both models is larger, by almost an order of magnitude, than the Jones and Bland-Hawthorn (2001) data. A similar conclusion is reached at $z = 0.2$ (upper-right panel in Fig. 4.1), on comparing the models to most of the observational data. However, there is a significant scatter in observations of the faint end of the LF. At redshifts $z \gtrsim 1$ (bottom panels in Fig. 4.1), the models bracket the observational estimates, with the Bow06 model tending to underpredict the observational LF, whereas the Bau05 model over predicts it. Despite the imperfect agreement, these model LFs "bracket" the observed LFs for the redshifts relevant to space mission surveys propsed, so we proceed to use them for the purposes of this work.

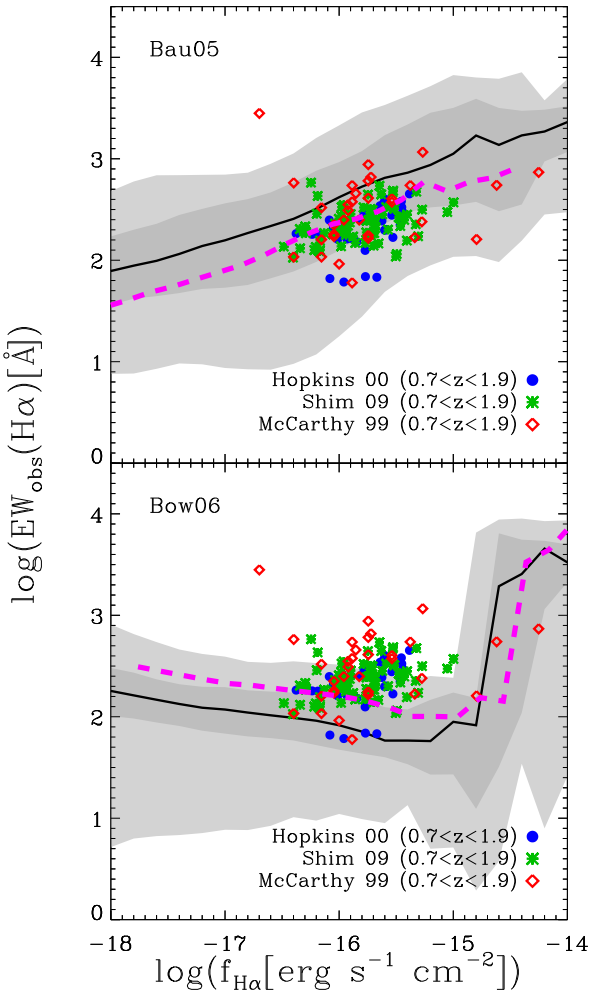

Figure 4.2: The distribution of Hα equivalent width in the observer frame as a function of Hα flux, over the redshift interval $0.7 < z < 1.9$. The top panel shows the predictions of the `Bau05` model and the bottom panel shows the `Bow06` model, calculated as described in the text. The black line shows the median EW at each flux. The shaded regions enclose 68% (dark grey) and 95% (light grey) respectively of the `GALFORM` predictions around the median (black circles). The blue circles show observational data from Hopkins et al. (2000), green asterisks show data from Shim et al. (2009) and red diamonds show data from McCarthy et al. (1999), as indicated by the key. The magenta dashed lines show the `GALFORM` predictions for the median equivalent width after applying the empirically derived continuum flux and line luminosity rescalings described in Section 4.4.

4.3.2 Hα equivalent width (EW) distribution

Broadly speaking the EW of the Hα line depends on the current SFR in a galaxy (which determines the Hα emission), and its stellar mass (to which the continuum luminosity is more closely related). We compare the model predictions for the EW of Hα versus Hα flux with observational results in Fig. 4.2. The observational data cover a wide redshift interval, $0.7 < z < 1.9$ (McCarthy et al., 1999; Hopkins et al., 2000; Shim et al., 2009). In order to mimic the observational selection when generating model predictions, we go through the following two steps. First, we run the models for a set of redshifts covering the above redshift range. Second, we weight the $EW_{\rm obs}$ distribution at a given flux by dN/dz, the redshift distribution of Hα emitters over the redshift range, to take into account the change in the volume element between different redshifts (see Section 4.4 for details of the calculation of dN/dz).

Fig. 4.2 shows the $EW_{\rm obs}$ distribution predicted by the `Bau05` model (top panel) and the `Bow06` model (bottom panel). The models predict different trends of $EW_{\rm obs}$ with Hα flux. In the `Bau05` model, the typical EW increases with Hα flux, with a median value close to $EW_{\rm obs} \sim 100$Å at $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -18$, reaching $EW_{\rm obs} \sim 2000$Å at $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -14$. In contrast, the `Bow06` model predicts a slight decline of $EW_{\rm obs}$ with Hα flux until very bright fluxes are reached, with median $EW_{\rm obs} \sim 100$Å in the range $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = [-18, -15]$. For $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -15$, the `Bow06` model predicts a sharp increase of the median $EW_{\rm obs}$ to ~ 3000Å. The 95% interval of the $EW_{\rm obs}$ found in `GALFORM` galaxies (the light grey region in Fig. 4.2) covers almost 2 orders of magnitude in both models, except in the plateau found in the brightest bin of the `Bow06` model, where the distribution covers 3 orders of magnitude. The `Bau05` model matches the observed distribution of equivalent widths the best, particularly after the rescaling of continuum and line luminosities discussed in the next section (after which the median EW versus Hα distribution shifts from the solid black to the dashed magenta line). It is interesting to note that the "shifted" relations (see §4) give a better match to the observations for both models (although the `Bau05` model remains a better fit), particularly as the shift was derived with reference to the H-band galaxy number counts (for the continuum) and to the $z \sim 1$ Hα LF, rather than to the EW data.

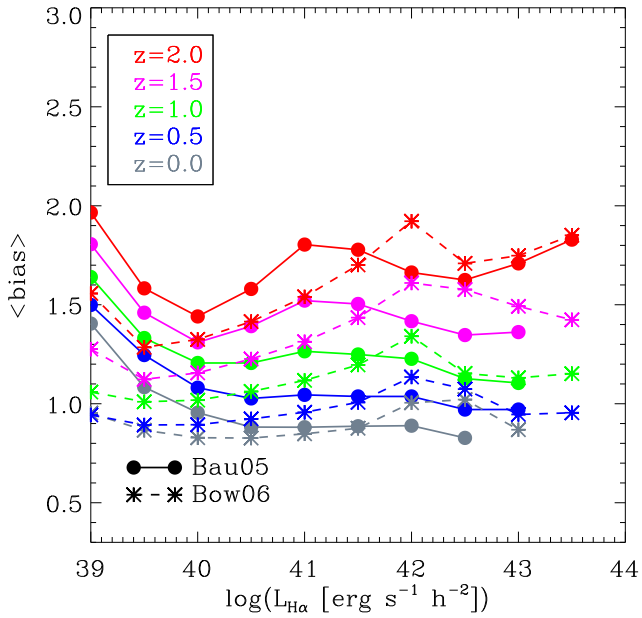

Figure 4.3: The effective bias parameter as a function of Hα luminosity for redshifts spanning the range $0 < z < 2$. The `Bau05` model results are shown using circles connected with solid lines and the `Bow06` model results are shown with asterisks connected by dashed lines. Each colour corresponds to a different redshift, as indicated by the key.

4.3.3 Clustering of Hα emitters: effective bias

The clustering bias, b, is defined as the square root of the ratio of the galaxy correlation function to the correlation function of the dark matter (Kaiser, 1984). As we shall see in Section 4.3, the clustering bias is a direct input into the calculation of the effective volume of a galaxy survey, which quantifies how well the survey can measure the large scale structure of the Universe. Simulations show that the correlation functions of galaxies and dark matter reach an approximately constant ratio on large scales (see for example Angulo et al. 2008a; note, however, that small departures from a constant ratio are apparent even on scales in excess of $100h^{-1}$Mpc).

In this section we compute the effective bias of samples of Hα emitting galaxies. There are theoretical prescriptions for calculating the bias factor of dark matter haloes as a function of mass and redshift (Cole and Kaiser, 1989; Mo and White, 1996; Sheth et al., 2001). These have been extensively tested against the clustering of haloes measured in N-body simulations and have been found to be reasonably accurate (Gao et al., 2005; Wechsler et al., 2006;

Model	$C_{H\alpha}$	$C_{\rm cont}$
Bau05	0.35	0.73
Bow06	1.73	0.42

Table 4.1: Luminosity rescaling factors for the Hα line and the stellar continuum. Column 2 shows $C_{H\alpha}$, the factor used to adjust the predicted Hα flux as described in the text. This factor is only applied to the Hα line. Column 3 shows C_{cont}, the correction factor applied to the stellar continuum, as derived by forcing the model to match the observed H-band counts at $H_{\rm AB} = 22$. This factor is applied to the entire stellar continuum of the model galaxies.

Angulo et al., 2008b). Here we use Sheth et al. (2001). The effective bias is computed by integrating over the halo mass the bias factor corresponding to the dark matter halo which hosts a galaxy multiplied by the abundance of the galaxies of the chosen luminosity, as shown in Chapter 3 (see also Baugh et al., 1999; Le Delliou et al., 2006; Orsi et al., 2008).

Fig. 4.3 shows the predicted galaxy bias, $b_{\rm eff}$, as a function of Hα luminosity over the redshift interval $0 < z < 2$. There is a clear increase in the value of the effective bias with redshift; at $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = 40$, $b_{\rm eff} \approx 0.8$ at $z = 0$, compared with $b_{\rm eff} \approx 1.5$ at $z = 2$. Although the median mass of haloes which host $H\alpha$ emitters decreases with increasing redshift, the characteristic mass of collapsing dark matter haloes, M^*, decreases even faster. Haloes with mass in excess of M^* have a bias factor $b > 1$ and as the ratio $M_{\rm halo}/M^*$ increases the bias also increases. Hence, the halos in which $H\alpha$ emitters are found at higher redshift are more strongly biased than their low redshift counterparts. Both models show an upturn in the effective bias with decreasing luminosity faintwards of $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = 40$. There is little dependence of bias on luminosity brightwards of $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = 40$, up to $z = 2$ because there is a wide spread in the mass of the haloes hosting Hα emitters of a given luminosity, and the median halo mass does not change significantly with luminosity. The predictions of the two models for the effective bias are quite similar. There are currently few observational measurements of the clustering of Hα emitters. Geach et al. (2008) inferred a spatial correlation length of $r_0 = 4.2^{+0.4}_{-0.2} h^{-1}$Mpc for their sample of 55 Hα emitters at $z = 2.23$. This corresponds to a bias of $b \approx 1.7$ in the `Bau05` model cosmology, which is in very good agreement with the predictions plotted in Fig. 4.3.

4.4 The effectiveness of redshift surveys for measuring dark energy

In this section we assess the relative merits of using Hα or H-band selection to construct future redshift surveys aimed at measuring the dark energy equation of state. The first step is to produce a mock catalogue that can reproduce currently available observations. We discuss how we do this in Section 4.1. We then present predictions for the clustering of Hα emitters and H-band selected galaxies in Section 4.2. We quantify the performance of the two selection methods in terms of how well the resulting surveys can measure the large-scale structure of the Universe in Section 4.3.

4.4.1 Building accurate mock catalogues

Our goal in this section is to build mock catalogues for future redshift surveys which agree as closely as possible with currently available observational data. We have already seen that the models are in general agreement with observations of the Hα luminosity function, and will see in the next subsection how well the models match the H-band number counts. In our normal mode of operation, we set the model parameters with reference to a subset of local observations and see how well the model then agrees with other observables. This allows us to test the physics of the model; if the model cannot reproduce a dataset adequately, perhaps some ingredient is missing from the model (e.g. for an application of this principle to galaxy clustering, see Kim et al., 2009). Here our primary aim is not to develop our understanding of galaxy formation physics but to produce a synthetic catalogue which resembles the real Universe as closely as possible. To achieve this end we allow ourselves the freedom to rescale the model stellar continuum and emission line luminosities, independently. This preserves the ranking of the model galaxies in luminosity. This approach is more powerful than an empirical model as we retain all of the additional information predicted by the semi-analytical model, such as the clustering strength of the galaxies. In any case, an empirical calculation of the clustering of the galaxy samples of interest in this work is simply not possible, given the paucity of available clustering measurements. Any empirical estimate would in reality be heavily model dependent, and would be ad hoc compared to the semi-analytical approach. Hereafter we will refer to the adjusted `Bau05` and `Bow06` models as `Bau05(r)` and `Bow06(r)` respectively, to avoid confusion. We also consider a sparsely sampled version

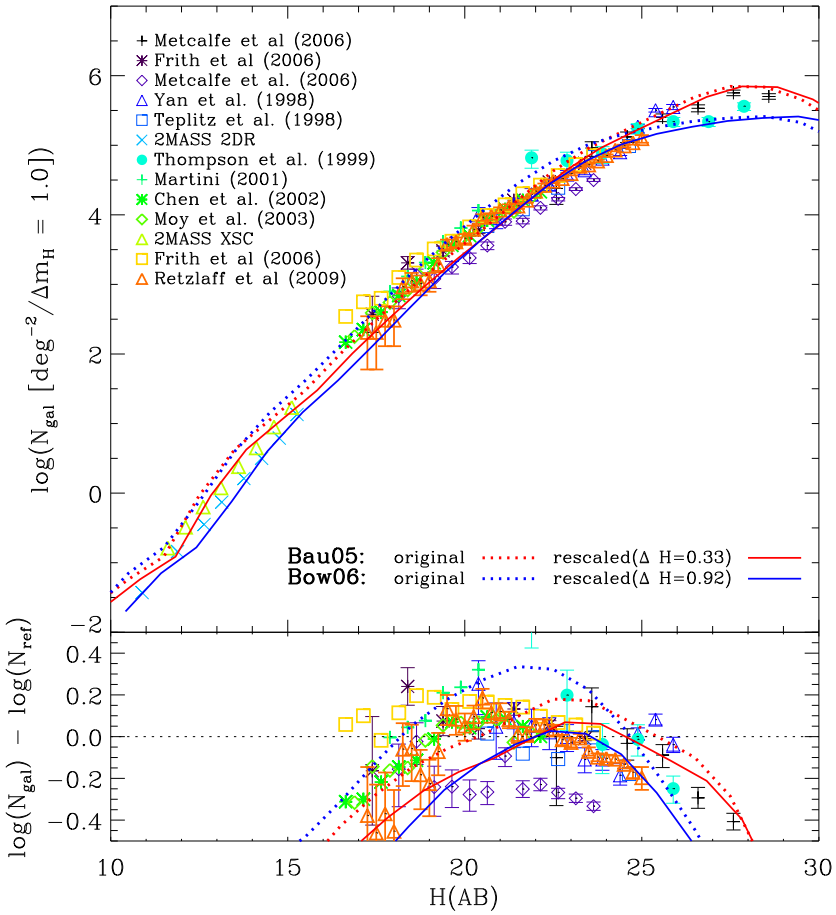

Figure 4.4: Number counts in the H band. The upper panel shows the differential counts on a log scale. The lower panel shows the counts after dividing by a power law $N_{\rm ref} \propto H_{\rm AB}^{0.32}$ to expand the dynamic range on the y-axis. The symbols show the observational data, as shown by the key in the upper panel. The lines show the model predictions. The dotted lines show the original GALFORM predictions for the Bau05 model (blue) and the Bow06 model (red). The solid curves show the rescaled GALFORM predictions after rescaling the model galaxy luminosities to match the observed number counts at $H_{\rm AB} = 22$.

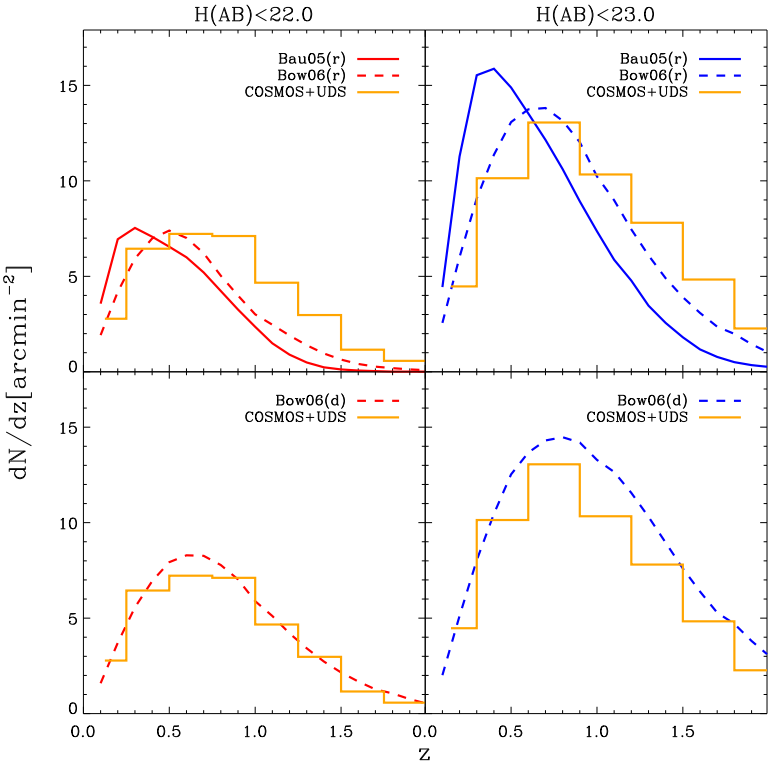

Figure 4.5: The redshift distribution of galaxies with $H_{\rm AB} = 22$ (left column) and $H_{\rm AB} < 23$ (right column). The top panels show the predictions after rescaling the model luminosities to better match the number counts as explained in the text. Red and blue lines show the model predictions for $H_{\rm AB} < 22$ and $H_{\rm AB} < 23$ respectively. Solid lines show the `Bau05(r)` model and the dashed lines show the `Bow06(r)` model. The lower panel shows the redshift distribution obtained from the `Bow06` model by diluting the galaxies, randomly selecting 0.63 of the sample, the `Bow06(d)` model (recall this is a purely illustrative case with no physical basis; see §4.1.1). In both panels, the histogram shows an estimate of the redshift distribution derived from spectroscopic observations in the COSMOS and UDF fields (Cirasuolo et al., 2010, ; Euclid-NIS Science Team, private communication).

of the Bow06 model, which we refer to as Bow06(d) (see §4.1.1).

H-band selected mock catalogues

In Fig. 4.4, we first compare the model predictions without *any* rescaling of the luminosities against a compilation of observed number counts in the H-band, kindly provided by Nigel Metcalfe. Observational data are taken from the following sources, shown with different symbols: Black plus-signs from Metcalfe et al. (2006); purple asterisks from Frith et al. (2006); purple diamonds from Metcalfe et al. (2006); blue triangles from Yan et al. (1998); blue squares from Teplitz et al. (1998); cyan crosses from the second data release of the 2MASS Survey [1]; green circles from Thompson et al. (1999); green plus-signs from Martini (2001); green asterisks from Chen et al. (2002); green diamonds from Moy et al. (2003); green triangles from the 2MASS extended source catalogue[2], orange squares from Frith et al. (2006), and orange triangles from Retzlaff et al. (2010)

There is a factor of three spread in the observed counts around $H_{\rm AB} = 20 - 22$. The unscaled models agree quite well with the observations at $H_{\rm AB} = 20$ but overpredict the counts at $H_{\rm AB} = 22$, the likely depth of a slit-based redshift survey from space. There are two ways in which the model predictions can be brought into better agreement with the observed counts at $H_{\rm AB} = 22$; first, by rescaling the luminosities of the model galaxies to make them fainter in the H-band or second, by artificially reducing, at each magnitude, the number density of galaxies. The first correction could be explained as applying extra dust extinction to the model galaxies; as we will see later on, the typical redshift of the galaxies is $z \sim 0.5$–1, shifting the observer frame H into the rest frame R to V-band.

The second correction has no physical basis and is equivalent to taking a sparse sampling of the catalogue at random, i.e. making a dilution of the catalogue. Galaxies are removed at random without regard to their size or redshift. (Note that the dissolution of galaxies invoked by Kim et al. 2009 only applies to satellite galaxies within haloes, and is mass dependent, and hence is very different from the random dilution applied here.) The motivation behind this second approach is that the shape of the original redshift distribution of the model is preserved. We found that on diluting at random the number of galaxies in the Bow06 model

[1] http://www.ipac.caltech.edu/2mass/releases/second/#skycover

[2] http://www.ipac.caltech.edu/2mass/releases/allsky/doc/sec2_3d3.html

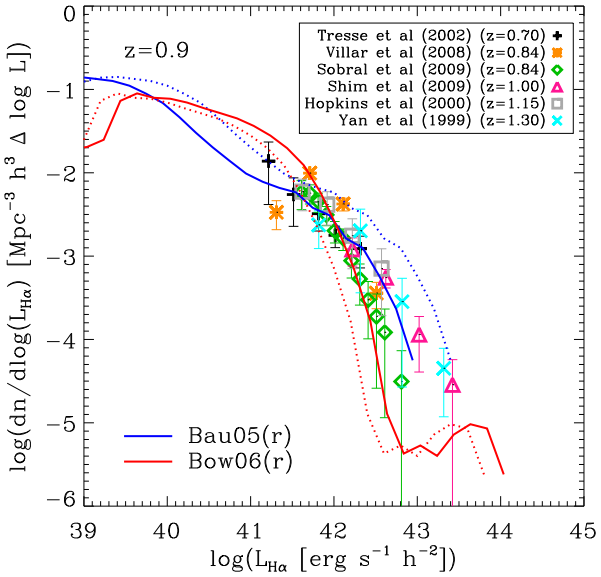

Figure 4.6: The Hα LF at $z = 0.9$. The symbols show observational data, with the sources indicated in the key. The dotted curves show the original predictions for the Hα luminosity function, as plotted in Fig. 4.1. The solid curves show the model predictions after rescaling the Hα luminosity to better match the observed LF at $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = 42$, which corresponds to a flux limit of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -15.3$ at this redshift.

by a factor of 0.63, we can reproduce much better the shape and amplitude of the observed $\mathrm{d}N/\mathrm{d}z$ distribution for $H < 22$ galaxies. As we shall see, the first approach, rescaling the model galaxy luminosities, produces a significant change in the shape of the predicted redshift distribution.

The agreement with the observed counts is improved at $H_{\mathrm{AB}} = 22$ by shifting the `Bow06` galaxy magnitudes faintwards by +0.92 magnitudes; the `Bau05` model requires a more modest dimming of +0.33 magnitudes (see Table 4.1).

The redshift distribution of H-band selected galaxy samples provides a further test of the models. In Fig. 4.5, the model predictions are compared against an estimate of the redshift distribution compiled using observations from the COSMOS survey and the Hubble Ultra-Deep Field for $H_{\mathrm{AB}} < 22$ and $H_{\mathrm{AB}} < 23$ (Cirasuolo et al. 2010; Cirasuolo, Le Fevre and McCracken, private communication). If we focus on the lower panels first, which shows $\mathrm{d}N/\mathrm{d}z$ in the randomly diluted `Bow06` model, denoted as `Bow06(d)`, it is apparent that the original `Bow06` model predicted the correct shape for the redshift distribution of sources, but

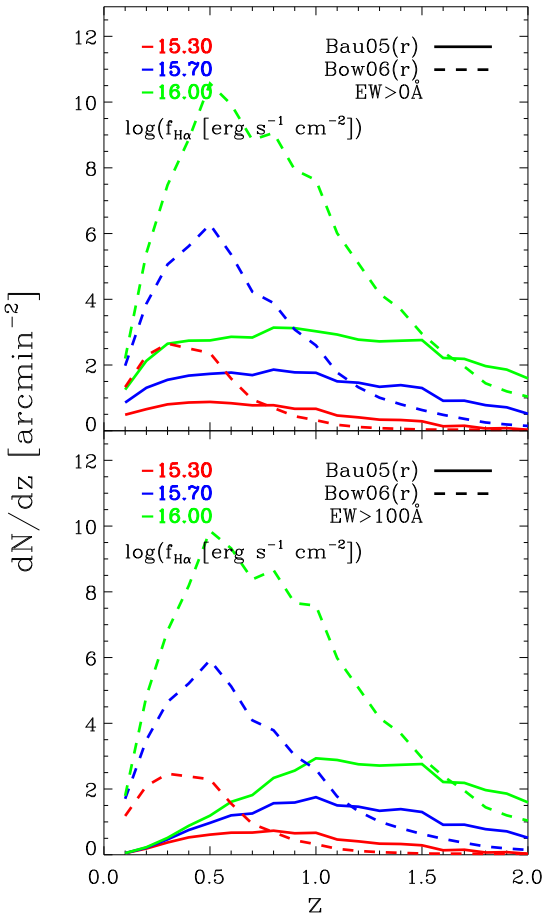

Figure 4.7: The redshift distribution of Hα selected galaxies for 3 different flux limits: log($F_{H\alpha}$[erg s^{-1} cm^{-2}]) > -15.3, -15.7 and -16.0 shown in red, blue and green respectively. The solid lines show the `Bau05(r)` prediction and the dashed lines show the `Bow06(r)` predictions. In the top panel, galaxies contributing to the redshift distribution have no cut imposed on the equivalent width of Hα . In the bottom panel, the model galaxies have to satisfy the Hα flux limit and a cut on the observed equivalent width of Hα of $EW_{\rm obs} > 100$Å.

with simply too many galaxies at each redshift. In the upper panel of Fig. 4.5, we see that the models with the shifted H-band luminosities give shallower redshift distributions than the observed one. The difference between the predicted dN/dz after dimming the luminosities or diluting the number of objects has important implications for the number density of galaxies as a function of redshift, which in turn is important for the performance of a sample in measuring the large-scale structure of the Universe.

Hα-selected mock catalogues

The original model predictions for the Hα luminosity function were presented in Fig. 4.1. The models cross one another *and* match the observed Hα LF at a luminosity of $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) \sim 41.5$. At $z = 0.9$, this corresponds to a flux of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -15.8$. The flux limit attainable by Euclid is likely to be somewhat brighter than this, although the precise number is still under discussion. For this reason, we chose to force the models to agree with the observed Hα LF at $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = 42$ at $z = 0.9$, which corresponds to a flux limit of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -15.3$ (see Fig. 4.6). Before rescaling, the model LFs differ by a factor of three at $\log(L_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) \sim 42$. In the rescaling, the Hα line luminosity is boosted in the `Bow06` model and reduced in the case of the `Bau05` model (see Table 4.1 for the correction factors used in both cases). The latter could be explained as additional dust extinction applied to the emission line, compared with the extinction experienced by the stellar continuum. The former correction, a boost to the Hα luminosity in the `Bow06` model, is harder to explain. This would require a boost in the production of Lyman-continuum photons (e.g. as would result on invoking a top-heavy IMF in starbursts or an increase in the star formation rate). This would require a revision to the basic physical ingredients of the model and is beyond the scope of the current thesis.

After making this correction to the Hα line flux in the models, we next present the predictions for the redshift distribution of Hα emitters. Fig. 4.7 shows dN/dz for flux limits of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = [-15.7, -16.0, -16.3]$. The redshift distribution of the `Bow06(r)` model peaks around $z \sim 0.5$ and declines sharply approaching $z \sim 2$, whereas the `Bau05(r)` dN/dz are much broader. The lower panel of Fig. 4.7 shows the redshift distribution after applying the flux limits and a cut on the observed equivalent width of $EW_{\mathrm{obs}} = 100$Å. (Note that the dN/dz is not sensitive to low EW cuts; similar results to the $EW_{\mathrm{obs}} > 0$ Å case are obtained with 10Å in both models). In the rescaled model, the equivalent width changes be-

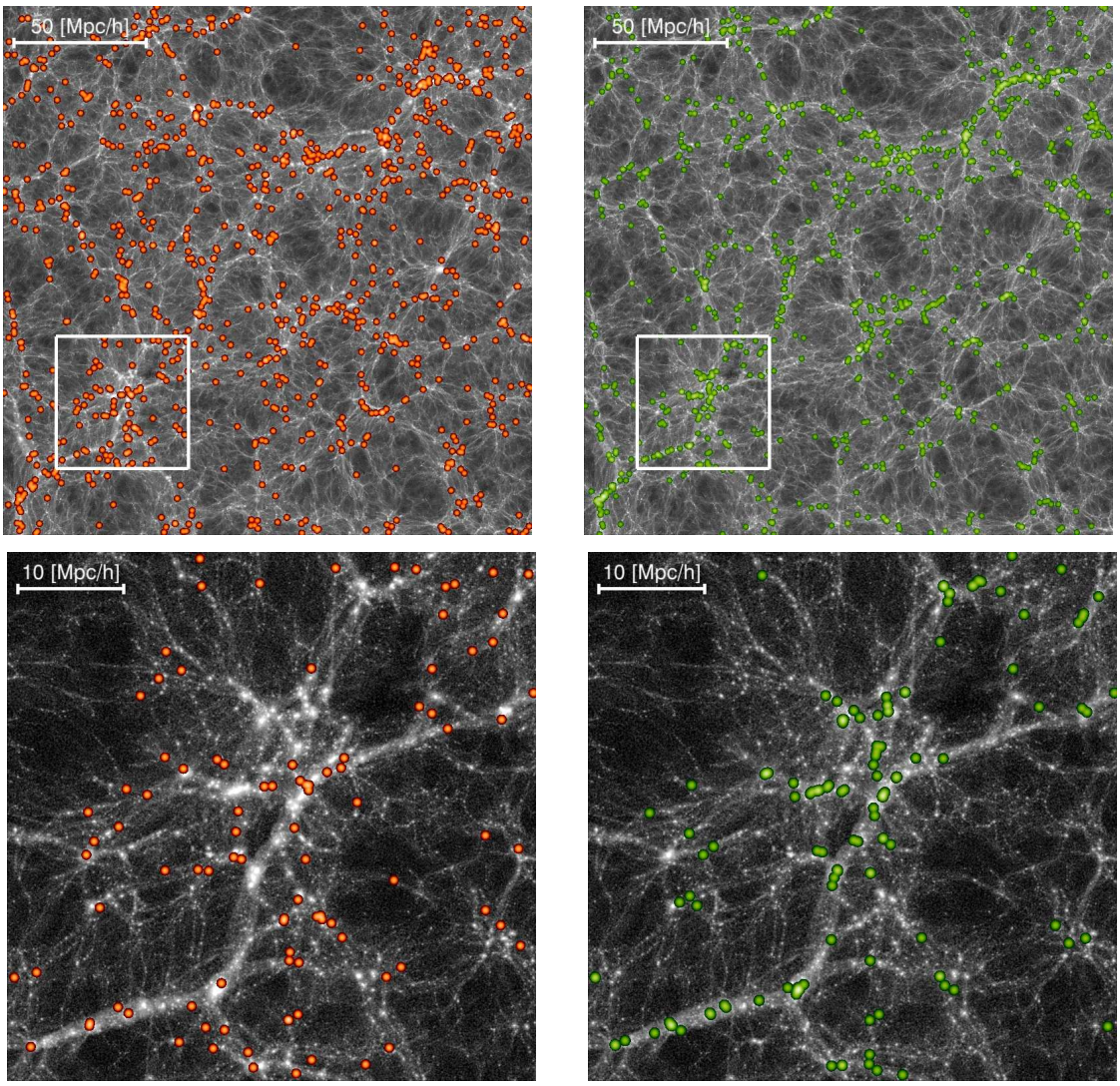

Figure 4.8: The spatial distribution of galaxies and dark matter in the `Bow06(r)` model at $z = 1$. Dark matter is shown in grey, with the densest regions shown with the brightest shading. Galaxies selected by their Hα emission with $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16.00$ and and $EW_{\mathrm{obs}} > 100$Å are shown in red in the left-hand panels. Galaxies brighter than $H_{\mathrm{AB}} = 22$ are shown in green in the right-hand panels. Each row shows the same region from the Millennium simulation. The first row shows a slice of $200h^{-1}$Mpc on a side and $10h^{-1}$Mpc deep. The second row shows a zoom into a region of $50h^{-1}$Mpc on a side and $10h^{-1}$Mpc deep, which corresponds to the white square drawn in the first row images. Note that all of the galaxies which pass the selection criteria are shown in these plots.

cause the Hα line flux has been adjusted and because the continuum has been altered (by the same shift as applied to the H-band). Adding the selection on equivalent width results in a modest change to the predicted dN/dz in the `Bow06(r)` model. In the `Bau05(r)` model, the dN/dz shifts to higher redshifts. There is no observational data on the redshift distribution of Hα emitters to compare against the model predictions. Geach et al. (2010) make an empirical estimate of the redshift distribution, by fitting a model for the evolution of the luminosity function to observational data. The luminosity of the characteristic break in the luminosity function, L_*, is allowed to vary, while the faint end slope and normalisation are held fixed. The resulting empirical LF looks similar to the original `Bau05` model at $z = 0.9$, and the two have similar redshift distributions. The Hα redshift distributions in the `Bow06(r)` models are shallower than the empirical estimate; the `Bau05(r)` model has a similar shape to the empirical redshift distribution, but with a lower normalisation. It is important to realise that the approach of Geach et al. is also model dependent, and the choices of model for the evolution of the luminosity function and of which observational datasets to match are not unique and will have an impact on the resulting form of the redshift distribution.

4.4.2 The clustering of Hα and H-band selected samples

The semi-analytic galaxy formation model predicts the number of galaxies hosted by dark matter haloes of different mass. In the cases of Hα emission, which is primarily sensitive to ongoing star formation, and H-band light, which depends more on the number of long-lived stars, different physical processes determine the number of galaxies per halo. The model predicts contrasting spatial distributions for galaxies selected according to their Hα emission or H-band flux. We compare in Fig. 4.8 the spatial distribution of Hα emitters with fluxes $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ and $EW_{\mathrm{obs}} > 100$Å (red circles) with that of an H-band selected sample with $H_{\mathrm{AB}} < 22$ (green circles), in the `Bow06(r)` model which is set in the Millennium Simulation. The upper panels of Fig. 4.8 show how the different galaxy samples trace the underlying cosmic web of dark matter. The lower panels of Fig. 4.8 show a zoom into a massive supercluster. There is a marked difference in how the galaxies trace the dark matter on these scales. The Hα emitters avoid the most massive dark matter structures. At the centre of massive haloes, the gas cooling rate is suppressed in the model due to AGN heating of the hot halo. This reduces the supply of gas for star formation and in turn cuts the rate of production of Lyman continuum photons, and hence the Hα emission. The H-band

selected galaxies, on the other hand, sample the highest mass dark matter structures.

To study the difference in the spatial distribution of galaxies in a quantitative way, we compare the clustering predictions from the models with observational data. We use the same method explained in Section 3.3 to calculate the effective bias, i.e. using the Sheth et al. (2001) analytical bias, and use this to derive the correlation length, r_0, a measure of the clustering amplitude, which we define as the pair separation at which the correlation function equals unity. The correlation function of galaxies, $\xi_{\rm gal}$, is related to the correlation function of dark matter, $\xi_{\rm dm}$, by $\xi_{\rm gal} = b^2\xi_{\rm dm}$. The effective bias is *approximately* constant on large scales (e.g. Angulo et al. 2008a). We use the Smith et al. (2003) prescription to generate a nonlinear matter power spectrum in real space. This in turn is Fourier transformed to obtain the two-point correlation function of the dark matter, $\xi_{\rm dm}$. We can then derive $\xi_{\rm gal}$ for any survey configuration by multiplying $\xi_{\rm dm}$ by the square of the effective bias, and then we read off the correlation length as the scale at which the correlation function is equal to unity.

Fig. 4.9 shows the correlation length in comoving units for both Hα and H-band samples at different redshifts, compared to observational estimates. Differences in the bias predicted by the two models (as shown in Fig. 4.3) translate into similar differences in r_0. The correlation length declines with increasing redshift for Hα emitters in the `Bau05(r)` model, since the increase of the effective bias with redshift is not strong enough to balance the decline of the amplitude of clustering of the dark matter. For the range of flux limits shown in the top panel of Fig. 4.9 ($-16 < \log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) < -17$), r_0 changes from $\sim 5-7\ h^{-1}$Mpc at $z = 0.1$ to $r_0 \sim 3.5\ h^{-1}$Mpc at $z = 2.5$. On the other hand, the `Bow06(r)` model shows a smooth increase of r_0 which depends on flux and redshift. At bright flux limits r_0 evolves rapidly at high redshift, reaching $r_0 = 4.3h^{-1}$Mpc at $z = 2.5$. At fainter luminosities the change in correlation length with redshift is weaker.

The currently available observational estimates of the clustering of near infrared selected galaxy samples mainly come from angular clustering. A number of assumptions are required in order to derive a spatial correlation length from the angular correlation function. First, a form must be adopted for the distribution of sources in redshift. Second, some papers quote results in terms of proper separation whereas others report in comoving units. Lastly, an evolutionary form is sometimes assumed for the correlation function (Groth and Peebles, 1977). In this case, the results obtained for the correlation length depend upon the choice of

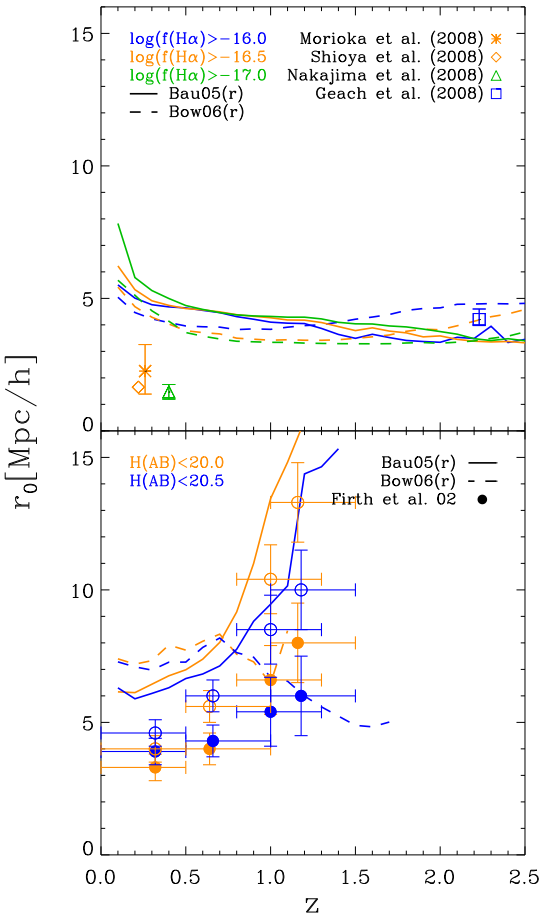

Figure 4.9: The correlation length, r_0, as a function of redshift for selected Hα and H-band samples. Solid and dashed lines show the predictions of the `Bau05` and `Bow06` models respectively. The top panel shows the predictions for different Hα limiting fluxes, $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > [-16.0, -16.5, -17.0]$ in green, orange and blue respectively. Observational data is shown with symbols. The bottom panel shows the model predictions for $H_{AB} < [20., 20.5]$ in orange and blue respectively. In this case there are two sets of observational estimates, based on different assumptions for the evolution of clustering with redshift.

evolutionary model. Hence, estimates of the spatial correlation length derived from angular clustering data are model dependent. Moreover, in the majority of cases, the errors on the inferred correlation length include neither the impact of different model choices nor the contribution of sample variance due to the small volumes typically available.

Estimates of the correlation length of Hα emitters are available at a small number of redshifts from narrow band surveys, as shown in Fig. 4.9 (Morioka et al., 2008; Shioya et al., 2008; Nakajima et al., 2008; Geach et al., 2008). As remarked upon in the previous paragraph, these surveys are small and sampling variance is not always included in the error bar quoted on the correlation length (see Orsi et al. 2008 for an illustration of how sampling variance can affect measurements of the correlation function made from small fields). The models are in reasonable agreement with the estimate by Geach et al. (2008) at $z = 2.2$, but overpredict the low redshift measurements. The $z = 0.24$ measurements are particularly challenging to reproduce in any viable hierarchical clustering model. The correlation length of the dark matter in the ΛCDM model is around $5h^{-1}$Mpc at this redshift, so the $z = 0.24$ result implies an effective bias of $b < 0.5$. Gao & White (2007) show that even the lowest mass dark matter haloes at the resolution limit of the Millennium Simulation, $M \sim 10^{10}h^{-1}M_{\odot}$, do not reach this level of bias, unless the 20% of the youngest haloes of this mass are selected. In the `Bow06(r)` model, the Hα emitters populate a range of halo masses, with a spread in formation times, and so the effective bias is closer to unity. Another possible explanation for the discrepancy is that the observational sample could be contaminated by objects which are not Hα emitters and which dilute the clustering signal. (For reference we note that the correlation length of the dark matter in the cosmology of the Millennium Simulation is $r_0 = 2.8h^{-1}$Mpc at $z = 1$ and $r_0 = 1.6h^{-1}$Mpc at $z = 2$.)

The bottom panel of Fig. 4.9 shows the correlation length evolution for different H-band selections, compared to observational estimates from Firth et al. (2002). Note that the samples analysed by Firth et al. are significantly brighter than the typical samples considered in this work ($H_{\rm AB} = 20$ versus $H_{\rm AB} = 22$). Firth et al. use photometric redshifts to isolate galaxies in redshift bins before measuring the angular clustering. Two sets of observational estimates are shown for each magnitude limit, corresponding to two choices for the assumed evolution of clustering. Again the models display somewhat stronger clustering than the observations would suggest at low redshift. The `Bau05(r)` model predicts a clustering length which increases with redshift. The `Bow06(r)` model, on the other hand, predicts a peak in

the correlation length around $z \sim 0.7$, with a decline to higher redshifts. This reflects the form of the luminosity - halo mass relation for galaxy formation models with AGN feedback (Kim et al., 2009). The slope of the luminosity - mass relation changes at the mass for which AGN heating becomes important. Coupled with the appreciable scatter in the predicted relation, this can result in the brightest galaxies residing in haloes of intermediate mass.

Whilst the comparisons between models and clustering measurements presented in this section admittedly seem less than impressive, we continue to use the model clustering predictions in the remainder of this work. As we have already remarked, the currently available correlation length data is extracted from angular clustering and hence is itself model dependent. The errors quoted on the correlation lengths do not generally take this into account, nor do they include the impact of sample variance, which we have previously demonstrated can be significant for samples of the size under consideration, as shown in Chapter 3 (see also Orsi et al., 2008). There *is* no empirical way using currently available data to estimate the clustering strength of the samples of interest in this work. Any such attempt would require substantial extrapolation from the uncertain existing data and would therefore become model dependent. In our opinion, the semi-analytical approach with its physical basis offers a more reliable route to take to make clustering predictions for future space missions.

4.4.3 Redshift-space distortions

The amplitude of gravitationally induced bulk flows is sensitive to the rate at which perturbations grow, which depends on the expansion history of the universe and the nature of the dark energy (Wang, 2008; Guzzo et al., 2008). Bulk flows can be measured by their impact on the correlation function of galaxies when plotted as a function of pair separation perpendicular and parallel to the line of sight, $\xi(r_\sigma, r_\pi)$ (Hawkins et al., 2003; Ross et al., 2007). We now restrict our attention to the `Bow06(r)` model, since this is set in the Millennium Simulation and we can measure the clustering of the model galaxies directly. As the Millennium simulation has periodic boundary conditions, we can estimate the correlation function as follows:

$$\xi(r_\sigma, r_\pi) = \frac{DD_{\sigma,\pi}}{N\bar{n}\Delta V_{\sigma,\pi}} - 1, \tag{4.1}$$

$$\Delta V_{\sigma,\pi} = 2\pi r_\sigma \Delta r_\sigma \Delta r_\pi, \tag{4.2}$$

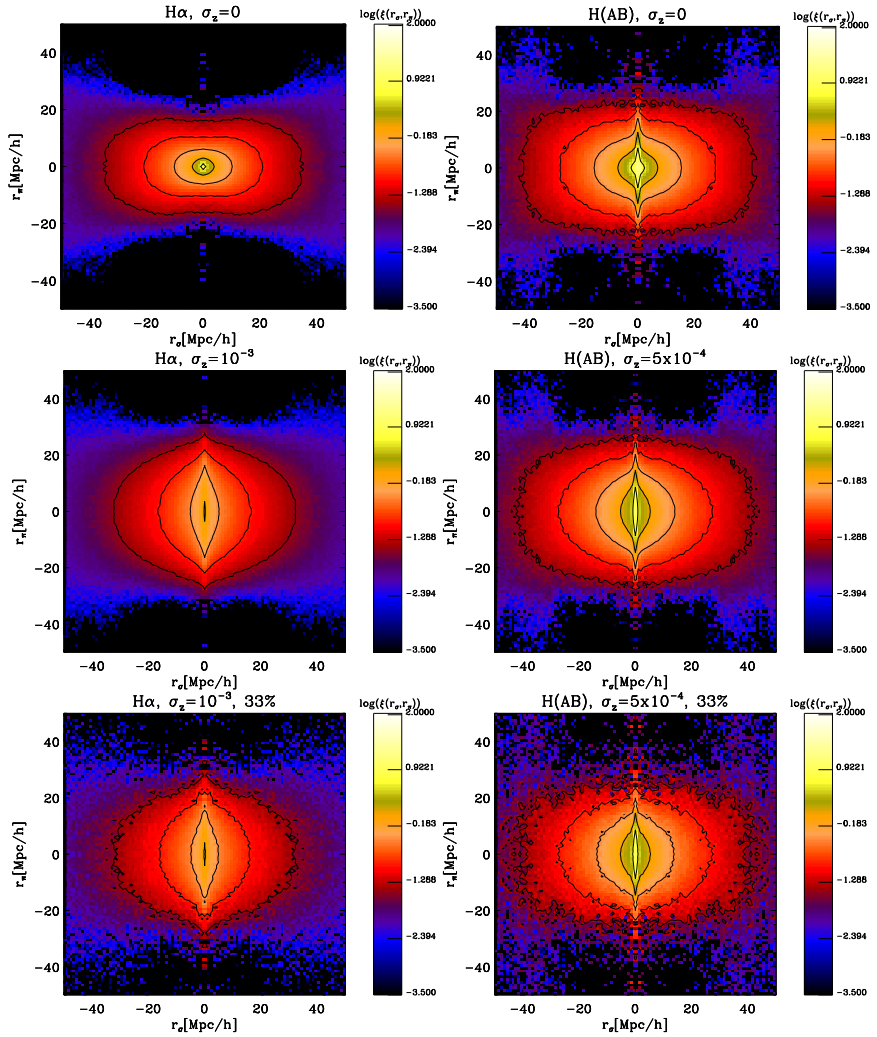

Figure 4.10: The two point correlation function, measured in redshift space, plotted in bins of pair separation parallel (r_π) and perpendicular (r_σ) to the line of sight, $\xi(r_\sigma, r_\pi)$, for Hα emitters (left-hand panels) and H-band selected (right-hand panels) galaxies in the Millennium simulation. The samples used are those plotted in Fig. 4.8. The pair counts are replicated over the four quadrants to enhance the visual impression of deviations from circular symmetry. The Hα catalogue has a limiting flux of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ and an equivalent width cut of $EW_{\mathrm{obs}} > 100$Å; the H-band magnitude limit is $H_{\mathrm{AB}} = 22$. The contours show where $\log(\xi(r_\sigma, r_\pi)) = [0.5, 0.0, -0.5, -1.0, -1.5]$, from small to large pair separations. The upper panels show the correlation function measured in fully sampled catalogues without redshift errors. The middle panels show how redshift errors change the clustering pattern. Representative errors for the two redshift measurements are used: $\sigma_z = 10^{-3}$ for the slitless case (Hα emitters), and $\sigma_z = 5 \times 10^{-4}$ for the slit based measurement (H-band selected). In the upper and middle panels, all the galaxies are used to compute the correlation function. In the bottom panels, only 33% of the galaxies are used in each case, which is indicative of the likely redshift success rate for a survey from space.

where $DD_{\sigma,\pi}$ is the number of distinct galaxy pairs in a bin of pair separation centred on (r_σ, r_π), Δr_σ and Δr_π are the widths of the bins in the r_σ and r_π directions, respectively, N and $\bar{n}$ are the total number and the number density of galaxies in the sample, and $\Delta V_{r_\sigma, r_\pi}$ corresponds to the volume enclosed in an annulus centred on (r_σ, r_π). Note that to avoid any confusion, here we refer to the line of sight separation as r_π and use π to denote the mathematical constant.

In redshift surveys, the radial distance to a galaxy is inferred from its redshift. The measured redshift contains a contribution from the expansion of the Universe, along with a peculiar velocity which is induced by inhomogeneties in the density field around the galaxy. Thus the position inferred from the redshift is not necessarily the true position. The distortion of the clustering pattern resulting from peculiar velocities is referred to as the redshift space distortion. On large scales, coherent motions of galaxies from voids towards overdense regions lead to a boost in the clustering amplitude (Kaiser, 1987):

$$\frac{\xi(s)}{\xi(r)} = 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2, \tag{4.3}$$

where $\xi(s)$ is the spherically averaged, redshift space correlation function, and $\xi(r)$ is its equivalent in real space (i.e. without the contribution of peculiar velocities). Eq. (4.3) holds in linear perturbation theory in the distant observer approximation when gradients in the bulk flow and the effect of the velocity dispersion are small (Cole et al., 1994; Scoccimarro, 2004). Strictly speaking, these approximations apply better on large scales. The parameter β is related to the linear growth rate, D, through

$$\beta_{\rm lin} = \frac{1}{b}\frac{{\rm d}\ln D}{{\rm d}\ln a}, \tag{4.4}$$

$$\approx \frac{\Omega_{\rm m}(z)^\gamma}{b}, \tag{4.5}$$

where a is the expansion factor. The approximation in Eq. (4.5) is valid for an open cosmology, in which γ is traditionally approximated to 0.6 (Peebles, 1980). Lahav et al. (1991) showed that this approximation should be modified in the case of a CDM model with a cosmological constant, to display a weak dependence on Λ. Lue et al. (2004) pointed out that the value of γ allows one to differentiate between modified gravity and dark energy, since $\beta(z) \simeq \Omega_m(a)^{2/3}/b$ for DGP gravity models, while $\beta(z) \simeq \Omega_m(a)^{5/9}/b$ for a flat Universe with a cosmological constant.

On small scales, the randomised motions of galaxies inside virialised structures lead to

a damping of the redshift space correlation function and a drop in the ratio $\xi(s)/\xi(r)$(Cole et al., 1994).

We include redshift space distortions in the model by applying the distant observer approximation and taking one of the cartesian axes of the simulation cube as the line of sight. The peculiar velocity of a galaxy along the chosen axis is added to its comoving position along the same axis, after applying the appropriate scaling to change from velocity units. The impact of peculiar velocities on the clustering of galaxies is clearly seen in $\xi(r_\sigma, r_\pi)$. The top panels of Fig. 4.10 show the correlation function of Hα emitters selected to have $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ and $EW_{\mathrm{obs}} > 100$Å (left) and H-band selected galaxies with $H_{\mathrm{AB}} < 22$ (right). In the top and middle rows of Fig. 4.10, all galaxies are used down to the respective flux limits. To obtain clustering in redshift space, we use the distant observer approximation and give the galaxies a displacement along one of the cartesian axes, as determined by the component of the peculiar velocity along the same axis. Without peculiar velocities, contours of constant clustering amplitude in $\xi(r_\sigma, r_\pi)$would be circular. In redshift space, the clustering of H-band selected galaxies exhibits a clear signature on small scales of a contribution from high velocity dispersion systems – the so called "fingers of God". This effect is less evident in the clustering of the Hα sample, as these galaxies avoid massive haloes, as shown in Fig. 4.8. On large scales, the contours of equal clustering are flattened due to coherent flows. Similar distortions have been measured in surveys such as the 2dFGRS (Hawkins et al. 2003) and the VLT-VIMOS deep survey (Guzzo et al. 2008).

In practice, the measured correlation functions will look somewhat different to the idealised results presented in the top row of Fig. 4.10. The redshift measurements will have errors, and the errors for slitless spectroscopy are expected to be bigger than those for slit-based spectroscopy (Euclid-NIS team, private communication). We model this by adding a Gaussian distributed velocity, v_r, to the peculiar velocities following $\delta z = (1+z)v_r/c$. The dispersion of the Gaussian is parametrized by $\sigma_z \equiv \langle \delta z^2 \rangle^{1/2}/(1+z)$. We show the impact on the predicted clustering of adding illustrative redshift uncertainties to the position measurements in the middle and bottom panels of Fig. 4.10. For Hα-emitters, we chose a fiducial error of $\sigma_z = 10^{-3}$, based on simulations by the `Euclid` NIS team. The errors on the slit-based redshifts are expected to be at least a factor of 2 times smaller than the slitless errors, so we set $\sigma_z = 5 \times 10^{-4}$ for the H_{AB} selected sample. The impact of the redshift errors is most prominent in the case of the Hα sample, where the contours of constant clustering become

more elongated along the line-of-sight direction.

A measure of how well bulk flows can be constrained can be gained from the accuracy with which β can be measured (Eq. (4.4)). We estimate β by applying Eq. (4.3) to the ratio of the redshift space to real space correlation function on pair separations between $15-30h^{-1}$Mpc, which is close to the maximum pair separation out to which we can reliably measure clustering in the Millennium simulation volume. The introduction of redshift errors forces us to apply Eq. (4.3) to the measurements from the Millennium simulation on larger scales than in the absence of errors. We note that the ratio is noisy even for a box of the volume of the Millennium, and we therefore average the ratio by treating each of the cartesian axes in turn as the line of sight direction. The real space correlation function is difficult to estimate on large scales, so a less direct approach would be applied to actual survey data (see e.g. Guzzo et al., 2008). (For a comprehensive discussion of how to estimate β see Branchini et al., in preparation.) Hence, our results will be on the optimistic side of what is likely to be attainable with future surveys. Ideally, we would like to apply Eq. (4.5) on as large a scale as possible. Kaiser's derivation assumes that the perturbations are in the linear regime.

We solve the integral for the growth rate D in Eq. (4.4) (see Lahav et al., 1991) and use this exact result with the value of the bias b measured for each galaxy sample to get the theoretical value $\beta_{\rm lin}$. Table 4.2 shows the comparison between β_m, the measured value of β in the simulation, and target theoretical value $\beta_{\rm lin}$. Two different selection cuts are chosen for both Hα and H-band samples to cover a range of survey configurations: $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > [-15.4,-16.0]$ for Hα samples and $H_{AB} < [22,23]$ for the magnitude limited samples. All the mock catalogues studied return a value for β_m which is systematically below the expected value, $\beta_{\rm lin}$.

When redshift errors are omitted and a 100% redshift success rate is used, both selection methods seem to reproduce the expected value of $\beta_{\rm lin}$ to within better than $\sim 10\%$. When redshift errors are included, the spatial distribution along the line of sight appears more elongated than it would be if the true galaxy positions could be used. This leads to an increase in the small scale damping of the clustering. However, at the same time contours of constant clustering amplitude are pushed out to larger pair separations in the radial direction. This results in an increase in the ratio of redshift space to real space clustering and an increase in the recovered value of β. When including the likely redshift errors, the values of β_m found are slightly higher than those without redshift errors. This small boost in the value of β_m is

Table 4.2: Values of β estimated from the ratio of the redshift space to real space correlation function for the fiducial samples at $z = 1$. We consider Hα emitters with fluxes $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > [-15.4, -16]$ and H-band selected galaxies with $H_{AB} < [22, 23]$. The table is divided into two parts. The first half assumes a redshift success rate of 100% and the second a 33% redshift success rate. Each segment is divided into two, showing the impact on β of including the expected redshift uncertainties: $\sigma_z = 10^{-3}$ for Hα emitters and $\sigma_z = 5 \times 10^{-4}$ for H-band selected samples. Column (1) shows $\beta_{\rm lin}$, the exact theoretical value of β obtained when using Eq. (4.4). Column (2) shows β_m, the value of β measured in the simulation including the 1 σ error. Column (3) shows the fractional error on β_m using the Millennium volume. Column (4) shows the fractional error on β_m obtained when using mock catalogues from the BASICC simulation.

	(1)	(2)	(3)	(4)
	β_{lin}	β_m	$(\delta\beta_m/\beta_{lin})$ Millennium	$(\delta\beta_m/\beta_{lin})$ BASICC
Sampling rate = 100%				
$\log(F(H\alpha)) > -15.4, \sigma_z = 0$	0.761	0.684 ± 0.153	0.201	0.125
$\log(F(H\alpha)) > -16.0, \sigma_z = 0$	0.821	0.766 ± 0.027	0.034	0.021
$H(AB) < 22, \sigma_z = 0$	0.521	0.491 ± 0.026	0.051	0.019
$H(AB) < 23, \sigma_z = 0$	0.565	0.536 ± 0.013	0.023	0.013
$\log(F(H\alpha)) > -15.4, \sigma_z = 10^{-3}$	0.761	0.768 ± 0.170	0.224	0.122
$\log(F(H\alpha)) > -16.0, \sigma_z = 10^{-3}$	0.821	0.825 ± 0.058	0.071	0.081
$H(AB) < 22, \sigma_z = 5 \times 10^{-4}$	0.521	0.527 ± 0.029	0.057	0.012
$H(AB) < 23, \sigma_z = 5 \times 10^{-4}$	0.565	0.569 ± 0.012	0.022	0.008
Sampling rate = 33%				
$\log(F(H\alpha)) > -15.4, \sigma_z = 0$	0.634	0.123 ± 0.447	0.704	0.449
$\log(F(H\alpha)) > -16.0, \sigma_z = 0$	0.807	0.680 ± 0.104	0.129	0.033
$H(AB) < 22, \sigma_z = 0$	0.516	0.482 ± 0.049	0.095	0.036
$H(AB) < 23, \sigma_z = 0$	0.568	0.569 ± 0.029	0.051	0.018
$\log(F(H\alpha)) > -15.4, \sigma_z = 10^{-3}$	0.634	0.300 ± 0.216	0.341	0.341
$\log(F(H\alpha)) > -16.0, \sigma_z = 10^{-3}$	0.807	0.749 ± 0.118	0.146	0.078
$H(AB) < 22, \sigma_z = 5 \times 10^{-4}$	0.516	0.494 ± 0.061	0.118	0.023
$H(AB) < 23, \sigma_z = 5 \times 10^{-4}$	0.568	0.603 ± 0.028	0.050	0.012

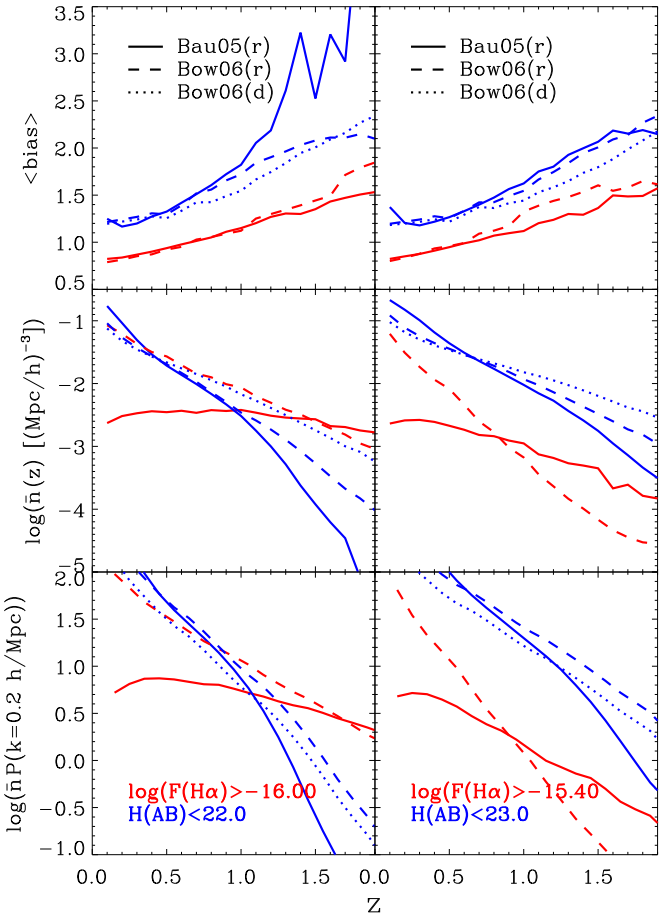

Figure 4.11: The effective bias (top panel), number density of galaxies (middle panel) and the product $\bar{n}P$ (bottom panel) as functions of redshift, where P is measured at wavenumber $k = 0.2$ Mpc/h. The solid lines show the predictions for the `Bau05(r)` model and the `Bow06(r)` model is shown using dashed lines. The two columns show different Hα and H-band selections: In the first column the Hα sample is defined by a limiting flux of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ and $EW_{\mathrm{obs}} > 100$Å (red curves). The magnitude limited sample has $H_{AB} < 22$ (blue curves). In the second column the Hα sample has $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -15.4$ and $EW_{\mathrm{obs}} > 100$Å, and the H-band sample has $H(AB) < 23$. In all panels the redshift success rate considered is 100%.

greatest in the Hα sample, because of the larger redshift errors than in the H-band sample.

We have also tested the impact of applying different redshift success rates on the determination of β_m. The lower part of Table 4.2 shows the impact of a 33% redshift success rate. For $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -15.4$, our results for β_m shows that it is unlikely to get a robust estimate of β at this flux limit, because the smaller number density makes the correlation functions very noisy, thus making β_m impossible to be measured correctly. In contrast, the impact of a 33% of success rate in the $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ sample is negligible. The β_m values calculated using the H-band catalogues are also mostly unaffected. When redshift uncertainties are considered, as before, the β_m values are closer to the theoretical β_{lin}. Hence redshift uncertainties will contribute to the uncertainty on β_m, but they still permit an accurate determination of β, provided they do not exceed $\sigma_z = 10^{-3}$.

The noisy correlation functions for the configurations with $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -15.4$ and sampling rate of 33% produce measurements of β_m with large errors. The mock catalogues used so far in this section were created from the Millennium simulation, which has $V_{\mathrm{Mill}} = 500^3[\mathrm{Mpc/h}]^3$. This volume is almost three orders of magnitude smaller than the volume expected in a large redshift survey from a space mission like `Euclid`(see next section). In order to test the impact of using this limited volume when measuring β_m and its error, we plant the `Bow06(r)` model into a larger volume using the BASICC N-body simulation (Angulo et al., 2008a), which has a volume almost 20 times larger than the Millennium run ($V_{\mathrm{BASICC}} = 1340^3[\mathrm{Mpc/h}]^3$). The errors on β_m shown in Table 4.2 are expected, to first order, to scale with the error on the power spectrum (see Eq. (4.6) below). If we compare two galaxy samples with the same number density but in different volumes, then the error on β_m should scale as $\delta\beta \propto 1/\sqrt{V}$, where V is the volume of the sample.

The only drawback of using the BASICC simulation is that the mass resolution is worse than in the Millennium simulation. Haloes with mass greater than $5.5 \times 10^{11} M_\odot/h$ can be resolved in the BASICC simulation. The galaxy samples studied here are hosted by haloes with masses greater than $\sim 8 \times 10^{10} M_\odot/h$, so if we only plant galaxies into haloes resolved in the BASICC run then we would miss a substantial fraction of the galaxies. To avoid this incompleteness, those galaxies which should be hosted by haloes below the mass resolution limit are planted on randomly selected ungrouped particles, i.e. dark matter particles which do not belong to any halo. This scheme is approximate and works best if the unresolved haloes have a bias close to unity, i.e. where the bias is not a strong function of mass. This

is almost the case in the application of this method to the BASICC run, so the clustering amplitude appears slightly boosted for all the configurations studied here. However, since we only want to study the variation in the error on β_m when using a larger volume, we apply the same method described above to measure β_m in the galaxy samples planted in the BASICC run.

As shown in the fourth column of Table 4.2, we find that for all the Hα configurations here studied the error on β_m obtained when using the BASICC simulation is a factor 1-6 smaller than that found with the Millennium samples. The H-band samples, on the other hand, have errors roughly ~ 4 times smaller in the BASICC volume compared to the Millennium volume, which is what we expect if we assume that the error on β_m scales with $1/\sqrt{V}$.

It is worth remarking that the scenario in which we calculate β_m should be regarded as idealised. In reality, the error σ_z will depend on the source flux in a rather complicated way. The effect of a redshift success rate below unity might not be equivalent to removing a random fraction of galaxies as we have assumed, but instead it could be related to, for example, line mis-identifications. Our calculations should be considered as a first attempt to get an idea of the uncertainties and relative merits expected for different survey configurations.

The `Euclid` survey will cover a geometrical volume of $\sim 90[\mathrm{Gpc/h}]^3$ with an effective volume of around half of this (see next section). We expect that `Euclid` should meausre β_m with an accuracy around 4 times smaller than that estimated for the galaxy samples planted into the BASICC simulation.

4.4.4 Effective survey volume

Ongoing and future surveys aim to measure the baryonic acoustic oscillation (BAO) signal in the power spectrum of galaxies. The primary consideration for an accurate power spectrum measurement is to maximize the survey volume in order to maximize the number of independent k-modes. However, because the power spectrum is measured using a finite number of galaxies there is an associated discreteness noise. The number density of galaxies in a flux limited sample drops rapidly with increasing redshift, which means that discreteness noise also increases. When the discreteness noise becomes comparable to the power spectrum amplitude, it is difficult to measure the clustering signal. This trend is encapsulated in the expression for the fractional error on the power spectrum derived by Feldman, Kaiser &

Peacock (1994):

$$\frac{\sigma}{P} \approx \frac{2\pi}{\sqrt{Vk^2\Delta k}}\left(1+\frac{1}{\bar{n}P}\right), \tag{4.6}$$

$$\approx \frac{2\pi}{\sqrt{V_{\rm eff}(k)k^2\Delta k}}, \tag{4.7}$$

where σ is the error on the power spectrum P, V is the geometrical survey volume and $\bar{n}$ is the number density of galaxies.

Consider expanding the survey by adding shells of redshift width δz, with fixed solid angle. As we have seen from Fig. 4.11, the number density of galaxies in the samples we are considering drops steeply with increasing redshift. We therefore need to compare the discreteness or shot noise of the galaxies in the shell with the clustering signal amplitude. If the ratio $\bar{n}P > 1$, then the clustering signal can be measured above the discreteness noise level, and the volume of the shell contributes usefully to the survey volume. On the other hand if $\bar{n}P < 1$, it is hard to measure the clustering of the galaxies in this shell and it contributes nothing to the statistical power of the survey. When the limit of $\bar{n}P < 1$ is reached, the effective volume reaches a plateau and adding further redshift shells does not improve the accuracy with which the power spectrum can be measured. The amplitude of the power spectrum compared to the discreteness noise of the galaxies used to trace the density field is therefore a key consideration when assessing the effectiveness of different tracers of the large scale structure of the Universe.

GALFORM gives us all the information required to estimate the effective volume of a survey with a given selection criteria (which defines the number density of galaxies, $\bar{n}(z)$, and the effective bias as a function of redshift). For simplicity, we use the linear theory power spectrum of dark matter, which is a reasonable approximation on the wavenumber scales studied here. The galaxy power spectrum is assumed to be given by $P_{\rm g}(k,z) = b(z)^2 P_{\rm dm}(k,z)$, where $b(z)$ is the effective bias of the galaxy sample. We calculate the fraction of volume utilized in a given redshift interval following Tegmark (1997),

$$V_{\rm eff}(k) = \int_{z_{\rm min}}^{z_{\rm max}} \left[\frac{\bar{n}(z)P_{\rm g}(k,z)}{1+\bar{n}(z)P_{\rm g}(k,z)}\right]^2 \frac{{\rm d}V}{{\rm d}z}{\rm d}z, \tag{4.8}$$

where all quantities are expressed in comoving coordinates. We calculate $V_{\rm eff}/V$ for a range of possible survey configurations considering different limits in flux, $EW_{\rm obs}$, magnitude limit and redshift success rate (see Table 4.3). The redshift range is chosen to match that

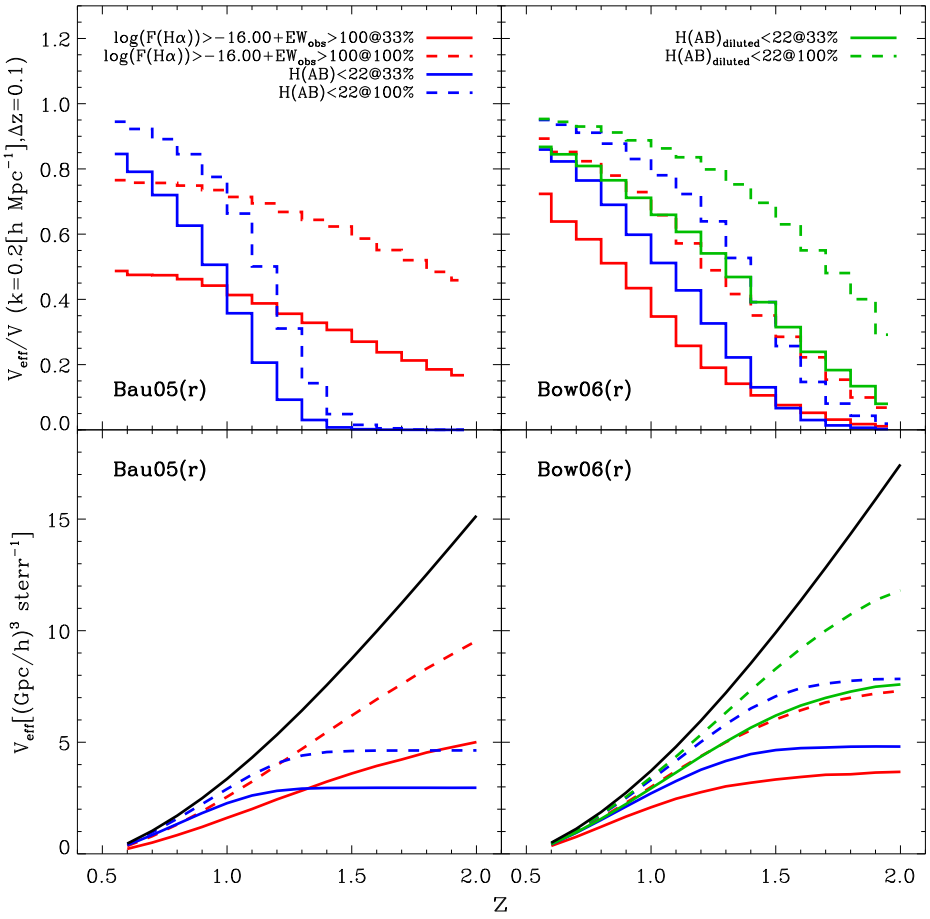

Figure 4.12: The effective volume of Hα- and H-band selected samples. The left-hand panels show results for `Bau05(r)` model and the right-hand panels show the `Bow06(r)` model; in the latter case, the effective volume for a randomly diluted sample of galaxies from the original `Bow06` model is also shown. The upper row shows the effective volume divided by the geometrical volume in redshift shells of width $\Delta z = 0.1$; the power spectrum at $k = 0.2h\mathrm{Mpc}^{-1}$ is used to compute the effective volume (see text). The lower panels show the cumulative effective volume per steradian starting from $z = 0.5$ and extending up to the redshift at which the curve is plotted. Red curves show the results for Hα selected galaxies with $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ and $EW_{\mathrm{obs}} > 100$Å. The solid red line shows the result of applying a redshift success of 33%, whereas the red dashed line assumes a 100% success rate. The blue lines show the results for an H-band magnitude selected survey with $H_{\mathrm{AB}} < 22$. As before, the solid blue line shows the results for a sampling rate of 33%, and the dashed line assumes 100% sampling. The green lines show the results using the `Bow06` model diluted (`Bow06(d)`) to match the observed number counts; as before solid and dashed show 33% and 100% success rates, respectively. The black solid curves in the bottom panels show the total comoving volume covering the redshift range shown.

expected to be set by the near-infrared instrumentation to be used in future surveys.

Fig. 4.11 shows the predictions from GALFORM which are required to compute the effective volume, for two illustrative Hα and H-band selected surveys, covering the current expected flux/magnitude limits of space missions. The bias predicted for H-band galaxies is at least $\sim 30\%$ higher than that for Hα-emitters in both panels of Fig. 4.11. This reflects the different spatial distribution of these samples apparent in Fig. 4.8, in which is it clear that Hα emitters avoid cluster-mass dark matter haloes. The middle panel of Fig. 4.11 shows the galaxy number density as a function of redshift for these illustrative surveys. For the Hα selection, the models predict very different number densities at low redshifts, as shown also in Fig 4.7. For $z > 1$ the Bow06(r) model predicts progressively more galaxies than the Bau05(r) model for the H-band selection. Overall, the number density of galaxies in the H-band sample at high redshift is much lower than that of Hα emitters. However, we remind the reader than these scaled models match the H-band counts but have a shallower redshift distribution than is suggested by the observations. The bottom panel of Fig. 4.11 shows the power spectrum times the shot noise, $\bar{n}P$, as a function of redshift. A survey which efficiently samples the available volume will have $\bar{n}P > 1$. The slow decline of the number density of Hα galaxies with redshift in the Bau05(r) model is reflected in $\bar{n}P > 1$ throughout the redshift range considered here, whereas in the Bow06(r) model, the Hα sample has a very steeply falling $\bar{n}P$ curve, with $\bar{n}P < 1$ for $z > 1.5$. The predictions of $\bar{n}P$ for the H-band are similar in both models, dropping below 1 at $z \sim 1.3 - 1.5$.

The predictions for the bias, number density and power spectrum of galaxies plotted in Fig. 4.11 are used in Eq. (4.8) to calculate the effective volume, which is shown in Fig. 4.12. The top panels show the differential $V_{\rm eff}/V$ calculated in shells of $\Delta z = 0.1$ for redshifts spanning the range $z = [0.5, 2]$. The bottom panels of Fig. 4.12 show the cumulative $V_{\rm eff}$ contained in the redshift range from $z = 0.5$ up to $z = 2$. We follow previous work and use the amplitude of the power spectrum at $k = 0.2h{\rm Mpc}^{-1}$, which roughly corresponds to the centre of the wavenumber range over which the BAO signal is measured. We show the result for the fiducial survey selections with different redshift success rates, 100% and 33%. In addition, for the H-band selected survey, we also show the results obtained with the alternative approach discussed in the previous section, in which the galaxies in the Bow06 sample are diluted by a factor of 0.63.

Table 4.3: The effective volume of Hα- and H-band selected surveys for different selection criteria. We evaluate a given survey configuration in terms of its effective volume in the redshift range $0 < z < 2$ (top) and $0.5 < z < 2$ (bottom), which is expressed as a fraction of the geometrical volume over the same redshift interval. The first column shows the galaxy selection method used, Hα for an Hα selected survey with a minimum flux limit and $EW_{\rm obs}$ cut or $H_{\rm AB}$ for an H-band magnitude limited survey. The second column shows the H-band magnitude limit chosen in a given configuration, where applicable. The third column shows the minimum Hα flux chosen, again where applicable, and the fourth column the minimum $EW_{\rm obs}$ cut applied. The fifth column shows the redshift success rate assumed. Columns 6, 7 and 8 show the fractional effective volume obtained for a given configuration in the `Bau05`, `Bow06` and the diluted version of the `Bow06` model respectively. Finally, columns 9, 10 and 11 show our estimate of the corresponding percentage error on the determination of w, the dark energy equation of state parameter, for the `Bau05`, `Bow06` and diluted `Bow06` models, respectively.

Selection	$H_{\rm AB}$ (mags)	log($F_{\rm H\alpha}$) (ergs^{-1}cm^{-2})	$EW_{\rm obs}$ (Å)	Sampling rate	$V_{\rm eff}/V$ `Bau05(r)`	$V_{\rm eff}/V$ `Bow06(r)`	$V_{\rm eff}/V$ `Bow06(d)`	Δw(%) `Bau05(r)`	Δw(%) `Bow06(r)`	Δw(%) `Bow06(d)`
$0 < z < 2$										
Hα	-	-15.40	100	0.33	0.08	0.09	-	1.2	1.1	-
Hα	-	-15.40	100	1.00	0.24	0.18	-	0.7	0.8	-
Hα	-	-15.40	0	1.00	0.24	0.18	-	0.7	0.8	-
Hα	-	-15.70	100	0.33	0.19	0.20	-	0.8	0.7	-
Hα	-	-15.70	100	1.00	0.44	0.39	-	0.5	0.5	-
Hα	-	-15.70	0	1.00	0.45	0.39	-	0.5	0.5	-
Hα	-	-16.00	100	0.33	0.34	0.41	-	0.6	0.5	-
Hα	-	-16.00	100	1.00	0.63	0.67	-	0.4	0.4	-
Hα	-	-16.00	0	1.00	0.64	0.67	-	0.4	0.4	-
H(AB)	21	-	-	0.33	0.13	0.13	0.22	1.0	0.9	0.7
H(AB)	21	-	-	1.00	0.18	0.21	0.38	0.8	0.7	0.5
H(AB)	22	-	-	0.33	0.23	0.30	0.45	0.7	0.6	0.5
H(AB)	22	-	-	1.00	0.33	0.47	0.68	0.6	0.5	0.4
H(AB)	23	-	-	0.33	0.41	0.57	0.68	0.5	0.4	0.4
H(AB)	23	-	-	1.00	0.59	0.78	0.86	0.4	0.3	0.3
$0.5 < z < 2$										
Hα	-	-15.40	100	0.33	0.06	0.06	-	1.4	1.4	-
Hα	-	-15.40	100	1.00	0.21	0.15	-	0.8	0.9	-
Hα	-	-15.40	0	1.00	0.21	0.15	-	0.8	0.9	-
Hα	-	-15.70	100	0.33	0.18	0.07	-	0.9	1.2	-
Hα	-	-15.70	100	1.00	0.43	0.17	-	0.5	0.8	-
Hα	-	-15.70	0	1.00	0.44	0.17	-	0.5	0.8	-
Hα	-	-16.00	100	0.33	0.33	0.21	-	0.6	0.7	-
Hα	-	-16.00	100	1.00	0.62	0.41	-	0.4	0.5	-
Hα	-	-16.00	0	1.00	0.63	0.41	-	0.4	0.5	-
H(AB)	21	-	-	0.33	0.09	0.10	0.19	1.2	1.1	0.8
H(AB)	21	-	-	1.00	0.14	0.18	0.35	1.0	0.8	0.6
H(AB)	22	-	-	0.33	0.19	0.27	0.43	0.8	0.6	0.5
H(AB)	22	-	-	1.00	0.30	0.44	0.67	0.6	0.5	0.4
H(AB)	23	-	-	0.33	0.38	0.55	0.67	0.6	0.4	0.4
H(AB)	23	-	-	1.00	0.57	0.77	0.86	0.5	0.4	0.3

In general, the effective volume is close to the geometrical volume at low redshifts. This is because $\bar{n}P \gg 1$ at these redshifts. In the top panels of Fig. 4.12, where the differential $V_{\rm eff}$/V is plotted in shells of $\Delta z = 0.1$, we see that shells at higher redshifts cover progressively smaller differential effective volumes. This is due to the overall decrease in the number density of galaxies beyond the peak in the redshift distribution (see Figs. 4.5, 4.7 and 4.11), which wins out over the more modest increase in the bias of the galaxies picked up with increasing redshift. The bottom panels of Fig. 4.12 show the same effect: at higher redshifts, the gain in effective volume is much smaller than the corresponding gain in the geometrical volume of the survey. We remind the reader that our calculation for the effective volume in the H-band using models with rescaled luminosities is likely to be an underestimate, as these models underpredict the observed high redshift tail of the redshift distribution. A better estimate is likely to be provided by the `Bow06(d)` model, in which the number of galaxies is adjusted by a making a random sampling, rather than by changing their luminosities. This case is shown by the green curves in Fig. 4.12.

The calculations presented in Fig. 4.12 are extended to a range of survey specifications in Table 4.3. This table shows calculations for two different redshift ranges: $0 < z < 2$ and $0.5 < z < 2$, and includes also the effect of applying different selection criteria and redshift success rates to Hα and H-band surveys. An Hα survey with a limiting flux of $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -15.4$, an equivalent width $EW_{\rm obs} > 100$Å and a sampling rate of 0.33, similar to the baseline spectroscopic solution for `Euclid`, would have a very small $V_{\rm eff}/V \sim 0.04$ for the redshift interval $z = 0.5 - 2$. In contrast, an H-band survey with $H_{\rm AB} < 22$ and a sampling rate of 0.33, an alternative spectroscopic solution for `Euclid`, has $V_{\rm eff}/V = 0.19 - 0.27$ or even up to $V_{\rm eff}/V = 0.43$ in the case of the diluted model. To reach a comparable effective volume, a Hα survey would need to reach a flux limit of at least $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -16$ (at the same equivalent width cut and redshift success rate).

The calculation of the effective volume also allows us to make an indicative estimate of the accuracy with which the dark energy equation of state parameter, w, can be measured for a given survey configuration. Angulo et al. (2008a) used large volume N-body simulations combined with the `GALFORM` model to calculate the accuracy with which the equation of state parameter w can be measured for different galaxy samples. They found a small difference ($\sim 10\%$) in the accuracy with which w can be measured for a continuum magnitude limited sample and an emission line sample with the same number density of objects. Their results

can be summarised by:

$$\Delta w(\%) = \frac{1.5\%}{\sqrt{V_{\rm eff}}}, \tag{4.9}$$

where $V_{\rm eff}$is in units of h^{-3}Gpc^3 and the constant of proportionality (in this case, 1.5) depends on which cosmological parameters are held fixed; in the present case models are considered in which the distance to the epoch of last scattering is fixed as the dark energy equation of state parameter varies. We obtain an estimate of the accuracy with which w can be measured by inserting $V_{\rm eff}$into Eq. 4.9, which is shown in Table 4.3, for the `Bau05` and `Bow06` models.

4.5 Discussion and Conclusions

In this chapter we have presented the first predictions for clustering measurements expected from future space-based surveys to be conducted with instrumentation sensitive in the near-infrared. We have used published galaxy formation models to predict the abundance and clustering of galaxies selected by either their Hα line emission or H-band continuum magnitude. The motivation for this exercise is to assess the relative performance of the spectroscopic solutions proposed for galaxy surveys in forthcoming space missions which have the primary aim of constraining the nature of dark energy. Our comparison is idealised in that we only consider the effective volume sampled by the survey strategies, and do not address issues of cost or whether or not a particular flux limit and sampling rate is achievable in practice.

The physical processes behind Hα and H-band emission are quite different. Hα emission is sensitive to the instantaneous star formation rate in a galaxy, as the line emission is driven by the number of Lyman continuum photons produced by massive young stars. Emission in the observer frame H-band typically probes the rest frame R-band for the proposed magnitude limits and is more sensitive to the stellar mass of the galaxy than to the instantaneous star formation rate.

The `GALFORM` code predicts the star formation histories of a wide population of galaxies, and so naturally predicts their star formation rates and stellar masses at the time of observation. Variation in galaxy properties is driven by the mass and formation history of the host dark matter halo. This is because the strength of a range of physical effects depend on halo properties such as the depth of the gravitational potential well or the gas cooling time. This point is most striking in our plot of the spatial distribution of Hα and H-band selected galax-

ies, Fig. 4.8. This figure shows remarkable differences in the way that these galaxies trace the underlying dark matter distribution. Hα emitters avoid the most massive dark matter haloes and trace out the filamentary structures surrounding them. The H-band emitters, on the other hand, are preferentially found in the most massive haloes. This difference in the spatial distribution of these tracers has important consequences for the redshift space distortion of clustering.

In this chapter we have studied two published galaxy formation models, those of Baugh et al. (2005) and Bower et al. (2006). The models were originally tuned to reproduce a subset of observations of the local galaxy population and also enjoy notable successes at high redshift. We presented the first comparison of the model predictions for the properties of Hα emitters, extending the work of Le Delliou et al. (2005, 2006) and Orsi et al. (2008) who looked at the nature of Lyman-alpha emitters in the models. Observations of Hα emitters are still in their infancy and the datasets are small. The model predicitions bracket the current observational estimates of the luminosity function of emitters. In addition, the `Bau05` model is in reasonable agreement with the observed distribution of equivalent widths.

The next step towards making predicitions of the effectiveness of future redshift surveys is to construct mock catalogues from the galaxy formation models (see Baugh, 2008). Using the currently available data, we used various approaches to fine tune the models to reproduce the observations as closely as possible. The main technique was to rescale the line and continuum luminosities of model galaxies; another approach was to randomly dilute or sample galaxies from the catalogue. This allowed us to better match the number of observed galaxies. The resulting mocks gave reasonable matches to the available clustering data around $z \sim 2$ for the Hα samples. Our goal in this chapter was to make faithful mock catalogues. The nature of Hα emitters in hierarchical models will be pursued in a future work.

The ability of future surveys to measure the large scale structure of the Universe can be quantified in terms of their effective volumes. The effective volume takes into account the effect of the discreteness of sources on the measurement of galaxy clustering. If the discreteness noise is comparable to the clustering signal, it becomes hard to extract any useful clustering information. Once this point is reached, although the available geometrical volume is increased by going deeper in redshift, in practice there is little point as no further statistical power is being added to the clustering measurements. The error on a power spectrum or correlation function measurement scales as the inverse square root of the effective volume.

In the case of flux-limited samples, the number density of sources falls rapidly with increasing redshift beyond the median redshift. Even though the effective bias of these galaxies tends to increase with redshift, it does not do so at a rate sufficient to offset the decline in the number density. The GALFORM model naturally predicts the abundance and clustering strength of galaxies needed to compute the effective volume of a galaxy survey.

The differences in the expected performance of Hα and H-band selected galaxies when measuring the power spectrum is related to the different nature of the galaxies selected by these two methods. Hα emitters are active star forming galaxies, which makes them have smaller bias compared to H-band selected galaxies. Their redshift distribution is also very sensitive to the details of the physics of star formation: The effect of a top-heavy IMF in bursts in the Bau05 model boosts the number density of bright emitters, making the redshift distrubtion of Hα emitters very flat and slowly decreasing towards high redshifts, in contrast to the predictions of the Bow06 model, where a sharp peak at $z \sim 0.5$ and a rapid decrease for higher redshifts is found. H-band galaxies are less sensitive to this effect, and the redshift distributions are similar in both models. This is why the balance between the power spectrum amplitude (given by the effective bias) and the number density is translated in two different effective volumes for Hα and H-band selected galaxies.

Although there are differences in detail between the model predictions, they give similar bottom lines for the effective volumes of the survey configurations of each galaxy selection. Comparing the spectroscopic solutions in Table 4.3, a slit based survey down to $H_{\rm AB} = 22$ would sample 4-10 times the effective volume which could be reached by a slitless survey to $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -15.4$, taking into account the likely redshift success rate. To match the performance of the H-band survey, an Hα survey would need to go much deeper in flux, down to $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) = -16$.

We have also looked at the accuracy with which Hα emitters and H-band selected galaxies will be able to measure the bulk motions of galaxies and hence the rate at which fluctuations are growing, another key test of gravity and the nature of dark energy. All of the samples we considered showed a small systematic difference between the measured growth rate and the theoretical expectation, at about the 1σ level. The error on the growth rate from an Hα survey with $\log(F_{H\alpha}[\mathrm{erg\ s^{-1}\ cm^{-2}}]) > -15.4$ was found to be about three times larger than that for a sample with $H_{AB} < 22$.

Chapter 5

Radiative transfer of Lyα *photons*

5.1 Introduction

In chapter 3 we described the importance and usefulness of Lyα emitters as tracers of the dark matter distribution, particularly in the high redshift Universe. This is mostly due to the prominent spectral feature at 1216Å (in the rest frame), which arises from the downward transition between the levels $n = 2$ and $n = 1$ of hydrogen atoms in the interstellar medium of galaxies. Several observational techniques have been developed to search for the Lyα line in galaxies, with narrow band surveys being the most successful and common choice (Hu et al., 1998; Kudritzki et al., 2000; Gawiser et al., 2007; Nilsson et al., 2007b; Gronwall et al., 2007; Ouchi et al., 2008). Magnification of Lyα emission by galaxy lensing has also helped to detect candidates for very high redshift Lyα emitters (e.g. Stark et al., 2007). Integral field units are also being used to construct large surveys of Lyα emitters (Blanc et al., 2007; Hill et al., 2008)

From a theoretical point of view, perhaps the main uncertainty when modelling Lyα emission is the assumption about the fraction of Lyα photons which escape from the galaxy, f_{esc}. Lyα photons are very vulnerable to even small amounts of dust, due to the many scatterings they undergo before escaping from an HI region. Furthermore, other factors also play an important role, such as the kinematics, composition, temperature and geometry of the interstellar medium. This makes the modelling of the escape of Lyα photons a very challenging task.

There has been important progress over the last few years in modelling Lyα emitters in a cosmological setting. The first consistent hierarchical galaxy formation model which includes Lyα emission is the one described in Chapter 3 (see also Le Delliou et al., 2005, 2006; Orsi

et al., 2008), which makes use of the GALFORM semianalytical model. In this model, the simple assumption of a fixed escape fraction $f_{esc} = 0.02$, regardless of any galaxy property, allowed us to predict remarkably well the abundances and clustering of Lyα emitters in a wide range of redshifts and luminosities.

Kobayashi et al. (2007, 2010) attempted to include an empirical model of the escape of Lyα emitters in a different semianalytical model, predicting the correct abundances, UV luminosities and EWs of Lyα emitters. Nagamine et al. (2006, 2008) modelled Lyα emitters in cosmological SPH simulations, introducing a tunable escape fraction and duty cycle or stochasticity parameter $C_{Ly\alpha}$ (Lyα emission is assumed to *switch off* or be completely obscured in a random fraction of galaxies determined by $C_{Ly\alpha}$). Dayal et al. 2010a and Dayal et al. 2010b combined an SPH simulation with a radiative transfer model of ionizing photons to link the escape of continuum photons to the escape of Lyα at high redshifts. Their detailed modelling of the attenuation of the Lyα flux by the IGM allowed them to predict Lyα and UV luminosity functions in agreement with observations. However, a large degeneracy was found between the fraction of neutral hydrogen and the ratio of continuum to Lyα escape fractions assumed.

A more physical approach to modelling the escape of Lyα photons requires a treatment of the radiative transfer processes that photons undergo when travelling through an HI region. The scattering and destruction of Lyα photons have been extensively studied in the past, due to its many applications in astrophysical media. Due to the great complexity of the problem, numerical methods have been developed since the sixties to study specific problems, such as the mean number of scatterings a Lyα photon will experience before leaving a medium of a given optical depth, and the emerging flux from extremely thick media (Osterbrock, 1962; Avery and House, 1968; Adams, 1972).

Nowadays, numerical methods allow us to study the line profiles and escape fractions of Lyα photons in a variety of physical configurations. The standard method is to use a Monte Carlo algorithm, in which the path of a set of photons is followed one at a time through each scattering event, until the photon either escapes or is absorbed by a dust grain. These calculations have been succesfully applied to study the properties of Lyα emitters in different scenarios: Ahn 2003 (see also Ahn 2004) developed an outflow model to predict and

characterise the Lyα line profiles coming from starbursts at high redshifts. Verhamme et al. shown in a series of papers (Verhamme et al., 2006; Schaerer and Verhamme, 2008; Verhamme et al., 2008) that photons escaping from a spherically symmetric shell with a homogeneous hydrogen number density can reproduce the observed Lyα profiles of a sample of observed Lyman-break galaxies. Also Laursen et al. studied the variety of Lyα profiles and escape fractions obtained from a sample of galaxies taken from an SPH simulation (Laursen and Sommer-Larsen, 2007; Laursen et al., 2009a,b). Zheng and Miralda-Escudé 2002 (see also Dijkstra et al. 2006; Barnes and Haehnelt 2010) applied a Lyα radiative transfer code to study the properties of Damped Lyα Absorption systems (DLAs). Hansen and Oh (2006) used a similar code to study the effect of a multiphase medium (where dust is concentrated in clouds rather than being homogeneously distributed) on the emerging Lyα flux and escape fraction. Tasitsiomi (2006) applied their Lyα radiative transfer code to an SPH simulation to study the emergent Lyα flux of a $z \sim 8$ galaxy.

Despite this variety of theoretical work, there has not been any attempt to incorporate a detailed Lyα radiative transfer model in a fully fledged galaxy formation model. That is the goal of the next two chapters of this thesis: To develop a Lyα radiative transfer code in order to combine it with the GALFORM semi-analytical model to get an insight of the physical properties that affect the escape fraction of Lyα photons and the observed properties of Lyα emitters.

In the following, we will describe the physics and numerical implementation of the radiative transfer of Lyα photon when crossing a dusty HI region. The result of coupling this calculation to the GALFORM code will be discussed in the next chapter. The Monte Carlo radiative transfer code developed here is similar to other codes developed in the past, and mostly follows previous work by Zheng and Miralda-Escudé (2002); Verhamme et al. (2006); Dijkstra et al. (2006) and Laursen et al. (2009a).

5.2 Basics of Lyα radiative transfer

It is convenient to express frequencies, ν, in terms of

$$x \equiv \frac{(\nu - \nu_0)}{\Delta\nu_D}, \tag{5.1}$$

where $\Delta\nu_D = \nu_{\rm th}\nu_0/c$, c is the speed of light, and $\nu_{\rm th}$ is the thermal velocity of the hydrogen atoms in the gas, which is given by

$$\nu_{\rm th} = \left(\frac{2k_B T}{m_p}\right)^{1/2}, \tag{5.2}$$

where k_B is the Boltzmann constant, T is the gas temperature, m_p is the proton mass and ν_0 is the central frequency of the Lyα line, $\nu_0 = 2.47 \times 10^{15}$Hz.

When a Lyα photon interacts with an hydrogren atom, the scattering cross section, in the rest frame of the atom is given by

$$\sigma_\nu = f_{12}\frac{\pi e^2}{m_e c}\frac{\Gamma/4\pi^2}{(\nu-\nu_0)^2 + (\Gamma/4\pi)^2}, \tag{5.3}$$

where $f_{12} = 0.4162$ is the Lyα oscillator frequency, and $\Gamma = A_{12} = 6.25 \times 10^8 s^{-1}$ is the Einstein coefficient for the Lyα transition ($n = 2$ to $n = 1$).

The optical depth of a Lyα photon with frequency ν is determined by convolving this cross section with the velocity distribution of the gas,

$$\tau_\nu(s) = \int_0^s \int_{-\infty}^{+\infty} n(V_z)\sigma(\nu, V_z)\, {\rm d}V_z {\rm d}l, \tag{5.4}$$

where V_z denotes the velocity component along the photon's direction. Atoms are assumed to have a Maxwell-Boltzmann velocity distribution. In Doppler units, the optical depth can be written as

$$\tau_x(s) = \sigma_H(x) n_H s = 5.868 \times 10^{-14} T_4^{-1/2} N_H \frac{H(x,a)}{\sqrt{\pi}}, \tag{5.5}$$

where n_H is the hydrogen density, N_H the corresponding hydrogen column density, T_4 the temperature in units of 10^4K and a is the Voigt parameter, defined as

$$a = \frac{\Gamma/4\pi}{\Delta\nu_D} = 4.7 \times 10^{-4} T_4^{-1/2} \tag{5.6}$$

The Hjerting function $H(x,a)$ (Hjerting, 1938) describes the Voigt scattering profile,

$$H(x,a) = \frac{a}{\pi}\int_{-\infty}^{+\infty} \frac{e^{-y^2}\, {\rm d}y}{(y-x)^2 + a^2}, \tag{5.7}$$

which is often approximated by a central resonant core and power-law “damping wings" for frequencies below/above a certain boundary frequency x_c, which typically ranges between $2.5 < x_c < 4$. As a consequence, photons with frequencies close to the line centre have a large scattering cross section compared to those with frequencies in the wings of the profile. Hence, photons will be more likely to escape a medium when they have a frequency away from the line centre.

Scattering events are considered to be *coherent* (the frequency of the photon is the same before and after the scattering event) only in the rest frame of the atom, but not in the observer's frame. Thus, the thermal motion of the atom, plus any additional bulk motion of the gas, will potentially change the frequency of the photons, giving them the chance to escape from the resonant core. We will study this in detail in the next section.

5.3 A Monte Carlo radiative transfer code

Our goal is to understand the transfer of Lyα radiation in a large variety of physical configurations, so that we can apply our results to galaxies predicted by `GALFORM`. Of particular interest are the emergent spectrum of the Lyα line and the escape fraction of Lyα photons f_{esc}.

The above quantities have been computed analytically for some specific cases. Harrington (1973) computed the emergent spectrum of Lyα photons generated at the line centre escaping from an optically thick, static, homogeneous and dust-free infinite slab. Almost 20 years later, Neufeld (1990) generalised the previous result allowing the Lyα photons to be generated with frequencies away from the line centre. Later on, Dijkstra et al. (2006) followed a similar procedure to that of Harrington and Neufeld to compute the emergent spectrum from an optically thick, static, homogeneous, dust-free sphere.

As for the escape fraction in the presence of dust, less analytical progress has been made due to the complexity of the task. Neufeld (1990) computed the escape fraction of Lyα photons from an optically thick, static, homogeneous dusty slab. Using a Monte Carlo code, Hansen and Oh (2006) computed f_{esc} in a variety of multi-phase media, and provided fitting formulae based on the Neufeld expression for f_{esc}.

Evidently, the above results are not suitable to study more general cases, which is why we have developed our own Monte Carlo numerical code to compute the escape of Lyα photons in a wide variety of possible scenarios.

Monte Carlo radiative transfer codes work on a 3D grid in which each cell will contain information about the neutral hydrogen density n_H, the temperature of the gas T, the bulk velocity $v_{\rm bulk}$, and the probability of emitting a Lyα photon, which could, for instance, depend on the distribution of sources, for instance. Once a Lyα photon is created, a random direction and frequency are given to it, and the code must follow its trajectory and compute each scattering event of the photon until it either escapes or is absorbed by a dust grain. If the photon escapes, then its final frequency is recorded. In the following, we will call ξ_n a random number in the range $0 < \xi_n < 1$, where the subscript $n = 1, 2...$ reffers to different random numbers when more than one is used in a given calculation.

5.3.1 Initial direction and frequency

The initial direction of the photon is randomly selected using the transformations

$$\theta = \cos^{-1}(2\xi_1 - 1), \tag{5.8}$$

$$\phi = 2\pi\xi_2, \tag{5.9}$$

where ξ_1 and ξ_2 are two random numbers, and θ and ϕ are in radians. The initial frequency of the photon in the rest frame of the medium (which may not be static) will be, in Doppler units, $x_i' = 0$. In the observer's rest frame, the initial frequency of the photon will be

$$x_i = x_i' + \boldsymbol{n_i} \cdot \boldsymbol{v_{\rm bulk}}/v_{th}, \tag{5.10}$$

where $\boldsymbol{v_{\rm bulk}}$ is the bulk velocity vector of the gas at the location of the emission, and $\boldsymbol{n_i}$ is the direction of emission in cartesian coordinates, given by

$$\boldsymbol{n_i} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{5.11}$$

5.3.2 Distance travelled

The location of the interaction (with either a dust grain or a hydrogen atom) is calculated as follows. The optical depth $\tau_{\rm int}$ the photon will travel is determined from the probability

distribution

$$P(\tau) = 1 - e^{-\tau}, \tag{5.12}$$

and so

$$\tau_{\rm int} = -\ln(1-\xi). \tag{5.13}$$

This optical depth corresponds to a distance travelled s given by

$$\tau(s) = \tau_x(s) + \tau_d(s), \tag{5.14}$$

where $\tau_x(s)$ and $\tau_d(s)$ are the optical depths due to hydrogen atoms and dust grains respectively. The length of the path travelled is determined by finding the distance s where $\tau(s) = \tau_{\rm int}$ by setting

$$s = \frac{\tau_{\rm int}}{n_H \sigma_x + n_d \sigma_d}, \tag{5.15}$$

where n_d and σ_d, the number density of dust grains and cross-section for interaction with dust, are described below.

The cross-section for scattering with a hydrogen atom depends on $H(x)$ (see Eq. 5.5), which looks like a Gaussian in the core and a power law in the wings:

$$H(x,a) \sim \begin{cases} e^{-x^2}, & \text{core} \\ \frac{a}{\sqrt{\pi}x^2}, & \text{wings.} \end{cases} \tag{5.16}$$

These two approximations break down in the transition domain, which is why we choose to compute the integral in Eq. (5.7) numerically and store the values of $H(x,a)$ in a table for a wide range of x and a. For frequencies larger than the one used to compute $H(x,a)$ numerically ($x_{max} = 500$), we use the power law approximation given in Eq. (5.16). This alternative both improves the performance (it is faster to compute the expression of (5.16) rather than interpolate values from a very long look-up table) and extends the validity of the code over a very wide range in x.

Sometimes the distance travelled makes the photon leave the cell from which it originated. In this case, the physical conditions can change, and, thus, this may affect the distance the photon was originally intended to travel. For instance, if the number density of hydrogen drops between two adjacent cells, then the total distance travelled by the photon will be larger than that originally computed, since the optical depth in the low density cell

translates into a larger travel distance than in the higher density cell. Temperature gradients and relative bulk motions between two adjacent cells have a similar effect on the total distance travelled. To overcome this problem, we compute how much optical depth is used after crossing each cell, until there is no more *available* optical depth to use. Numerically, when the distance computed is larger than the distance of the photon to the edge of the cell it is pointing to, then we let the photon travel to that edge, and compute the optical depth expended in this part of the motion. Then we recompute a new distance using the conditions in the new cell using Eq. (5.14), and repeat the procedure until all the original optical depth assigned to the photon according to Eq. (5.13) has been used.

The final location of the photon corresponds to the point where it interacts with either a hydrogen atom or a dust grain. To find out which type of interaction the photon experiences, we compute the probability $P_H(x)$ of being scattered by a hydrogen atom, given by

$$P_H(x) = \frac{n_H \sigma_H(x)}{n_H \sigma_H(x) + n_d \sigma_d}. \tag{5.17}$$

We generate a random number ξ and compare it to P_H. If $\xi < P_H$, then the photon interacts with the hydrogen atom, otherwise, it interacts with dust.

5.3.3 Dust scattering and absorption

Observationally, the extinction A_V is found to be proportional to the column density of hydrogen both within the Milky Way and in the Magellanic clouds (Massa and Fitzpatrick, 1986; Fitzpatrick and Massa, 2007), although with different proportionality coefficients.

Nevertheless, the cross section of dust interaction $\sigma_d(\lambda)$ can be expressed as an effective cross section per hydrogen atom. This eliminates any dependence of the cross section on the dust size distribution, grain shapes, etc, and relies merely on observed extinction curves. The optical depth of dust τ_d can be written as

$$\tau_d = \sigma_d N_d, \tag{5.18}$$

where σ_d is the cross section of dust particles (absorption plus scatterings), and N_d the column density of dust, $N_d \propto \Sigma_{\rm dust}$. On the other hand, the column density of hydrogen, $N_H \propto \Sigma_{\rm gas}$. To relate both we assume that the mass of dust can be related to the mass of the gas and the metallicity through

$$M_{\rm dust} = \epsilon M_{\rm gas} Z_{\rm gas}, \tag{5.19}$$

where $\epsilon = \delta_*/Z_*$, and δ_* is the dust-to-gas ratio at the solar metallicity $Z_* = 0.02$. We can write then

$$N_d \propto N_H Z_{\rm gas}, \tag{5.20}$$

which makes

$$\frac{\tau_d}{N_H} \propto Z_{\rm gas}. \tag{5.21}$$

The ratio τ_d/N_H can be obtained at solar metallicity from a tabulated extinction curve. If

$$\frac{\tau_d}{N_H}(Z_\odot) = E_\odot, \tag{5.22}$$

then we find

$$\tau_d = \frac{E_\odot}{Z_\odot} Z_{\rm gas} N_H. \tag{5.23}$$

Now, the optical depth of dust can be split in two separate terms, accounting for absorption and scatter by dust grains: $\tau_d = \tau_a + \tau_s$. Both are related by the albedo A, which is the probability of a photon being scattered when interacting with a dust grain. Thus, if A is known we have

$$\tau_a = (1-A)\frac{E_\odot}{Z_\odot} Z_{\rm gas} N_H. \tag{5.24}$$

Notice that τ_d and A are assumed not to depend on the photon frequency, since over the typical frequency range photons travel both quantities do not vary significantly, even if thermal motion of dust grains are considered.

When interacting with a dust grain, a Lyα photon can be either absorbed or scattered. This depends on the albedo of dust particles. At the wavelength of Lyα, the albedo is usually $A \sim 0.4$, depending on the extinction curve used. If the Lyα photon is absorbed, then it is lost forever. If not, then it will be scattered. The new direction will depend on a probability distribution for the elevation angle θ, whereas for the azimuthal angle ϕ the scattering will be symmetric. The scattering angle θ can be obtained from Henyey and Greenstein (1941) phase function

$$P_{HG}(\mu) = \frac{1}{2}\frac{1-g^2}{(1+g^2-2g\mu)^{3/2}}, \tag{5.25}$$

where $\mu = \cos\theta$ and $g = \langle\mu\rangle$ is the asymmetry parameter. If $g = 0$, Eq.(5.25) reduces to isotropic scattering. $g = 1(-1)$ implies complete forward (backward) scattering. In general g depends on the wavelength. For Lyα, $g = 0.73$.

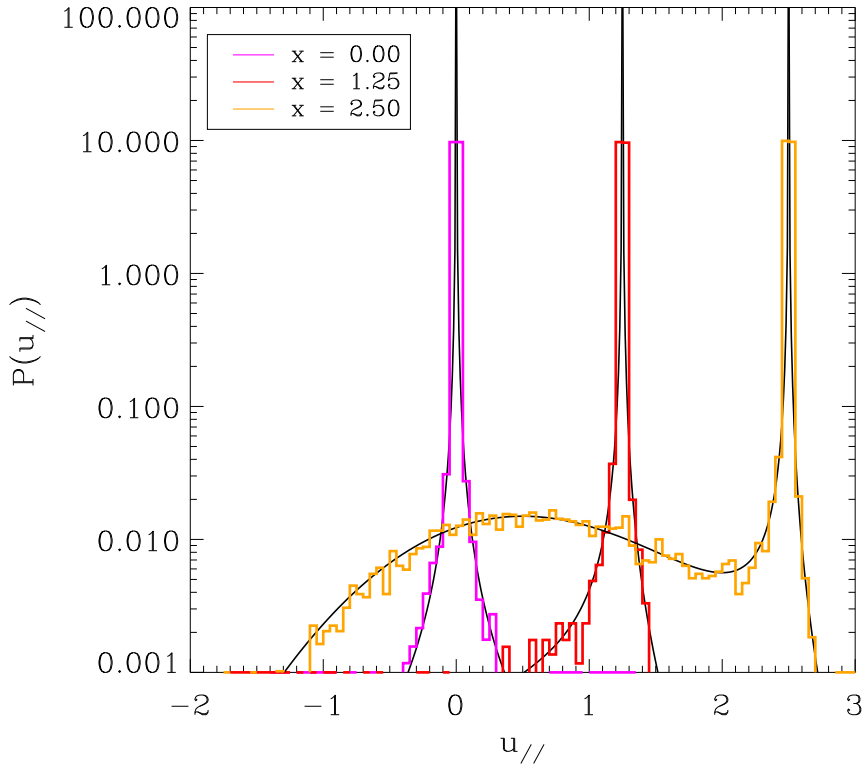

Figure 5.1: Probability distribution function of velocities of the scattering atom $u_{\parallel}$ parallel to the photon's direction for 3 values of the frequency of the incoming photon. The solid curves show the distribution from Eq. (5.27), and the histograms show the numerical results from the method described in section 5.3.4.

If the photon is interacting with dust, then we generate a random number ξ_1 to determine whether it is going to be absorbed or scattered, comparing this number to A. If the photon is absorbed, then it is lost. If it is scattered, then a new direction must be drawn.

5.3.4 Hydrogen scattering

Scattering by hydrogen atoms is more tricky. Inside an HI region, atoms move in random directions with velocities given by the Maxwell-Boltzmann distribution. Each of these atoms will *see* the same photon moving with a different frequency, due to the Doppler shift caused by their velocities. Since the cross section for scattering depends on the frequency of the photon, the probability for an atom to interact with a photon will depend on a combination of the frequency of the photon and the velocity of the atom.

The total velocity of the atom, $\boldsymbol{v}$, that scatters the Lyα photon, is given by the sum of the

bulk velocity of the gas plus the thermal velocity

$$\boldsymbol{v} = \boldsymbol{v}_{\rm bulk} + \boldsymbol{v}_{th}. \tag{5.26}$$

In the directions perpendicular to the photon's direction, $\boldsymbol{n_i}$, the two perpendicular components of the velocity of the atom $v_{\perp,1,2}$ follow a Gaussian distribution. The velocity component parallel to the photon's direction, $v_{\parallel}$, will depend on x. The normalized probability distribution for $u_{\parallel} \equiv v_{\parallel}/v_{th}$ in scattering events is found to be

$$f(u_{\parallel}) = \frac{a}{\pi H(a,x)} \frac{e^{-u_{\parallel}^2}}{(x-u_{\parallel})^2 + a^2}. \tag{5.27}$$

Eq.(5.27) is not analytically integrable, so to draw values from this probability distribution we make use of the *rejection method*. Following Zheng and Miralda-Escudé (2002), this is done as follows. We need a random number u from

$$f(u) \propto \frac{e^{-u^2}}{(x-u)^2 + a^2}, \tag{5.28}$$

so we choose as a comparison function

$$g(u) \propto \frac{1}{(x-u)^2 + a^2}. \tag{5.29}$$

A first random number u is chosen from $g(u)$, and then we keep it if a second random number ξ_1 is smaller than e^{-u^2}. Unfortunately, when $x \gg 1$, the chance of acceptance becomes very small. To increase this fraction, the comparison distribution is modified to be

$$g(u) \propto \begin{cases} [(x-u)^2 + a^2]^{-1}\,, & u \le u_0 \\ e^{-u_0^2}[(x-u)^2 + a^2]^{-1}\,, & u > u_0. \end{cases} \tag{5.30}$$

The value of u_0 is chosen as a function of x to minimize the fraction of generated values that will be discarded, and will be discussed next. The acceptance fractions now are e^{-u^2} and $e^{-u^2}/e^{-u_0^2}$ in the regions $u \le u_0$ and $u > u_0$ respectively. Once u_0 is known, we must choose which region to use. To do this a random number ξ is compared to p, defined as

$$p = \frac{\displaystyle\int_{-\infty}^{u_0} g(u)\, du}{\displaystyle\int_{-\infty}^{+\infty} g(u)\, du} \tag{5.31}$$

$$= \left(\theta_0 + \frac{\pi}{2}\right)\left[(1 - e^{-u_0^2})\theta_0 + (1 + e^{-u_0^2})\frac{\pi}{2}\right]^{-1}, \tag{5.32}$$

where

$$\theta_0 = \tan^{-1}\frac{u_0 - x}{a}. \tag{5.33}$$

u is generated from $u = a\tan\theta + x$, where θ is a random number distributed between $[-\pi/2, \theta_0]$ and $[\theta_0, \pi/2]$ for $R \leq p$ and $R > p$ respectively. Then, another random number ξ determines whether the generated value of u is accepted by comparing it with the corresponding fraction of acceptance.

Following Laursen et al. (2009a), a value of u_0 valid for a wide range of temperatures and frequencies with an acceptance-to-rejection ratio of order unity is achieved when

$$u_0 = \begin{cases} 0 & \text{for } 0 \leq x < 0.2 \\ x - 0.01a^{1/6}e^{1.2x} & \text{for } 0.2 \leq x < x_{\rm cw}(a) \\ 4.5 & \text{for } x \geq x_{\rm cw}(a). \end{cases} \tag{5.34}$$

Here $x_{\rm cw}$ defines the boundary between the core and the wings of the Voigt profile, i.e., where

$$\frac{e^{-x^2}}{\sqrt{\pi}} = \frac{a}{\pi x^2}. \tag{5.35}$$

The solution to this equation can be approximated by

$$x_{\rm cw}(a) = 1.59 - 0.60\log a - 0.03\log^2 a. \tag{5.36}$$

The two velocity components perpendicular to $\boldsymbol{n_i}$, $u_{\perp 1,2}$ are drawn from a Gaussian distribution with zero mean and standard deviation $2^{-1/2}v_{th}$. Following Dijkstra et al. (2006), these components are calculated using the Box-Muller method (see Press et al., 1992) in the following way

$$u_{\perp 1} = \sqrt{-\ln(\xi_1)}\cos(2\pi\xi_2) \tag{5.37}$$

$$u_{\perp 2} = \sqrt{-\ln(\xi_1)}\sin(2\pi\xi_2), \tag{5.38}$$

where both velocities are in units of v_{th}.

Figure 5.1 shows a comparison between the probability distribution of Eq. (5.27) and the numerical algorithm described here for 100,000 values of $u_{\parallel}$. The method reproduces the expected distribution remarkably well, even at the resonance peaks shown.

We will assume that in the frame of the atom, the frequency of the outgoing photon is the same as the incident frequency. In reality, it will differ slightly due to the *recoil effect*, which

accounts for the transfer of small amounts of momentum from the photon to the atom during the scattering process. The average fractional amount of energy transferred per scattering can be written as (Field, 1959)

$$g = \frac{h\Delta\nu_D}{2kT} \quad (5.39)$$

$$= 2.6\times10^{-4}(13\mathrm{kms}^{-1}/v_{th}), \quad (5.40)$$

where h is the Planck constant. For the applications studied in this thesis, the recoil effect has been proven to be negligible (see also the discussion by Adams, 1971).

The new direction $\boldsymbol{n_o}$ is given by a dipole distribution, with the symmetry axis defined by the incident direction $\boldsymbol{n_i}$

$$P(\theta) = \frac{3}{8}(1+\cos^2\theta), \quad (5.41)$$

where θ is the polar angle to the direction $\boldsymbol{n_i}$. The azimuthal angle of the outgoing photon is random and uniform in $0 \leq \phi < 2\pi$.

Finally, the new frequency x' of the photon is given by

$$x' = x - \boldsymbol{n_i}\cdot\boldsymbol{u} + \boldsymbol{n_o}\cdot\boldsymbol{u} \quad (5.42)$$

$$= x - u_{\parallel} + \boldsymbol{n_o}\cdot\boldsymbol{u} \quad (5.43)$$

5.3.5 The resonant scattering calculation in detail

Once the total velocity of the atom is chosen according to the above probability distributions, we first perform a Lorentz transform of the direction and frequency of the photon to the rest frame of the atom. The direction and frequency of the scattered photon are then transformed back to the laboratory frame. Below we will set out how this is done.

The photon's initial direction is

$$\hat{\boldsymbol{n}}_{\boldsymbol{i}} = \sin\theta\cos\phi\,\hat{\imath} + \sin\theta\sin\phi\,\hat{\jmath} + \cos\theta\,\hat{k}, \quad (5.44)$$

where $\hat{\imath}, \hat{\jmath}$ and $\hat{k}$ are fixed in the lab frame. The atom's velocity satisfies

$$\hat{\boldsymbol{n}}_{\boldsymbol{i}}\cdot\boldsymbol{u} = u_{\parallel} = u\cos\alpha, \quad (5.45)$$

where α is the angle between the direction of the photon and the atom. To calculate α we need to know u, which satisfies

$$u^2 = u_{\parallel}^2 + u_{\perp,1}^2 + u_{\perp,2}^2. \quad (5.46)$$

Note here that u is the velocity of the atom in units of the thermal velocity, v_{th}, Eq. (5.2). In other words,

$$u = \frac{v}{v_{th}}, \tag{5.47}$$

where v[km s^{-1}].

We must now define $\boldsymbol{u}$ in the coordinate system $\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$. At this point $\boldsymbol{u}$ could be pointing in any direction within the cone formed by an angle α from $\hat{\boldsymbol{n}}_{\boldsymbol{i}}$. To fully specify the atom's direction, we define a set of vectors in the plane perpendicular to $\hat{\boldsymbol{n}}_{\boldsymbol{i}}$:

$$\boldsymbol{u} = u_{\parallel}\hat{\boldsymbol{n}}_i + u_{\perp,1}\hat{\boldsymbol{n}}_{\perp,1} + u_{\perp,2}\hat{\boldsymbol{n}}_{\perp,2} \tag{5.48}$$

We fix their direction as

$$\hat{\boldsymbol{n}}_{\perp,1} = \sin\phi\hat{\boldsymbol{\imath}} - \cos\phi\hat{\boldsymbol{\jmath}}, \tag{5.49}$$

$$\hat{\boldsymbol{n}}_{\perp,2} = \cos\phi\cos\theta\hat{\boldsymbol{\imath}} + \sin\phi\cos\theta\hat{\boldsymbol{\jmath}} - \sin\theta\hat{\boldsymbol{k}}. \tag{5.50}$$

We can now express $\hat{\boldsymbol{u}}$ as

$$\begin{aligned}\hat{\boldsymbol{u}} &= \frac{1}{u}\Big\{\Big[u_{\parallel}\sin\theta\cos\phi + u_{\perp,1}\sin\phi + u_{\perp,2}\cos\phi\cos\theta\Big]\hat{\boldsymbol{\imath}} \\ &\quad + \Big[u_{\parallel}\sin\theta\sin\phi - u_{\perp,1}\cos\phi + u_{\perp,2}\sin\phi\cos\theta\Big]\hat{\boldsymbol{\jmath}} \\ &\quad + \Big[u_{\parallel}\cos\theta - u_{\perp,2}\sin\theta\Big]\Big\}\hat{\boldsymbol{k}} \\ &\equiv u_i\hat{\boldsymbol{\imath}} + u_j\hat{\boldsymbol{\jmath}} + u_k\hat{\boldsymbol{k}}\end{aligned} \tag{5.51}$$

Since we will perform a Lorentz transformation, we also need to know the coordinates of the vector perpendicular to the velocity of the atom, $\hat{\boldsymbol{u}}_{\perp}$ lying in the plane formed by $\hat{\boldsymbol{n}}_i$ and $\hat{\boldsymbol{u}}$, which can be defined in the following way:

$$\hat{\boldsymbol{n}}_i = \cos\alpha\hat{\boldsymbol{u}} + (1-\cos^2\alpha)^{1/2}\hat{\boldsymbol{u}}_{\perp} \tag{5.52}$$

Solving for $\hat{\boldsymbol{u}}_{\perp}$ we find

$$\begin{aligned}\hat{\boldsymbol{u}}_{\perp} &= \frac{1}{(1-\cos^2\alpha)^{1/2}}[(\sin\theta\cos\phi - \cos\alpha u_i)\hat{\boldsymbol{\imath}} \\ &\quad + (\sin\theta\sin\phi - \cos\alpha u_j)\hat{\boldsymbol{\jmath}} \\ &\quad + (\cos\theta - \cos\alpha u_k)\hat{\boldsymbol{k}}] \\ &\equiv u_{\perp,i}\hat{\boldsymbol{\imath}} + u_{\perp,j}\hat{\boldsymbol{\jmath}} + u_{\perp,k}\hat{\boldsymbol{k}}.\end{aligned} \tag{5.53}$$

Finally, to perform the Lorentz transformation we need to know the photon's velocity in the directions parallel and perpendicular to the atom's velocity. These are simply

$$n_{i,\parallel}(a) = c\cos\alpha, \tag{5.54}$$

$$n_{i,\perp}(a) = c\sin\alpha. \tag{5.55}$$

So, in the atom's frame, the velocity components of the photon are

$$n'_{i,\parallel}(a) = \frac{c\cos\alpha - v}{1+(v\cos\alpha)/c}, \tag{5.56}$$

$$n'_{i,\perp}(a) = \frac{c\sin\alpha}{\gamma(1+(v\cos\alpha)/c)}, \tag{5.57}$$

where

$$\gamma = \left[1-\left(\frac{v}{c}\right)^2\right]^{-1/2} \tag{5.58}$$

$$\approx 1, \tag{5.59}$$

since $v \ll c$. The above equations can be rewritten as

$$n'_{i,\parallel}(a) = \frac{c\cos\alpha - v_{th}u}{1+u_\parallel v_{th}/c}, \tag{5.60}$$

$$n'_{i,\perp}(a) = \frac{c\sin\alpha}{(1+u_\parallel v_{th}/c)}. \tag{5.61}$$

Now the outgoing photon's direction, $\hat{\boldsymbol{n}}'_o$, satisfies

$$\hat{\boldsymbol{n}}'_o \cdot \hat{\boldsymbol{n}}'_i = \mu, \tag{5.62}$$

$$\hat{\boldsymbol{n}}'_o = \mu\hat{\boldsymbol{n}}'_i + \sqrt{(1-\mu^2)}\hat{\boldsymbol{n}}'_{i,p}, \tag{5.63}$$

where $\mu = \cos\theta$, and θ is the polar angle off the initial photon's direction, taken from the dipolar distribution, Eq.(5.41), and $\hat{\boldsymbol{n}}'_{i,p}$ is an arbitrary unit vector perpendicular to $\hat{\boldsymbol{n}}'_i$.

Since the scattering is isotropic in the azimuthal angle, its direction must be randomly drawn. The plane formed by the directions parallel and perpendicular to the atom's velocity contains one possible choice for this vector, which we will call $\hat{\boldsymbol{\zeta}}$. In addition, we will call χ the cosine of the angle between $\hat{\boldsymbol{n}}'_{i,p}$ and $\hat{\boldsymbol{\zeta}}$. In other words, if

$$\hat{\boldsymbol{n}}'_i = n'_{i,\parallel}(a)\hat{\boldsymbol{a}} + n'_{i,\perp}(a)\hat{\boldsymbol{b}}, \tag{5.64}$$

$$\hat{\boldsymbol{\zeta}} = n'_{i,\perp}(a)\hat{\boldsymbol{a}} - n'_{i,\parallel}(a)\hat{\boldsymbol{b}}, \tag{5.65}$$

where $\{\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\}$ is a set of orthogonal, unit vectors, parallel and perpendicular to the atom's velocity (in fact $\hat{\boldsymbol{a}} \equiv \hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{b}} \equiv \hat{\boldsymbol{u}}_\perp$), then $\hat{\boldsymbol{n}}'_{i,p}$ must satisfy

$$\hat{\boldsymbol{n}}'_{i,p} \cdot \hat{\boldsymbol{\zeta}} = \chi, \tag{5.66}$$

$$\hat{\boldsymbol{n}}'_{i} \cdot \hat{\boldsymbol{n}}'_{i,p} = 0, \tag{5.67}$$

If $\hat{\boldsymbol{n}}'_{i,p} = (p_1, p_2, p_3)$, then the above equations (plus the normalization condition) give

$$n'_{i,\perp}(a)p_1 - n'_{i,\parallel}(a)p_2 = \chi, \tag{5.68}$$

$$n'_{i,\parallel}(a)p_1 = -n'_{i,\perp}(a)p_2, \tag{5.69}$$

$$p_1^2 + p_2^2 + p_3^2 = 1. \tag{5.70}$$

Solving we find

$$\hat{\boldsymbol{n}}'_{i,p} = \frac{n'_{i,\perp}(a)\chi}{c}\hat{\boldsymbol{a}} - \frac{n'_{i,\parallel}(a)\chi}{c}\hat{\boldsymbol{b}} + \sqrt{1-\chi^2}\hat{\boldsymbol{c}}. \tag{5.71}$$

where $\hat{\boldsymbol{c}} \equiv \hat{\boldsymbol{a}} \times \hat{\boldsymbol{b}}$. Now $\hat{\boldsymbol{n}}'_o$ can be written in the basis $\{\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\}$ as

$$\begin{aligned} \hat{\boldsymbol{n}}'_o &= \frac{1}{c}\left(\mu n'_{i,\parallel}(a) + \sqrt{1-\mu^2}\frac{n'_{i,\perp}(a)}{c^2}\chi\right)\hat{\boldsymbol{a}} \\ &\quad + \frac{1}{c}\left(\mu n'_{i,\perp}(a) - \sqrt{1-\mu^2}\frac{n'_{i,\parallel}(a)}{c^2}\chi\right)\hat{\boldsymbol{b}} \\ &\quad + \left(\sqrt{(1-\chi^2}\sqrt{1-\mu^2}\right)\hat{\boldsymbol{c}}, \end{aligned} \tag{5.72}$$

$$\equiv n_{o1}\hat{\boldsymbol{a}} + n_{o2}\hat{\boldsymbol{b}} + n_{o3}\hat{\boldsymbol{c}}. \tag{5.73}$$

In the atom's frame, the photon has also changed its frequency due to the Doppler shift. The frequency seen by the atom corresponds to

$$\nu' = \nu_{\rm em}\gamma\left(1 - \frac{\boldsymbol{v}\cdot\hat{\boldsymbol{n}}_i}{c}\right), \tag{5.74}$$

$$\approx \nu_{\rm em}\left(1 - \frac{v_\parallel}{c}\right), \tag{5.75}$$

where ν' is the frequency seen by the atom and $\nu_{\rm em}$ is the frequency in the laboratory frame.

Now we need to perform a Lorentz transformation back to the laboratory frame. To do this we need to know $n'_{o,\parallel}$ and $n'_{o,\perp}$. Since $\hat{\boldsymbol{a}}$ is the unit vector along the direction of the atom, and $\{\hat{\boldsymbol{b}}, \hat{\boldsymbol{c}}\}$ are perpendicular to it, we can write

$$n'_{o,\parallel} = n_{o1} \tag{5.76}$$

$$n'_{o,\perp} = \sqrt{n_{o2}^2 + n_{o3}^2}. \tag{5.77}$$

The Lorentz transformation is then written as

$$n_{o,\parallel} = \frac{cn'_{o,\parallel} + v_{th}\nu}{1 + \dfrac{v_{th}\nu n'_{o,\parallel}}{c}}, \tag{5.78}$$

$$n_{o,\perp} = \frac{cn'_{o,\perp}}{1 + \dfrac{v_{th}\nu n'_{o,\parallel}}{c}}. \tag{5.79}$$

Now that we are back in the laboratory frame, it is possible to express n_o in terms of the unit vectors parallel and perpendicular to the atom's velocity:

$$\hat{\boldsymbol{n}}_{\boldsymbol{o}} = \frac{1}{c}\left(n_{o,\parallel}\hat{\boldsymbol{u}} + n_{o,\perp}\hat{\boldsymbol{u}}_{\perp,o}\right). \tag{5.80}$$

Since now $\hat{\boldsymbol{n}}_0$ lies on a different plane, $\hat{\boldsymbol{u}}_{\perp,o} \neq \hat{\boldsymbol{u}}_{\perp}$. However, $\hat{\boldsymbol{u}}_{\perp,o}$ is easy to read:

$$\hat{\boldsymbol{v}}_{\perp,o} = \frac{1}{(n_{o2}^2 + n_{o3}^2)^{1/2}}[n_{o2}\hat{\boldsymbol{b}} + n_{o3}\hat{\boldsymbol{c}}]. \tag{5.81}$$

So the only remaining task is to express $\hat{\boldsymbol{c}}$ in the basis $\hat{\imath}, \hat{\jmath}, \hat{k}$. Since $\hat{\boldsymbol{c}} = \hat{\boldsymbol{a}} \times \hat{\boldsymbol{b}}$ we get

$$\hat{\boldsymbol{c}} = (u_j u_{\perp,k} - u_k u_{\perp,j})\hat{\boldsymbol{\imath}} - (u_i u_{\perp,k} - u_k u_{\perp,i})\hat{\boldsymbol{\jmath}} + (u_i u_{\perp,j} - u_j u_{\perp,i})\hat{\boldsymbol{k}}. \tag{5.82}$$

Now it is possible to write

$$\begin{aligned} \hat{\boldsymbol{u}}_{\perp,o} &= \frac{1}{(n_{o2}^2 + n_{o3}^2)^{1/2}}\Big\{[n_{o2}v_{\perp,i} + n_{o3}(u_{\perp,k}u_j - u_k u_{\perp,j})]\hat{\boldsymbol{\imath}} \\ &\quad + [n_{o2}u_{\perp,j} - n_{o3}(u_i u_{\perp,k} - u_k u_{\perp,i})]\hat{\boldsymbol{\jmath}} \\ &\quad + [n_{o2}u_{\perp,k} + n_{o3}(u_i u_{\perp,j} - u_j u_{\perp,i})]\hat{\boldsymbol{k}}\Big\} \\ &\equiv u_{\perp,oi}\hat{\boldsymbol{\imath}} + u_{\perp,oj}\hat{\boldsymbol{\jmath}} + u_{\perp,ok}\hat{\boldsymbol{k}} \end{aligned} \tag{5.83}$$

So finally the outgoing direction of the photon can be written as

$$\hat{\boldsymbol{n}}_o = \frac{1}{c}[(n_{o,\parallel}u_i + n_{o,\perp}u_{\perp,oi})\hat{\boldsymbol{\imath}} + (n_{o,\parallel}u_j + n_{o,\perp}u_{\perp,oj})\hat{\boldsymbol{\jmath}} + (n_{o,\parallel}u_k + n_{o,\perp}u_{\perp,ok})\hat{\boldsymbol{k}}] \tag{5.84}$$

This gives the final direction of the photon, which replaces $\hat{\boldsymbol{n}}_i$ in Eq. (5.44). The frequency shift, is easy to calculate:

$$\nu_f = \nu'\left(1 + \frac{\nu \cdot \hat{\boldsymbol{n}}_{\boldsymbol{o}}}{c}\right), \tag{5.85}$$

$$= \nu_{\rm em}\left(1 - \frac{v_\parallel}{c}\right)\left(1 + \frac{\nu \cdot \hat{\boldsymbol{n}}_{\boldsymbol{o}}}{c}\right) \tag{5.86}$$

Now, for convenience, we express ν in terms of x, defined in Eq.(5.1):

$$x_f = \frac{\nu_{em}}{\Delta\nu_D}\left(1-\frac{v_\parallel}{c}\right)\left(1+\frac{\nu\cdot\hat{\boldsymbol{n}}_{\boldsymbol{o}}}{c}\right)-\nu_0, \tag{5.87}$$

$$= x_i+\frac{\nu_{em}}{\Delta\nu_D}\left(\frac{\nu\cdot\hat{\boldsymbol{n}}_{\boldsymbol{o}}}{c}-\frac{v_\parallel}{c}\right), \tag{5.88}$$

$$= x_i+\frac{\nu_{em}}{\nu_0}\left(\frac{\nu\cdot\hat{\boldsymbol{n}}_{\boldsymbol{o}}}{\nu_{th}}-\frac{v_\parallel}{\nu_{th}}\right), \tag{5.89}$$

$$= x_i+\frac{\nu_{em}}{\nu_0}(u\cdot\hat{\boldsymbol{n}}_{\boldsymbol{0}}-u_\parallel) \tag{5.90}$$

Usually $\nu_{em}\approx\nu_0$, in which case the expression reduces to what we already had in Eq.(5.43).

Finally, the dot product $\boldsymbol{u}\cdot\hat{\boldsymbol{n}}_{\boldsymbol{o}}$ is equal to

$$\boldsymbol{u}\cdot\hat{\boldsymbol{n}}_o=\frac{u}{c}[u_i(n_{o,\parallel}u_i+n_{o,\perp}u_{\perp,oi})+u_j(n_{o,\parallel}u_j+n_{o,\perp}u_{\perp,oj})+u_k(n_{o,\parallel}u_k+n_{o,\perp}u_{\perp,ok})] \tag{5.91}$$

5.3.6 Accelerating the code

The algorithm described above will follow the scattering events of a photon until it escapes (or is absorbed), and then the process starts again with a new photon travelling on a different path, and so on until we are satisfied with the number of photons generated. In practise, for the runs shown in this thesis the number of photons generated varies between a few thousand up to several hundred thousand, depending on the accuracy of the result we wish to achieve.

For the typical HI regions studied here, the number of scatterings that photons will undergo before escaping could be as high as several tens or hundreds of millions. If we want to model several thousand photons, then the total number of calculations grows enormously and the task could become computationally infeasible. However, most of the scattering events will occur when the photon is at the line centre, or very close to it, where the cross section for scattering peaks. Eq.(5.16) shows that the cross section for scatterings reduces as $\sim e^{-x^2}$ near the line centre, and as $\sim x^{-2}$ away from it. Whenever the photon falls near the centre it will experience so many scatterings that the actual distance travelled between each scattering event will be negligible, since in this case it will be most likely scattered by an atom with a velocity close to zero (see Fig. 5.1). Hence, the frequency after such scattering will remain in the resonant core. This motivates the possibility for accelerating the code performance by skipping those *inconsequential* scattering events.

Following Dijkstra et al. (2006), a critical frequency x_{crit} defines a transition from the resonant core to the wing. Whenever a photon is in the core (with $x < |x_{crit}|$) we can push it to the wings by allowing the photon to be scattered only by a rapidly moving atom. We do this by modifying the distribution of perpendicular velocities by a *truncated* Gaussian, i.e. a distribution which is a Gaussian for $u > x_{crit}$ but which is zero otherwise. The modified perpendicular velocities are then drawn from

$$u_{\perp 1} = \sqrt{x_{crit}^2 - \ln(\xi_1)}\cos(2\pi\xi_2) \quad (5.92)$$

$$u_{\perp 2} = \sqrt{x_{crit}^2 - \ln(\xi_1)}\sin(2\pi\xi_2). \quad (5.93)$$

When doing this, we allow the photon to redshift or blueshift away from the line centre, thus reducing the cross section for scattering and increasing the path length. For the configurations studied here, we found that a value of $x_{crit} = 3$ provides a good balance between accuracy and efficiency of the code, reducing the execution time by a factor 100 or more with respect to the non-accelerated case.

5.3.7 Output of the code

We allow the code to generate output in two modes:

- A *detailed* output file for each photon, in which its position and frequency is stored for every interaction followed by the code, until the photon either escapes from the HI region or is absorbed by a dust grain. Other properties are also stored, such as the velocity of the atom (if relevant) at the moment of scattering, and the running time until each interaction event. This output mode is obviously inefficient when studying a large set of photons, so it is used mainly for debugging purposes.

- A *short* output file, which stores the frequency and position of all photons at the moment they either escape or are absorbed by dust, along with other properties such as the time it took the code to compute each escape/destruction of photons. This is the preferred mode of output to compute quantities which require a large number of photons, such as the emergent spectrum or the escape fraction of photons.

5.4 Validation of the code

The flexibility of this numerical approach allows us to reproduce configurations for which analytical solutions are available. Hence, these analytical solutions are ideal to test the performance of the code. In the following we will describe the tests we have performed on our code, which prove its accuracy. Each comparison with an analytical solution is meant to test a different aspect of the code, which is why a proper validation requires to test the code against several analytical solutions.

5.4.1 The redistribution function

Hummer (1962) (and later on Lee, 1974) computed the redistribution in frequency of radiation scattered from moving atoms. Both authors studied several cases including coherent and non coherent scattering with isotropic or dipolar angular distributions and absorption profiles with zero natural line widths or Voigt profiles. In our case, we are interested in the redistribution function arising from coherent scattering in the rest frame of the atom, with a dipolar angular distribution and including radiation damping. The expression for the redistribution function is

$$R_{II-B}(x,x') = \frac{3\pi^{-3/2}}{8} a \int_{|\overline{x}-\underline{x}|/2}^{\infty} e^{-u^2} \int_{\overline{x}-u}^{\underline{x}+u} \left[3 - \left(\frac{x-t}{u}\right)^2 - \left(\frac{x'-t}{u}\right)^2 + 3\left(\frac{x-t}{u}\right)^2 \left(\frac{x'-t}{u}\right)^2 \right] \frac{\mathrm{d}t\ \mathrm{d}u}{t^2+a^2}, \quad (5.94)$$

where x and x' are the frequency of the photon before and after the scattering and $\overline{x}$ and $\underline{x}$ are the maximum and minimum between $|x|$ and $|x'|$ respectively. This expression cannot be computed analytically (although the integral over t is simple, the resulting expression is too complicated to be shown here), so it must be computed numerically. It is also possible to compute the redistribution function using our Monte Carlo code described above. Figure 5.2 shows the resulting redistribution function for 3 different initial frequencies using $\sim 10^5$ photons, and the remarkably good agreement between the Monte Carlo code and the analytical expression of Eq.(5.94).

Photons at the line centre are more likely to remain in the line centre after a scattering event, since the redistribution function is very narrow and peaks at $x_f = 0$. For a photon with an incoming frequency of $x_i = 5$, there is a non-zero chance to get closer to the line centre.

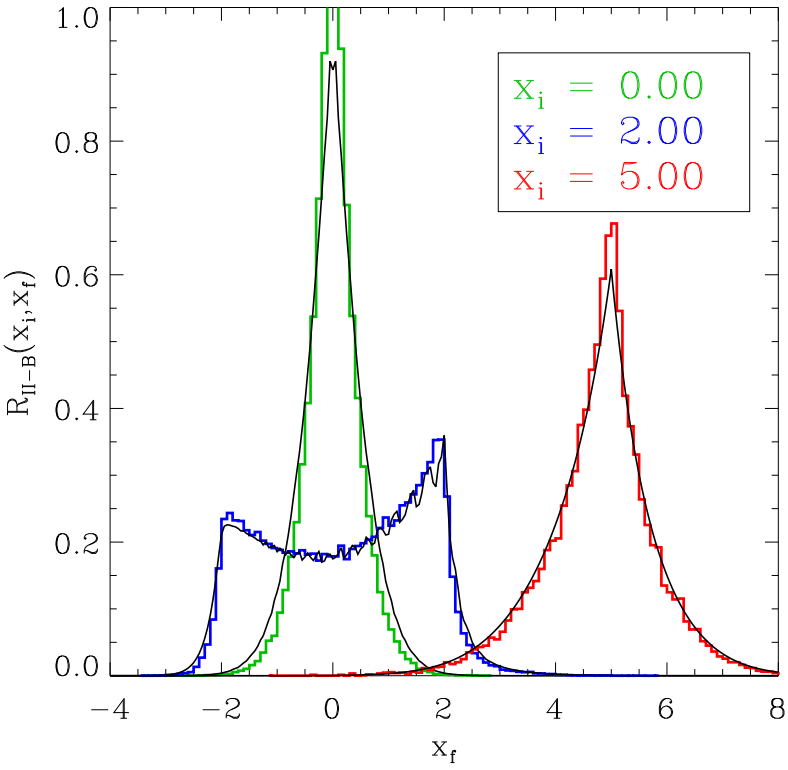

Figure 5.2: The redistribution function of Lyα photons scattered by hydrogen atoms for different initial frequencies. The histograms show the resulting frequency distribution from the Monte Carlo code, whereas the solid curves show a numerical integration of Eq. (5.94).

If this happens, subsequent scattering will increase the probability of approaching the line centre even further, until the photon will eventually reach the line centre. Once there, it is very difficult to change its frequency, which is why most of the scatterings tend to happen in the line centre.

5.4.2 Lyα spectrum from a static slab

The emergent spectrum from an optically thick, homogeneous static slab with photons generated at the line centre was first calculated by Harrington (1973), and the result was generalised by Neufeld (1990), allowing the generated photons to have any frequency. Harrington's solution for the emergent Lyα spectrum for a slab with thickness characterised by its

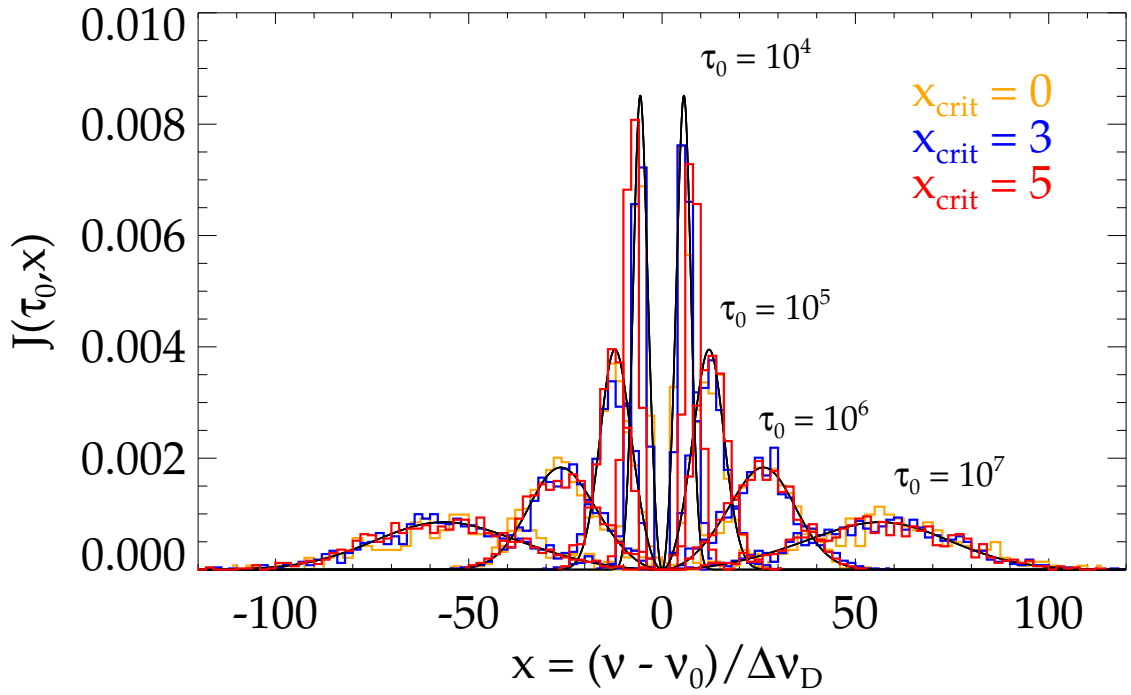

Figure 5.3: Lyα spectrum emerging from a homogeneous static slab at $T = 10$[K], for optical depths at the line centre of $\tau_0 = 10^4, 10^5, 10^6$ and 10^7, as shown in the plot. The profiles are symmetric around $x = 0$. The more optically thick the medium, the farther from the line centre the resulting peaks of each profile are found. The solid lines show Harrington (1973) analytical solution, and the orange, blue and red histograms show the results from the Monte Carlo code for a choice of $x_{crit} = 0, 3$ and 5 respectively.

optical depth at the line centre τ_0, is

$$J(\pm\tau_0, x) = \frac{\sqrt{6}}{24}\frac{x^2}{\sqrt{\pi}a\tau_0}\frac{1}{\cosh\left[\sqrt{\pi^3/54}x^3/(a\tau_0)\right]}. \quad (5.95)$$

The above expression is valid when $a\tau_0 \geqslant 10^3/\sqrt{\pi}$, or for $\tau_0 \geqslant 1.2\times10^6$ when $T = 10^4$K.

Figure 5.3 shows the emergent spectrum from a simulated homogeneous slab. The temperature of the medium was chosen to be $T = 10$K, since in this regime the analytical expression is accurate for optical depths down to $\tau_0 \sim 10^4$, which is faster to compute with the code.

The typical Lyα flux profile is double peaked, and is symmetrical with respect to the line centre. The centre of the peaks is displaced away from $x = 0$ by a value determined by τ_0. The higher the optical depth, the farther away from the line centre and the wider the profile will be. Figure 5.3 compares the analytic solution of Harrington (1973) with the ouput from the basic code (orange histogram), and two accelerated versions of it (blue and red histograms). Overall, it is clear that the non-accelerated version of the code reproduces the analytical formula over the range of optical depths shown here. When $x_{crit} = 3$ (the blue histogram in Fig. 5.3), the output is virtually indistinguishable from the non-accelerated version, but the running time has been decreased by a factor ~ 200. If we increase x_{crit} to 5 then the agreement with the analytical result breaks down at $\tau_0 = 10^4$, shifting the peaks slightly further away from the centre than the correct result. This occurs because some photons were expected to escape with frequencies of $x \leq 5$ at this optical depth, but since $x_{crit} = 5$ those photons were pushed into the wings of the absorption profile. The net effect is a shift in the resulting spectrum.

Figure 5.3 confirms that the choice of $x_{crit} = 3$ does not compromise the accuracy of the results, although larger values do.

5.4.3 Mean number of scatterings

Harrington (1973) also computed the mean number of scatterings expected before a Lyα photon escapes from an optically thick medium for the homogeneous slab. The result, already studied by Avery and House (1968) using a Monte Carlo technique, is

$$\langle N_{scat}\rangle = 1.612\tau_0. \quad (5.96)$$

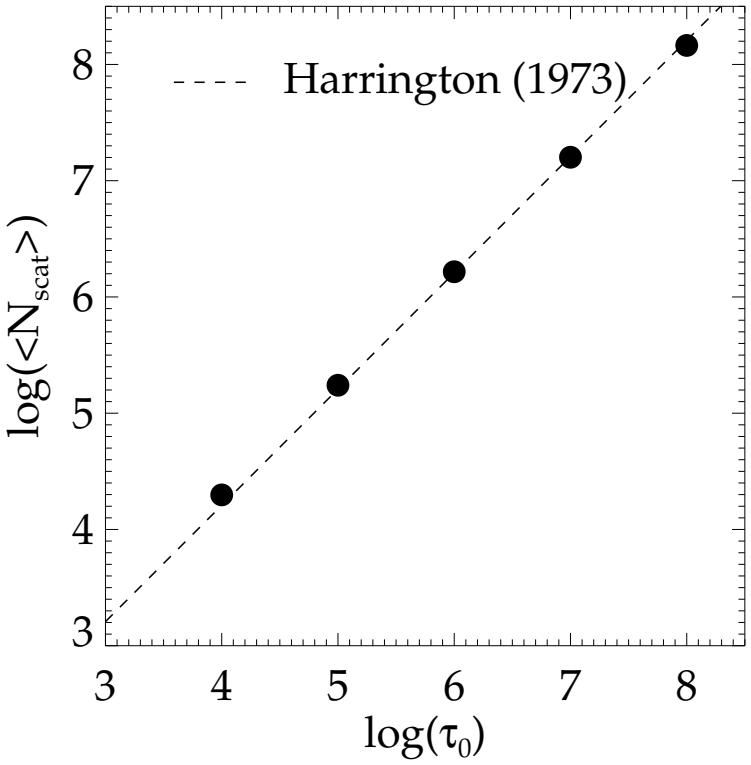

Figure 5.4: Mean number of scatterings as a function of the optical depth in the line centre of the medium. The circles show the results from the Monte Carlo code for configurations with different τ_0. The dashed line shows the analytical solution of Harrington (1973).

Fig. 5.4 shows a comparison between the mean number of scatterings computed using our code against the analytical prediction of Harrington (1973). The agreement is remarkably good.

5.4.4 Lyα spectrum from a static sphere

Following closely the methodology of Harrington (1973) and Neufeld (1990), Dijkstra et al. (2006) computed the emergent spectrum from a static sphere. Their result is

$$J(x,\tau_0) = \frac{\sqrt{\pi}}{\sqrt{24}a\tau_0}\left[\frac{x^2}{1+\cosh\left[\sqrt{2\pi^3/27}(|x^3|/a\tau_0)\right]}\right], \tag{5.97}$$

which looks very similar to the expression for the emergent spectrum from an homogeneous slab, Eq. (5.95). Fig. 5.5 shows a comparison between the analytic prediction and the output from the code at different optical depths. Again, there is a very good agreement between the two. The optical depths shown in Fig.5.5 were chosen to be different from those in Fig.

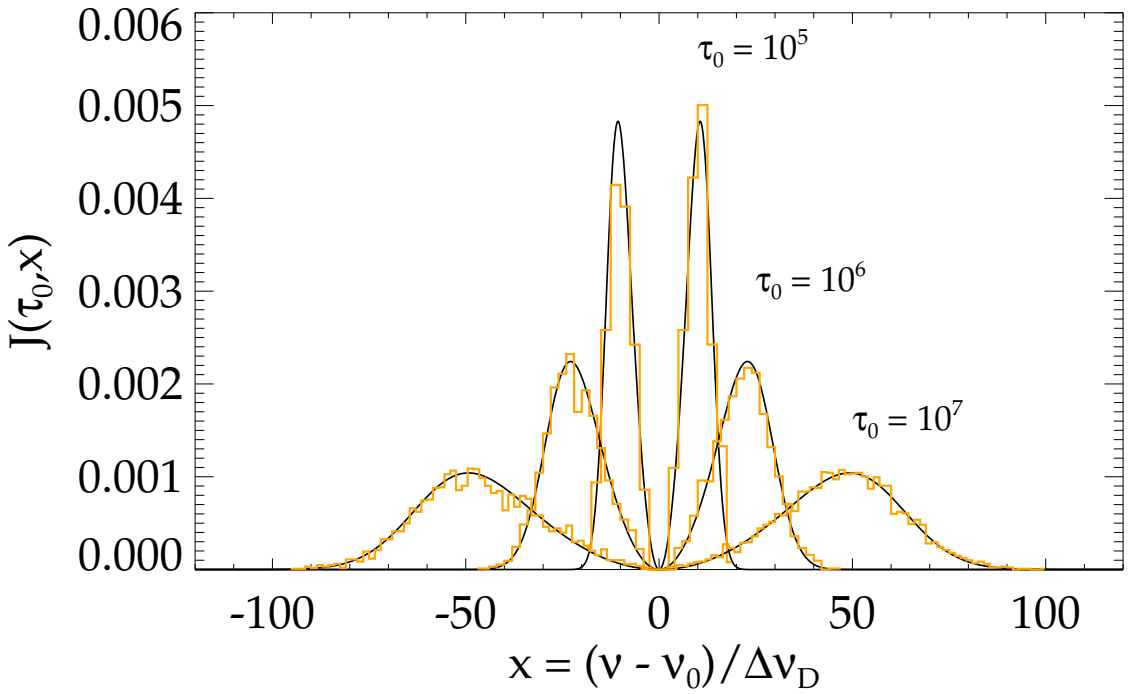

Figure 5.5: Lyα spectrum emerging from a homogeneous static sphere at $T = 10$K, for optical depths at the line centre of $\tau_0 = 10^5, 10^6$ and 10^7. The profiles are symmetric around $x = 0$. The thicker the medium, the farther from the line centre the resulting peaks of each profile are found. The solid lines show the analytical solution of Dijkstra et al. (2006) (Eq. 5.97), and the histograms show the results from the Monte Carlo code.

5.3 to show that the code is following closely the expected emergent spectrum for a range of optical depths spanning several orders of magnitude.

5.4.5 f_{esc} from a static dusty slab

Neufeld (1990) computed an analytical expression for the escape fraction of photons emitted from an homogeneous, dusty slab. The solution, valid for very high optical depths ($a\tau_0 > 10^3$), and in the limit $(a\tau_0)^{1/3} \gg \tau_a$, where τ_a is the optical depth of absorption by dust (Eq. 5.24), is

$$f_{esc} = \frac{1}{\cosh\left[\zeta'\sqrt{(a\tau_0)^{1/3}\tau_a}\right]}, \tag{5.98}$$

where $\zeta' \equiv \sqrt{3}/\zeta\pi^{5/12}$, and $\zeta \approx 0.525$ is a fitting parameter.

Fig. 5.6 shows a comparison between the escape fraction obtained from a series of simula-

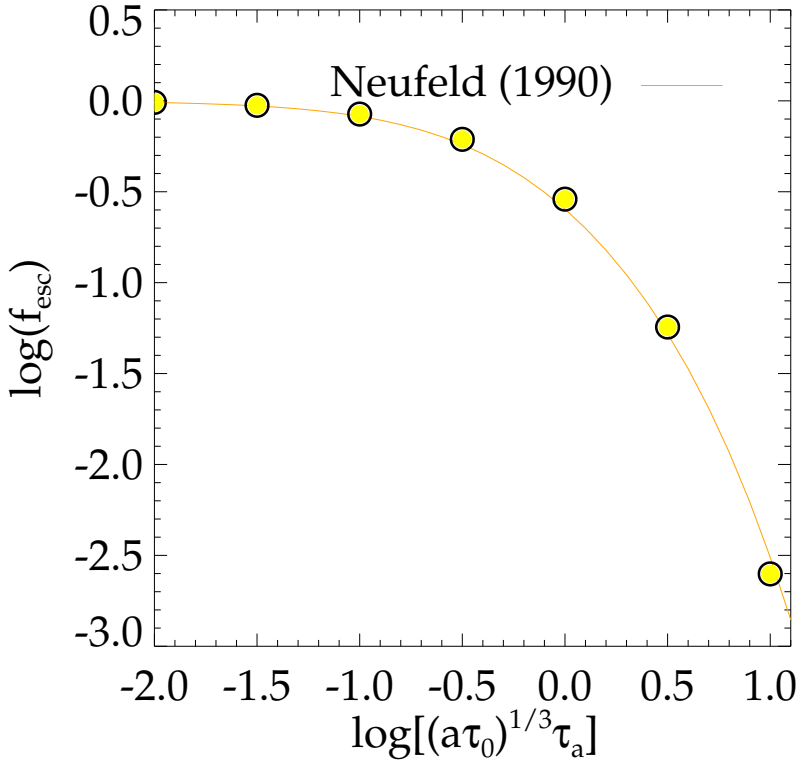

Figure 5.6: The escape fraction of Lyα photons from an homogeneous dusty slab. The optical depth of hydrogen scatterings at the line centre τ_0 is held constant at $\tau_0 = 10^6$, and different values of the optical depth of absorption τ_a are chosen. Circles show the output from the code, and the solid orange curve shows the analytical prediction of Neufeld (1990), Eq. (5.98)

tions, keeping τ_0 fixed and varying τ_a, with the analytical solution of Eq. (5.98). The escape fraction, as expected, decreases rapidly for increasing τ_a, which, for a fixed τ_0, translates into having a higher concentration of dust in the slab. It is worth noting that the escape fraction depends not only on the amount of dust, but also in the temperature of the gas (through the a parameter), the column density of hydrogen (which defines the number of scatterings in the medium, Eq. 5.96) and the amount of dust through τ_a. In a more general case, as we will see next, the mean number of scatterings is also regulated by the bulk motions of the gas, so in practice the escape fraction will depend on the bulk velocity of the gas as well.

5.4.6 Comparison with a similar code: Lyα spectrum from an expanding sphere

Another important parameter yet not taken into account is the bulk velocity of the medium. As will be shown in the next chapter, the velocity of the medium plays a crucial role shaping

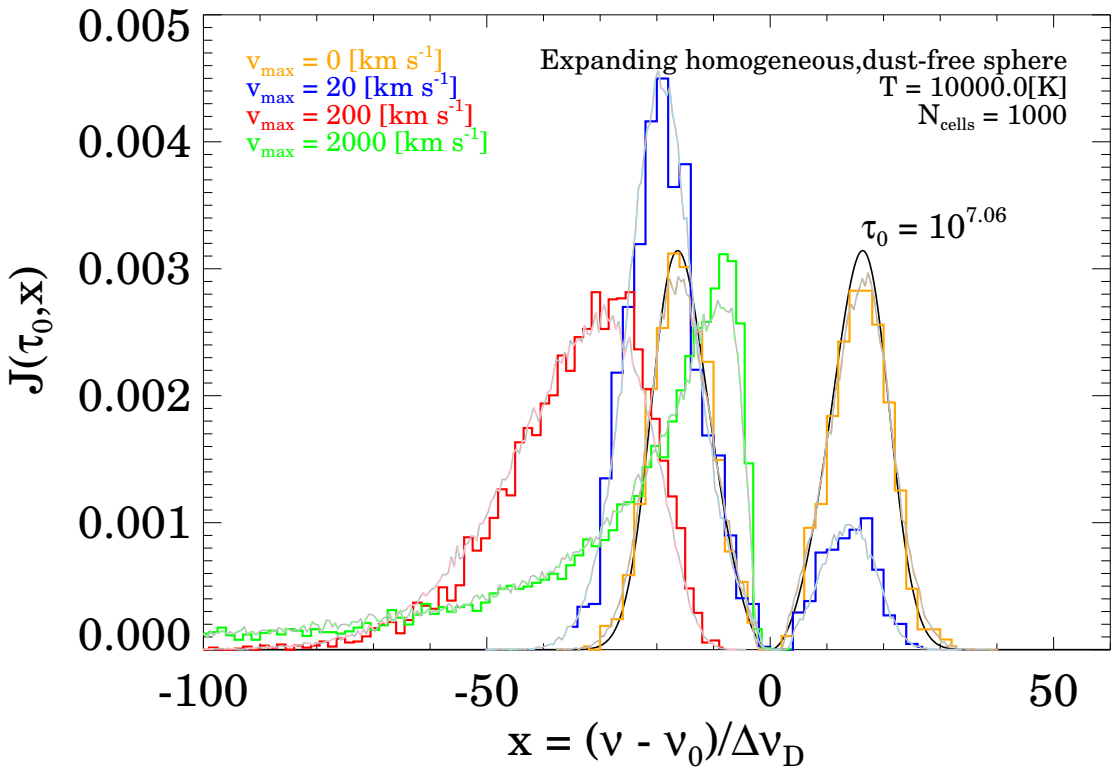

Figure 5.7: The emergent Lyα spectrum from a linearly expanding sphere with velocity zero at the centre and velocity at the edge $v_{max} = 0, 20, 200$ and 2000km/s shown in orange, blue, red and green respectively. The optical depth at the line centre is kept fixed at $\tau_0 = 10^{7.06}$. The analytical solution of Dijkstra et al. (2006) for the static case is shown in black. The coloured histograms show the output from the code. The coloured solid curves show the results obtained with the Laursen et al. (2009a) code (their Fig. 8).

the profile of the emergent spectrum and the escape fraction as well.

To study the effect of bulk motions in the gas, we model the case of an expanding homogeneous sphere, with a velocity at a distance $\boldsymbol{r}$ from the centre given by

$$v_{bulk} = H\boldsymbol{r}, \tag{5.99}$$

$$H = \frac{v_{max}}{R}, \tag{5.100}$$

where v_{max} is the velocity of the sphere at its edge, and R is the radius of the sphere.

There is no analytical solution for this configuration (except when $T = 0$, see Loeb and

Rybicki (1999)), so we decided to compare our results to those found by a similar Monte Carlo code. Fig. 5.7 shows a comparison between our code and the results obtained with the `MoCaLaTA` Monte Carlo code (Laursen et al., 2009a), kindly provided by Peter Laursen. The agreement between the two codes is encouraging. Moreover, the figure helps us understand the effect of bulk motions of the gas in the emergent spectrum. First, when $v_{max} = 0$ we obviously recover the static solution, Eq. (5.97). When $v_{max} = 20$km/s, the velocity of the medium causes photons to have a higher probability of being scattered by atoms with velocities dominated by the velocity of the medium. These atoms *see* the photons as being redshifted, and hence the peak of the spectrum is shifted slightly towards the red part of the spectrum, although still a fraction of photons appear to escape blueshifted.

When $v_{max} = 200$km/s, the blue peak is completely erased, and the peak is shifted even further to the red side. For very high velocities, such as $v_{max} = 2000$km/s, the velocity gradient makes the medium optically thin, and the average number of scatterings decreases drastically, and consequentially the photons have less chance of being redshifted far into the wings, thus shifting the peak back to the centre, but still with no photons in the blue side of the spectrum.

All the tests described in this section were used to validate our Monte Carlo code, but they also served as an example of the performance and range of possibilities open to study with our code. In the next chapter we will develop a model which will be suitable to be coupled with the output from `GALFORM`, so that we can assign an escape fraction to each galaxy and study statistical properties of Lyα emitters.

Chapter 6

Modelling the Lyα *emission of galaxies in a hierarchical Universe*

6.1 Introduction

In Chapter 3 we introduced Lyα emitters as a cosmological tool, particularly as a probe of the high redshift Universe. Galaxies detected by their Lyα emission have been used to test galaxy formation models (Le Delliou et al., 2005, 2006; Kobayashi et al., 2007; Nagamine et al., 2008; Dayal et al., 2010a) , study the spectrophotometric properties of galaxies at high redshifts (Gawiser et al., 2007; Gronwall et al., 2007; Nilsson et al., 2009), trace the large scale structure (Shimasaku et al., 2006; Gawiser et al., 2007; Kovač et al., 2007; Orsi et al., 2008; Francke, 2009; Ouchi et al., 2010) and to study possible constraints on the epoch of re-ionization of the Universe (Kashikawa et al., 2006; Dayal et al., 2010b; Ouchi et al., 2010).

Despite the many applications of the study of Lyα emitters, there has been only slow progress towards an understanding of the physical mechanisms driving the escape of Lyα radiation from the galaxy until it reaches the observer. It is evident that in order to interpret correctly the properties of Lyα emitters we must be able to understand the observed Lyα emission of these galaxies, i.e. both the fraction of the intrinsic Lyα luminosity which originates in the galaxy that we are able to observe and the shape of the Lyα line profile.

Recent observational studies have been able to infer the escape fraction f_{esc} of Lyα photons (Atek et al., 2008, 2009; Östlin et al., 2009; Kornei et al., 2010; Hayes et al., 2010a). This is generally done either by comparing the derived star formation rates from Lyα lumi-

nosities to those from ultraviolet continuum, or by comparing the observed line ratio between Lyα and other hydrogen recombination lines, such as Hα or Hβ. The first method relies heavily on the assumptions about the stellar evolution model used and the unattenuation of the ultraviolet continuum. The second method is more direct, since the intrinsic line ratios are a function of the Lyman continuum luminosity. The departure from case B recombination of the ratio of the Lyα intensity to one or more hydrogen recombination lines is then attributed to the escape fraction of Lyα differing from unity.

These measurements have revealed that the escape fraction of Lyα emitters can be anything from 10^{-3} to 1. The observational data mentioned above also suggest a correlation between the value of the escape fraction and the dust extinction, measured by $E(B-V)$, although the large scatter found in this relation suggests there is a variety of physical parameters determining the value of f_{esc}.

The first observations of local Lyα emitters suggested a strong correlation between the metallicity of galaxies and the Lyα luminosity (Meier and Terlevich, 1981; Hartmann et al., 1984, 1988), leading to the conclusion that dust, traced by the metallicity, is the most important factor driving the visibility of the Lyα line. However, a more complete analysis showed there was only a weak correlation between the two, and suggested instead the relevance of the neutral gas distribution and kinematics (e.g., Giavalisco et al., 1996). Further analysis of metal lines in local starbursts revealed evidence of the presence of outflows driving the escape of Lyα photons. The asymmetric, P-Cygni Lyα line profiles observed are consistent with Lyα photons escaping from an expanding shell of neutral gas (Thuan and Izotov, 1997; Kunth et al., 1998; Mas-Hesse et al., 2003), establishing outflows as the main mechanism driving the escape of Lyα photons from galaxies.

Observational data at higher redshifts also suggest that the Lyα line profiles observed in galaxies resemble those expected when photons escape through a galactic outflow (Shapley et al., 2003; Kashikawa et al., 2006; Kornei et al., 2010; Hu et al., 2010). Radiative transfer models have been able to reproduce successfully a sample of observed profiles assuming that photons escape from an expanding shell (Ahn, 2003, 2004; Verhamme et al., 2006; Schaerer, 2007; Verhamme et al., 2008). This method has proved useful at accounting for the variety

of Lyα line shapes, both in emission and absorption, giving rise to symmetric, single peaked profiles to asymmetric, P-Cygni, double peaked profiles and combinations of both.

It has also been suggested that the escape of Lyα photons could be boosted relative to the UV continuum when photons travel through a multi-phase ISM, i.e. a medium where dust is concentrated in clouds rather than being uniformly distributed (Neufeld, 1991; Hansen and Oh, 2006). In such a configuration, the probability of photons to interact with a dust grain is significantly reduced, since only the fraction of photons penetrating into a dust cloud are likely to be absorbed, whereas the rest, even when reaching the edge of a dust cloud, have the chance of being scattered off the dust cloud. In addition, a multi-phase ISM could also exhibit asymmetric Lyα frequency profiles (Hansen and Oh, 2006).

In this chapter we attempt to study the first case, i.e. when photons escape through an outflow of material. We leave the study of the impact of a multi-phase ISM on the visibility of the Lyα line for a future work. Previous studies of the escape of Lyα photons from an outflow focused on a sample of Lyman-break galaxies and attempted to fit the observed Lyα profiles (Schaerer, 2007; Verhamme et al., 2008). Instead, here we use the semi-analytical model GALFORM to study the properties of the Lyα emission of galaxies in a cosmological context. This represents a significant improvement in the modelling of Lyα emitters over the model presented in Chapter 3 (see also Le Delliou et al., 2005, 2006; Orsi et al., 2008), where a constant escape fraction of 0.02 was assumed to fit the observed luminosity function of Lyα emitters at $z = 3.0$.

In order to compute the escape of Lyα photons, we use the Monte Carlo radiative transfer code described in chapter 5 to model the physical conditions of an outflow. Having described the semi analytical model in chapter 2 and the Monte Carlo Lyα code in Chapter 5, we now focus directly on how we perform the coupling between both codes.

6.2 Model description

In our model, the physical properties of the medium used to compute the escape of Lyα photons depends on several predicted properties of galaxies from GALFORM. Throughout this

chapter, we use the Baugh et al. (2005) version of the GALFORM model. In this thesis we focus on two outflow models for the HI region surrounding the source of Lyα photons. Both models are similar, although they differ in their geometry and the way the properties of galaxies from GALFORM are used.

In the following we describe the details of both outflow models. We assume the temperature of the medium to be constant at $T = 10000$ K. For simplicity, the source of Lyα photons is at rest in the frame of the galaxy, in the centre, and emits photons only at the line centre. To ensure suitable time performance of the following calculations, we make use of the accelerated version of the Monte Carlo radiative transfer code with $x_{\rm crit} = 3$, as explained in Chapter 5.

6.2.1 Expanding Shell

Previous radiative transfer studies of Lyα line profiles (see, e.g. Ahn, 2003, 2004; Verhamme et al., 2006, and references therein) adopted an expanding super shell in the same way as we define it here. This model, hereafter named the **Shell** model, consists of a homogeneous, expanding, isothermal spherical shell. The **Shell** is described by an inner and outer radius $R_{\rm inn}$ and R_{out}, which satisfy $R_{\rm inn} = f_{th} R_{\rm out}$. The exact value of f_{th} is found to be irrelevant for the escape of Lyα photons, provided that $f_{th} \gtrsim 0.9$. Here we use $f_{th} = 0.9$. In addition, the medium is assumed to be expanding radially with constant velocity $\boldsymbol{V_{\rm exp}}$. The column density of the *Shell* is given by

$$N_H(r) = \frac{M_{\rm HI}}{4\pi m_H R_{\rm out}^2}, \tag{6.1}$$

where m_H is the mass of the hydrogen atom, $m_H = 1.672 \times 10^{-27}$g. The mass in the outflow is taken to be a fraction of the cold gas mass of the galaxy component from which the Lyα emission comes from. In GALFORM, the Lyα luminosity comes from the disk (in quiescent galaxies) or the bulge (in starbursts), but some galaxies also have contributions from both galactic components. Therefore, to compute the mass of hydrogen in the shell, $M_{\rm HI}$, we compute a luminosity-weighted hydrogen mass $\langle M_{\rm gas} \rangle$ between the disk and bulge. In other words,

$$M_{\rm HI} = f_M \langle M_{\rm gas,H} \rangle, \tag{6.2}$$

where f_M is the fraction of hydrogen gas that flows in the shell, and is a free parameter of the **Shell** model. Also,

$$\langle M_{\rm gas,H}\rangle = \frac{f_H}{L_{\rm Ly\alpha,tot}}(M_{\rm disk}L_{\rm Ly\alpha,disk} + M_{\rm bulge}L_{\rm Ly\alpha,bulge}), \tag{6.3}$$

where $f_H = 0.76$ is the fraction of gas in the form of hydrogen, $M_{\rm disk}$ and $M_{\rm bulge}$ are the mass of cold gas in the disk and bulge respectively, $L_{\rm Ly\alpha,disk}$ and $L_{\rm Ly\alpha,bulge}$ are the Lyα luminosities from the disk and bulge, and $L_{\rm Ly\alpha,tot} = L_{\rm Ly\alpha,disk} + L_{\rm Ly\alpha,bulge}$. Likewise, the outer radius of the shell $R_{\rm out}$ and its expansion velocity $V_{\rm exp}$ are calculated in an analogous way, i.e.

$$R_{\rm inn} = f_R\langle R_{1/2}\rangle, \tag{6.4}$$

$$V_{\rm exp} = f_V\langle V_{\rm circ}\rangle, \tag{6.5}$$

$$\langle R_{1/2}\rangle = \frac{R_{\rm disk}L_{\rm Ly\alpha,disk} + R_{\rm bulge}L_{\rm Ly\alpha,bulge}}{L_{\rm Ly\alpha,tot}}, \tag{6.6}$$

$$\langle V_{\rm circ}\rangle = \frac{V_{\rm disk}L_{\rm Ly\alpha,disk} + V_{\rm bulge}L_{\rm Ly\alpha,bulge}}{L_{\rm Ly\alpha,tot}}, \tag{6.7}$$

where $R_{\rm disk}$, $V_{\rm disk}$, $R_{\rm bulge}$ and $v_{\rm bulge}$ are the half-mass radius and circular velocity of the disk and bulge, respectively, and f_R, f_V are two extra free parameters of the model.

The metallicity of the outflow is taken to be $\langle Z_{\rm gas}\rangle$, i.e. the metallicity of the cold gas weighted by the cold gas mass of the disk and bulge. This value is used to compute the dust content in the outflow, which is assumed to be proportional to the gas mass content and metallicity, i.e.

$$M_{\rm dust} = \frac{\delta_*}{Z_\odot}M_{\rm gas}Z_{\rm gas}, \tag{6.8}$$

where the dust-to-gas ratio is taken to be $\delta_* = 1/110$ at the solar metallicity $Z_\odot = 0.02$. The optical depth of dust at the wavelength of Lyα can be defined as

$$\tau_d = \frac{E_\odot}{Z_\odot}N_H Z_{\rm gas}, \tag{6.9}$$

where $E_\odot$ is the ratio τ_d/N_H for solar metallicity at the wavelength of Lyα (1216 Å). Throughout this work we use the extinction curve and albedo from Silva et al. (1998), which are fit to the observed extinction and albedo in the local, Galactic ISM. For a dust albedo A, the optical depth of absorption is simply

$$\tau_a = (1-A)\tau_d. \tag{6.10}$$

At the wavelength of Lyα, $A = 0.39$.

6.2.2 Galactic Wind

Supernovae heat and accelerate the ISM through shocks and hence generate outflows from galaxies (see, e.g. Marlowe et al., 1995; Lehnert and Heckman, 1996; Strickland, 2002). Here we develop an outflow model, hereafter the **Wind** model, which relates the predicted mass ejection rate from galaxies due to supernovae to the density of the outflow. In `GALFORM`, this mass ejection rate is given by

$$\dot{M}_{ej} = \left[\beta_{reh}(V_c) + \beta_{sw}(V_c)\right] \psi, \tag{6.11}$$

where

$$\beta_{reh} = \left(\frac{V_c}{V_{hot}}\right)^{-\alpha_{hot}}, \tag{6.12}$$

$$\beta_{sw} = f_{sw}\min[1, (V_c/V_{sw})^{-2}], \tag{6.13}$$

$$\tag{6.14}$$

The parameters β_{reh} and β_{sw} define the two different modes of supernova feedback (the *reheating* and *superwind*), as described in Chapter 2. We assume this outflow consists of an isothermal, spherical flow expanding at constant velocity $V_{\rm out}$, with inner radius $R_{\rm inn}$. In a stationary spherical wind, the mass ejection rate is related to the velocity and density at any point of the wind via the equation of mass continuity, i.e.

$$\dot{M}_{\rm ej} = 4\pi r^2 V_{\rm exp} \rho(r). \tag{6.15}$$

Hence, the number density profile $n_{HI}(r)$ in the **Wind** model varies according to

$$n_{HI}(r) = \begin{cases} 0 & \mathrm{r} < \mathrm{R_{inn}} \\ \dfrac{\dot{M}_{\rm ej} f_{HI}}{4\pi m_H r^2 V_{\rm exp}} & \mathrm{r} \geq \mathrm{R_{inn}} \end{cases} \tag{6.16}$$

The column density N_H of the outflow is, then

$$N_H = \frac{\dot{M}_{\rm ej} f_{HI}}{4\pi m_H R_{\rm inn} V_{\rm exp}}, \tag{6.17}$$

where the inner radius of the wind, $R_{\rm inn}$, is computed in an analogous way to $R_{\rm out}$ in the **Shell** model, i.e.

$$R_{\rm inn} = f_R \langle R_{1/2} \rangle, \tag{6.18}$$

where f_R is a free parameter, and $\langle R_{1/2} \rangle$ is defined in Eq. (6.6). The expansion velocity and metallicity in the wind are computed using the same equations as in the **Shell** model.

In addition, for computational reasons, the radiative transfer code requires us to define an outer radius $R_{\rm out}$ for the outflow. However, since the number density of atoms decreases as $\sim r^{-2}$, we expect that at some point the optical depth will become so small that the photons will be able to escape regardless of the exact extent of the outflow. We found that an outer radius $R_{\rm out} = 20 R_{\rm inn}$ is large enough to achieve converged results, i.e. the escape fraction of Lyα photons does not vary if we increase the value of $R_{\rm out}$ further.

6.3 The effect of the UV background

For the **Wind** model, our radiative transfer code also allows the option of computing the resulting ionization of the medium by photons in the intergalactic UV background. The outflow as described above is completely neutral, but photoionization from the UV background could have the effect of reducing the abundance of neutral hydrogen modifying the density profile of the neutral gas.

It is generally believed that the extragalactic UV background is dominated by radiation from quasars and massive young stars from active star forming galaxies (Haardt and Madau, 1996, 2001; Meiksin, 2009). The mean intensity of the UV background at the observed frequency ν_0 and redshift z_0 is defined as

$$J_0(\nu_0, z_0) = \frac{1}{4\pi} \int_{z_0}^{\infty} {\rm d}z \frac{{\rm d}l}{{\rm d}z} \frac{(1+z_0)^3}{(1+z)^3} \epsilon(\nu, z) e^{-\tau_{eff}(\nu_0, z_0, z)}, \tag{6.19}$$

where z is the redshift of emission, $\nu = \nu_0(1+z)/(1+z_0)$, ${\rm d}l/{\rm d}z$ is the line element in a Friedmann cosmology, ϵ is the proper space-averaged volume emissivity and τ_{eff} is an effective optical depth due to absorption by the IGM. There is no explicit form of equation (6.19) since it must be computed iteratively by solving the cosmological radiative transfer equation (Peebles, 1993). For the analysis shown in this thesis we use the tabulated values of $J_0(\nu, z)$

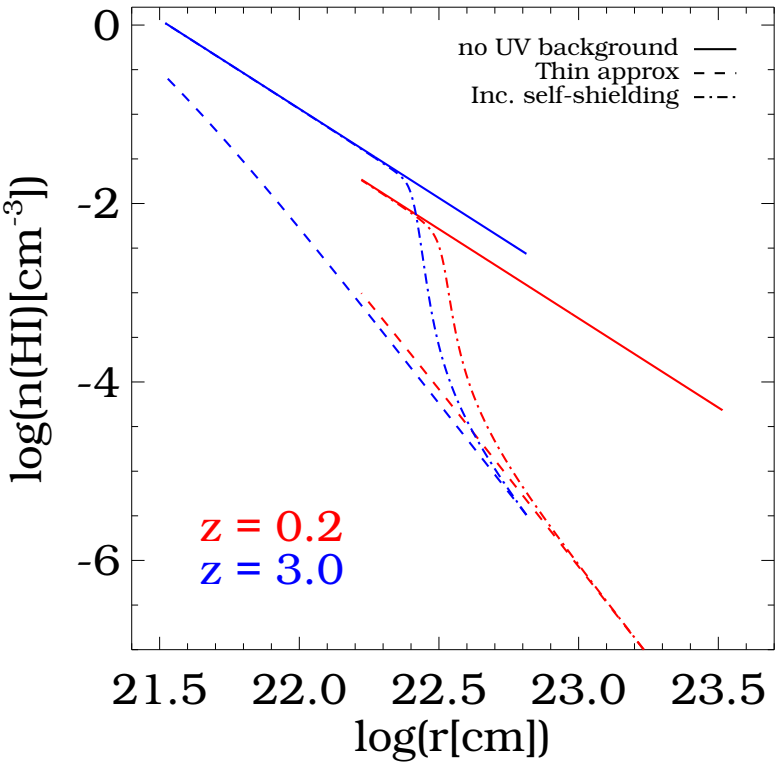

Figure 6.1: The effect of the UV background on the number density profiles for two randomly selected galaxies at $z = 0.2$ (red curves) and $z = 3.0$ (blue curves). The solid curves show the initial, unattenuated profiles. The dashed lines show the resulting profiles when including the UV background but without including self-shielding. The dotted-dashed lines show the effect of the UV background when self-shielding is included.

by Haardt and Madau (2001). Notice, however, that more recent calculations of the UV background flux (Bolton and Haehnelt, 2007; Meiksin, 2009) show that the calculation of Haardt and Madau (2001) may be underestimated for $z > 5$.

The fraction of ionized hydrogen $x \equiv n_{HII}/n_H$ varies according to a balance between radiative and collisional ionizations and recombinations inside the cloud:

$$\alpha_A n_e x = (\Gamma_H + \beta_H n_e)(1 - x), \tag{6.20}$$

where $\alpha_A = 4.18 \times 10^{-13}[\mathrm{cm^3 s^{-1}}]$ is the case A recombination coefficient at $T = 10^4 K$ (Osterbrock, 1989), and the photoionization rate $\Gamma_H(z)$ from the UV background is given by

$$\Gamma_H(z) = \int_{\nu_0}^{\infty} \frac{4\pi J_0(\nu, z)}{h\nu} \sigma_\nu(H) \mathrm{d}\nu, \tag{6.21}$$

and the collisional ionization rate $\beta_H(T)$ at $T = 10^4 K$, is (Cen, 1992)

$$\beta_H = 6.22 \times 10^{-16}[\mathrm{cm^3 s^{-1}}], \tag{6.22}$$

As the UV flux penetrates the outflow, it will be attenuated by the outer layers of neutral hydrogen (the fraction of atoms that remained neutral). The UV flux reaching an inner layer of the HI region is attenuated by this *self-shielding* process according to

$$J(\nu) = J_0(\nu) e^{-\tau(\nu)}, \tag{6.23}$$

where $J_0(\nu)$ is the original, un-shielded UV flux, and the optical depth $\tau(\nu)$ when UV photons travel a distance d inside the HI region (coming from outside) is given by

$$\tau(\nu) = \sigma_\nu(H) \int_{R_{\mathrm{out}}}^{d} n_H(r) \mathrm{d}r. \tag{6.24}$$

The photoionization rate is computed from the outer radius inwards. For each shell inside the outflow, Γ_H is computed taking into account the attenuation given by equations (6.23) and (6.24), making the photoionization rate smaller as photons penetrate inside the HI region.

Figure 6.1 shows two typical number density profiles and the effect of including the photoionization by the UV background. The optically *thin approximation* shown is obtained when using $J_0(\nu)$ in Eq. (6.21), which translates into a constant attenuation. The result of the full

calculation, including self-shielding, is very similar to the optically thin approximation for the outer layers of the outflow, since at large radii the number density of neutral hydrogen is very small. However, at some point there is a break and the outer layers start attenuating significantly the penetrating UV flux, thus bringing the number density profile back to its original form given by Eq. (6.16).

Figure 6.1 also shows that the effect of the UV background is different at these two redshifts: At $z = 0.2$, galaxies have in general larger radii compared to $z = 3$, so the inner radius of the outflow is significantly larger too. This makes the number density in the medium lower than at $z = 3$, thus making the outflow more permeable to the UV background. In detail, the effect of the UV background at modifying the number density profile of neutral hydrogen will depend on the strength of the UV flux itself, which peaks at $z \sim 1-2$ and has a sharp decline with decreasing redshifts and a slower decline at high redshifts (Haardt and Madau, 1996, 2001). The number density of the medium, on the other hand, will depend on the physical parameters of the outflow itself, i.e. the mass ejection rate, radius and velocity for the **Wind** model.

To assess the effect of the UV background on the Lyα emission properties of galaxies, we select a subsample of galaxies from `GALFORM` spanning the whole luminosity range at different redshifts. Then, we run the Monte Carlo code with and without the modification of the neutral hydrogen density profile due to the UV background, and we compare the escape fractions obtained. Figure 6.2 shows this comparison. We see that, on average, the escape fractions obtained when including the ionization of the UV background are higher. The difference is particularly noticeable at $z = 0.2$ and $z = 3.0$, due to a combination of the strength of the UV background at $z \sim 3$ (the photoionization rate $\Gamma_{HI}(z)$ resembles to some extent the shape of the global star formation density evolution, since quasars and hot stars are the main sources of UV radiation Haardt and Madau 1996, 2001), and the low densities of the outflows since their inner radius at $z = 0.2$ are usually larger by about an order of magnitude to the inner radii at $z > 3$, making, in general, the number densities lower. At $z = 5.7$ and $z = 6.6$ the UV flux is not strong enough to modify the neutral density profile of the outflows significantly, hence the escape fractions are only slightly larger when including the UV background.

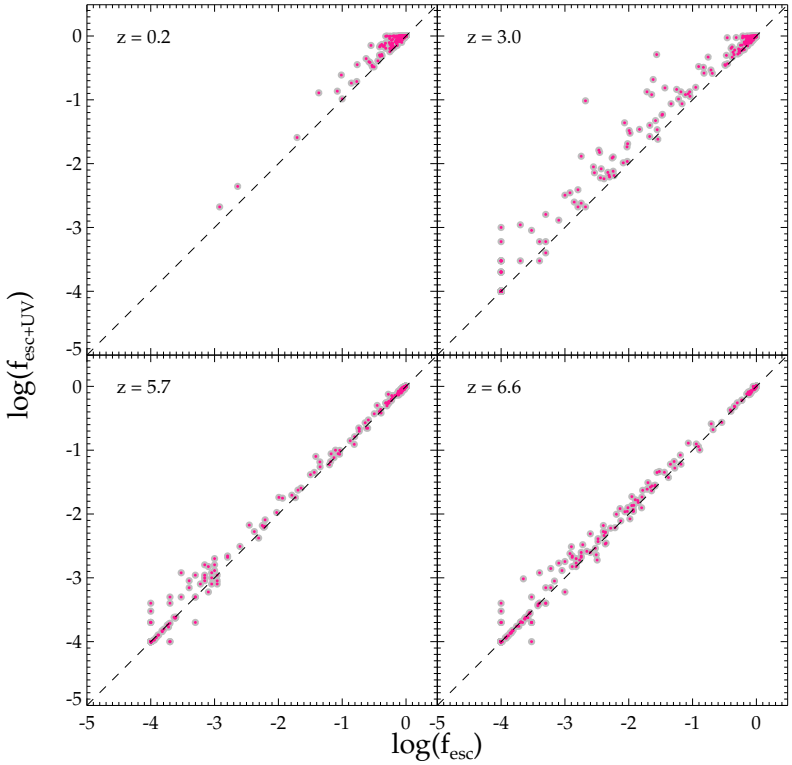

Figure 6.2: The effect of the UV background on the escape fraction for the **Wind** model. The y-axis shows the resulting f_{esc} when including the UV background, whereas the x-axis shows the result with the UV background switched off. The four panels show the escape fraction of a sub sample of galaxies taken from GALFORM at redshifts $z = 0.2$, $z = 3.0$, $z = 5.7$ and $z = 6.6$

Therefore, we include the ionization of the extragalactic UV background in the outflows of the **Wind** model following the method described above. Although the effect of the UV background can be noticeable in the escape fraction for some galaxies, as shown in figure 6.2, when studying statistical properties of Lyα emitters in the **Wind** model such as the Lyα luminosity function we notice that the results are very similar to what we obtain without including the ionization of the outflow by the UV background. This means that the Lyα escape is determined mainly by the intrinsic properties of the outflow and the extragalactic UV background plays a secondary role.

On the other hand, we do not attempt to include the effect of the UV background on the **Shell** model, since, in this case, the number density inside the outflow depends on the physical thickness of the shell, which in turn depends on f_{th}. As discussed in the previous section, this parameter is considered to have an arbitrary value provided that $f_{th} \gtrsim 0.9$. If we include the UV background in the **Shell** model, the Lyα properties would depend on the value of f_{th} assumed, which is an unnecessary complication to the model.

6.4 Comparison between the two outflow models

In summary, each outflow model requires the following parameters, provided by `GALFORM`, to compute the escape of Lyα photons for each galaxy: (i) The luminosity weighted half-mass radius $\langle R_{\mathrm{Ly}\alpha} \rangle$, (ii) The luminosity-weighted circular velocity $\langle V_{\mathrm{Ly}\alpha} \rangle$, and (iii) the mass-weighted metallicity of the gas Z_{gas}. In addition, the **Shell** model requires the mass of cold gas of the galaxy M_{gas}, whereas the **Wind** model requires the mass ejection rate due to supernovae $\dot{M}_{\mathrm{ej}}$.

Figure 6.3 shows the evolution of the above mentioned parameters in the redshift range $0.2 < z < 6.6$, as a function of the intrinsic Lyα luminosity $L_{Ly\alpha,0}$. It is worth noticing that these parameter values are extracted directly from `GALFORM`, so they do not depend on the details of the outflow model.

Figure 6.3a shows the relation of the cold gas mass with the intrinsic Lyα luminosity. As expected, overall the cold gas mass increases with $L_{Ly\alpha,0}$, since the latter is directly proportional to the star formation rate of galaxies. At low $L_{Ly\alpha,0}$ quiescent galaxies dominate the

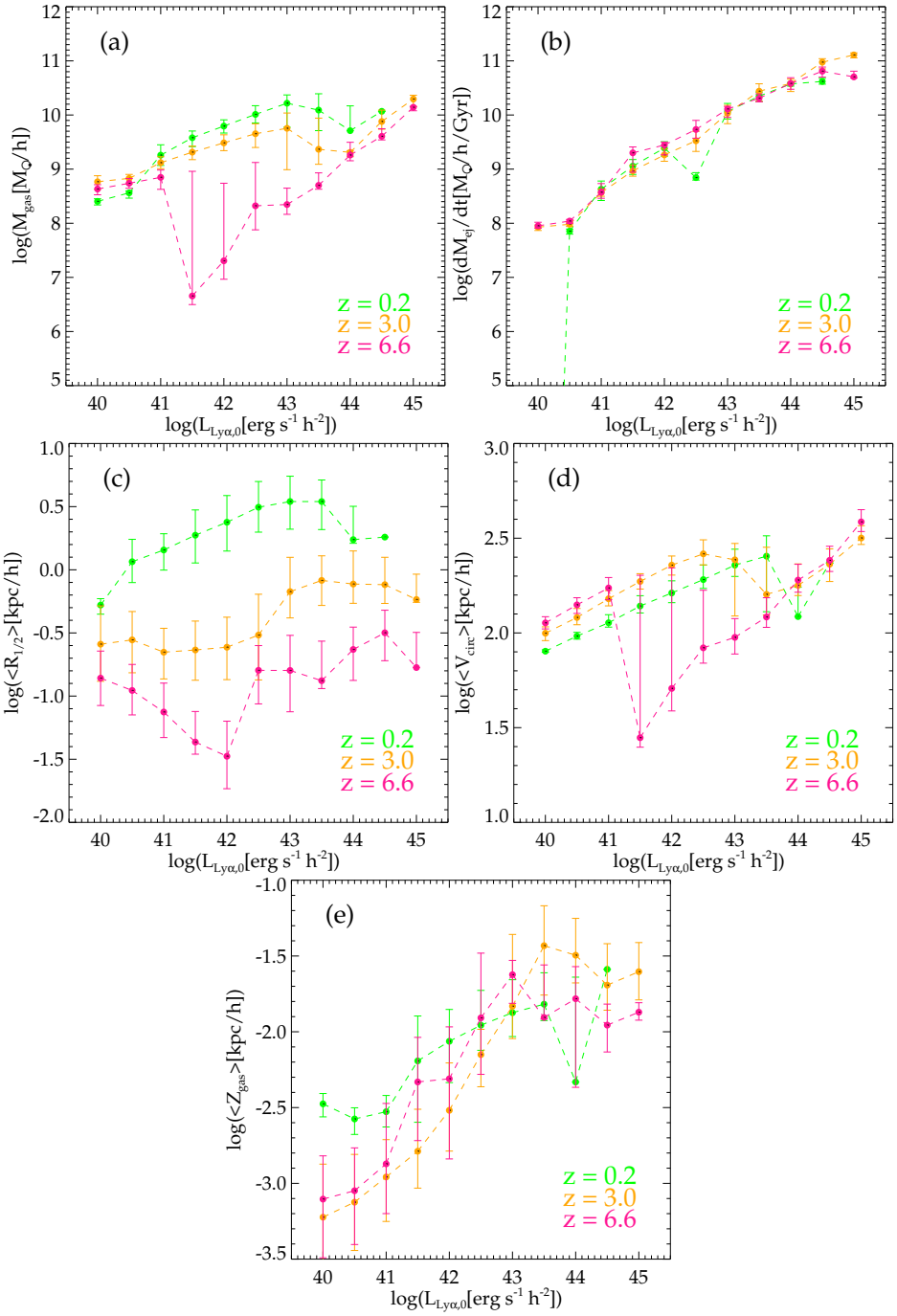

Figure 6.3: The evolution of the physical parameters used to define the outflow models for $z = 0.2$ (green), $z = 3.0$ (orange) and $z = 6.6$ (pink). Each panel shows, as a function of intrinsic Lyα luminosity, median values for (a) the cold gas mass of the galaxy $M_{\rm gas}$; (b) the mass ejection rate $\dot{M}_{\rm ej}$; (c) the luminosity-weighted half-mass radius $\langle R_{1/2} \rangle$; (d) the luminosity-weighted circular velocity $\langle V_{\rm circ} \rangle$; and (e) the mass-weighted metallicity of the gas $\langle Z_{\rm gas} \rangle$. Error bars show the 25-75 percentiles.

abundance of galaxies; however there is a luminosity range where bright quiescent galaxies and faint starbursts coincide in the same $L_{Ly\alpha,0}$ bin. In this luminosity range there is a discontinuity with the cold gas mass, since faint starbursts have smaller cold gas masses than bright quiescent galaxies. Luminosities above this region are dominated by starbursts.

The transition between quiescent and starbursts occurs at $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) \sim 44$ for $z = 0.2$ galaxies, and then it shifts to fainter luminosities for higher redshifts. At $z = 6.6$, starbursts start dominating at $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) > 41.5$. This reflects the overall change in the importance of starbursts relative to quiescent star formation with redshift.

The mass ejection rate is not found to evolve significantly with redshift, as shown in figure 6.3b, although it is tightly correlated with $L_{Ly\alpha,0}$. The mass ejection rate due to supernovae is directly proportional to the star formation rate, as shown in equations (6.12) and (6.13). Since the relation between $L_{Ly\alpha,0}$ and the star formation rate does not depend on redshift, it is not surprising that the mass ejection rate does not evolve with redshift either.

The luminosity-weighted half-mass radius $\langle R_{1/2} \rangle$ is the parameter that has the strongest evolution with redshift, as shown in figure 6.3c. Galaxies at $z = 0.2$ have typical sizes of $\langle R_{1/2} \rangle \sim$ 1kpc/h. The typical size of galaxies decreases rapidly with increasing redshift, falling by an order of magnitude or more at $z = 6.6$. In addition, $\langle R_{1/2} \rangle$ is found to correlate weakly with $L_{Ly\alpha,0}$.

The circular velocity of galaxies depends both on their mass and half-mass radius. Since the radius of galaxies correlates only weakly with the intrinsic Lyα luminosity, the luminosity-weighted circular velocity $\langle V_{\mathrm{circ}} \rangle$ is found to correlate with $L_{Ly\alpha,0}$, as shown in figure 6.3d, in the same way as the luminosity-weighted cold gas mass $\langle M_{\mathrm{gas}} \rangle$.

Finally, the mass-weighted metallicity $\langle Z_{\mathrm{gas}} \rangle$ correlates strongly with $L_{Ly\alpha,0}$ regardless of the redshift (figure 6.3e). The median metallicity of galaxies with $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) > 43.5$ is found to reach a maximum around $\langle Z_{\mathrm{gas}} \rangle \sim 0.01 - 0.03$.

The parameters shown in figure 6.3 are fed into the Monte Carlo radiative transfer code to produce the Lyα escape fraction f_{esc} and the line profile. The net Lyα luminosity of the

galaxy is simply

$$L_{\mathrm{Ly}\alpha} = \mathrm{f}_{\mathrm{esc}} \times L_{\mathrm{Ly}\alpha,0}, \tag{6.25}$$

where $L_{\mathrm{Ly}\alpha,0}$ is the intrinsic Lyα luminosity computed by GALFORM, as described in Chapter 3 of this thesis.

The escape fraction of a given galaxy depends on the choice of the outflow model. In addition to the geometrical configuration of each model, the input parameters used to compute the interior of the HI region are different as well. The number density in the **Shell** model is a function of the total cold gas mass of a galaxy and its size, whereas in the **Wind** model the number density depends on the mass ejection rate given by the supernova feedback model, the size and the circular velocity of the galaxy. This difference translates into different predicted properties of Lyα emitters, as will be shown in the next section. However, even when the column density, velocity of expansion and metallicity are the same, the two models will give different the escape fractions and line profile shapes due to their geometrical differences.

Figures 6.4, 6.5 and 6.6 illustrate the different line profiles obtained when matching the main outflow properties for outflows with different column densities. In order to make a fair comparison between the **Shell** and **Wind** outflow models, we compare configurations with the same column density, expansion velocities and metallicities. In addition, the inner radius in the **Wind** model is chosen to be the same as the outer radius in the **Shell** model.

The line profiles obtained can be characterised by the frequency distribution of photons according to the number of backscatterings (the number of times photons cross the inner empty region) they experience before escaping. The frequency at which backscattered photons peak depends on the ratio of the expansion velocity V_{exp} to the thermal (or, more generally, thermal and turbulent) velocity of the atoms, V_{th}, as shown by Eq. (5.10) of Chapter 5. When photons first hit the inner radius of the outflow, a fraction of them will be backscattered. These photons will reach the opposite side of the medium with frequency $x \approx -2V_{\mathrm{exp}}/V_{\mathrm{th}} \equiv 2x_{\mathrm{bs}}$. From those, a fraction will be backscattered again, reaching the medium at a different place with frequency $x \approx -3x_{\mathrm{bs}}$, and so on. This simple argument explains the position of the peaks in both models, as shown by the arrows in figures 6.4, 6.5 and 6.6, although each one contributes with a different proportion to the overall spectrum for

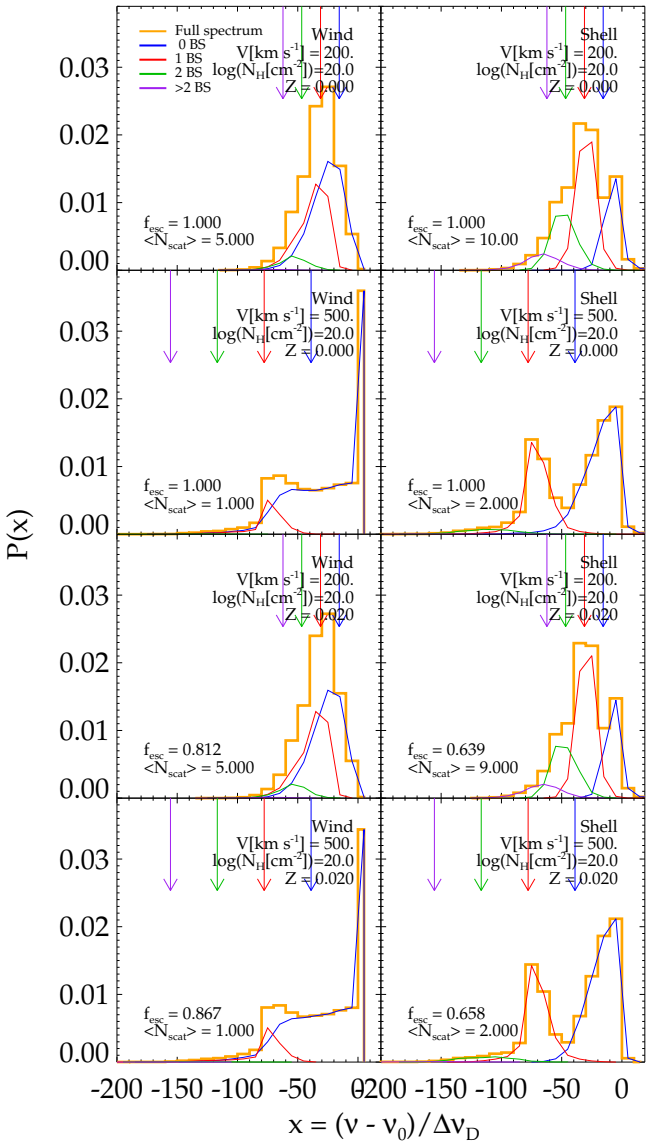

Figure 6.4: Comparison of the Lyα spectrum obtained with the **Wind** and **Shell** models for $N_H = 10^{20}[\mathrm{cm}^{-2}]$. The orange histogram shows the full spectrum for each case, whereas the blue, red, green and purple lines show the spectrum of photons which experienced 0, 1, 2 and 3 or more backscatterings before escaping, respectively. The coloured arrows show position of the frequencies $x = x_{\mathrm{bs}}, 2x_{\mathrm{bs}}, 3x_{\mathrm{bs}}$ and $4x_{\mathrm{bs}}$, where $x_{\mathrm{bs}} = -V_{\mathrm{exp}}/V_{\mathrm{th}}$, corresponding to the expected peaks for 0, 1, 2, and 3 backscatterings respectively (see text for details). Different values for the expansion velocity v_{exp} and the metallicity Z are used, as shown in the legend of each box. The temperature in all configurations is fixed at $T = 10000[K]$.

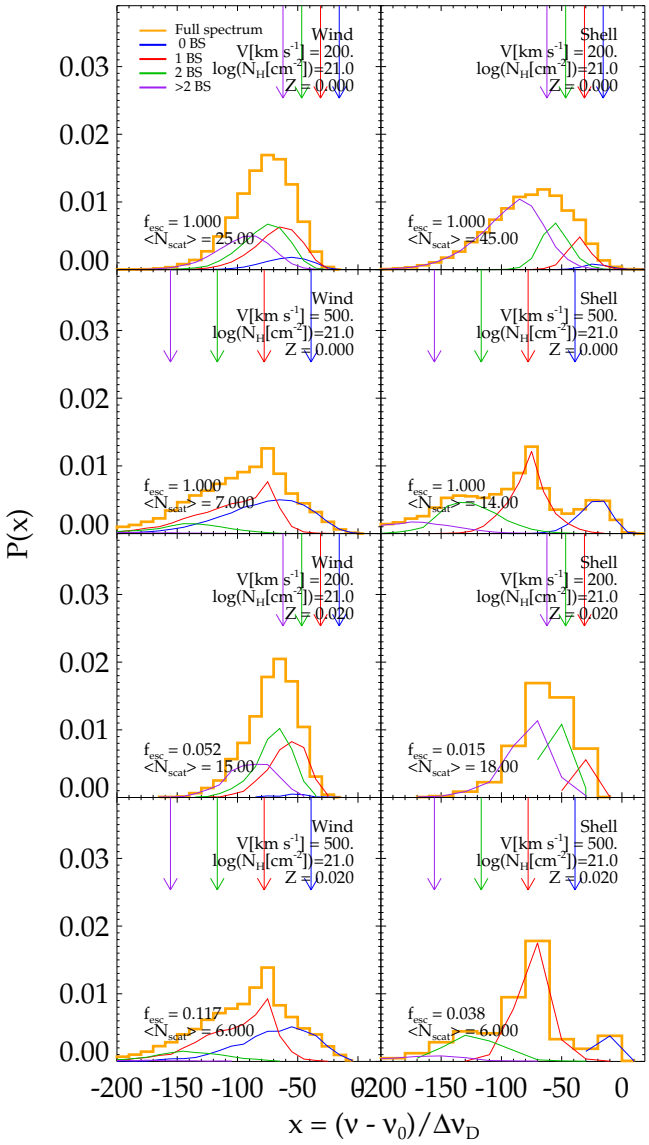

Figure 6.5: Similar to figure 6.4, but with Lyα spectra obtained for configurations with $N_H = 10^{21}[\mathrm{cm}^{-2}]$.

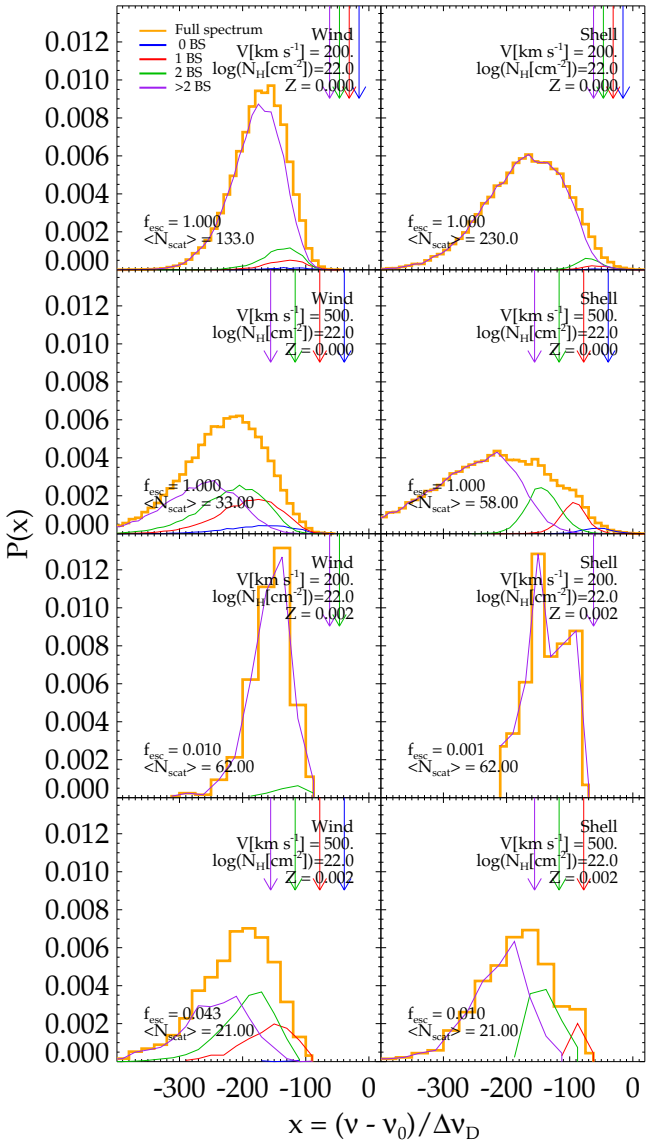

Figure 6.6: Similar to figure 6.4, but with Lyα spectra obtained for configurations with $N_H = 10^{22}[\mathrm{cm}^{-2}]$. Notice the wider range in the x-axis. The metallicity used in the configurations shown in the last two rows is $Z = 0.002$ instead of $Z = 0.02$, as in figures 6.4 and 6.5.

each model. Previous studies (Ahn, 2003, 2004; Verhamme et al., 2006) have also found this relation between the peaks of backscattered photons and $x_{\rm bs}$ in media with column densities of the order of $N_H \sim 10^{20}[{\rm cm}^{-2}]$, although they do not attempt to study the line profiles for higher optical depths as we do here. When studying very thick media we find that the peaks are displaced considerably from their expected position based on the above simple argument, as shown in figure 6.6, since the number of scatterings photons must undergo broadens the profiles and reddens the peaks position.

Figure 6.4 shows the line profiles from media with a column density of $N_H = 10^{20}[{\rm cm}^{-2}]$. The **Wind** model, with an expansion velocity of $V_{\rm exp} = 200[{\rm kms}^{-1}]$, presents an asymmetric, single peaked profile as opposed to the **Shell** model, which can be split into up to three visible different peaks, each one formed by photons being backscattered a different number of times.

For a higher expansion velocity of $V_{\rm exp} = 500[{\rm kms}^{-1}]$ the line profiles change drastically in both cases. In the **Wind** model there is a strong emission peak at the line centre of photons which did not undergo any scattering, followed by a flat plateau with a sharp break at the position of the first backscattering peak. For the same column density and expansion velocity, the **Shell** model splits into two asymmetric profiles, generated by photons with zero and one backscattering.

Figure 6.5 shows the same configurations as the previous case, but now the medium has a column density 10 times higher. At this larger column density photons escape after a larger number of scatterings, which is why the different profiles appear wider than in figure 6.4. Moreover, for media with column density $N_H = 10^{22}[{\rm cm}^{-2}]$ (figure 6.6), the Lyα profiles are broader than the previous two cases, and the position of the peaks have also shifted considerably. For very thick media only a small fraction of photons manage to escape the medium with fewer than 2 backscatterings.

In general, the fraction of photons escaping after different number of backscatterings depends on an *effective* optical depth $\tau_e \equiv \tau(x = -(N+1)V_{\rm exp}/V_{th})$, where N is the number of backscatterings, and the velocity of the medium. With each backscattering photons are redsfhifted and thus the optical depth is decreased. After a number of backscatterings photons will reach an effective optical depth $\tau_e \lesssim 1$, which will enable them to escape from the

medium.

For the configurations with $N_H = 10^{20}[\mathrm{cm}^{-2}]$ shown in figure 6.4, photons reach $\tau_e \lesssim 1$ after the first backscattering, which is why this peak is dominant in the overall line profile. On the other hand, for $N_H = 10^{22}[\mathrm{cm}^{-2}]$, photons reach $\tau_e \lesssim 1$ after three or more backscatterings, which is why this one becomes the dominant peak in figure 6.6.

Figures 6.4, 6.5 and 6.6 also show the effect of including dust in the media described above. Overall, dust seems to remove a similar fraction of photons at all escaping frequencies, and the shape of the obtained profiles is not significantly different from the dust-free version. In figures 6.4 and 6.5 we show media with a metallicity $Z = 0.02$. We found that for a column density $N_H = 10^{22}[\mathrm{cm}^{-2}]$ such value of the metallicity results in the absorption of all photons, which is why in figure 6.6 we use instead $Z = 0.002$ to illustrate the effect of dust attenuation. It is interesting to note that the configurations with dust in figures 6.5 and 6.6 have different metallicities but the same optical depth of dust (since it also depends on the column density).

However the escape fractions found are a factor 3-5 smaller in the configurations with $N_H = 10^{22}[\mathrm{cm}^{-2}]$ than in the configurations with $N_H = 10^{21}[\mathrm{cm}^{-2}]$, since the number of scatterings is higher in the former case, thus making photons more sensitive to the same optical depth of dust.

The shell model is found to be more sensitive to dust than the outflow model for configurations with the same column density, expansion velocity, metallicity and radius. This is a direct consequence of the median number of scatterings photons undergo inside the medium, which is found to be approximately twice as many in the shell compared to the outflow in all configurations studied. However, it must be kept in mind that for these calculations the median number of scatterings is obtained using the accelerated version of the code, so it does not corresponds to the physical median number of scatterings for the configurations studied.

6.5 A grid of configurations to compute the escape fraction

GALFORM typically generates samples numbering many thousands of galaxies brighter than a given flux limit at a number of redshifts. The task of running the radiative transfer code for each one of those galaxies is infeasible considering the time it takes the Monte Carlo code to reach completion, which varies from a few seconds up to several hours for some extreme configurations.

Therefore, this motivates the need to develop an alternative method to assign a Lyα escape fraction for each galaxy predicted by GALFORM. Instead of running the radiative transfer code to each galaxy, we construct a grid of configurations spanning the whole range of galaxy properties, as predicted by GALFORM.

The first step to construct the grid is to choose which parameters will be used. In principle, each outflow model (wind or shell) requires 4 input parameters from GALFORM: three of these, $\langle V_{\rm circ}\rangle, \langle R_{1/2}\rangle$ and $\langle Z_{\rm gas}\rangle$ are used by both models. In addition, $\dot{M}_{ej}$ is required in the **Wind** model, and $M_{\rm HI}$ in the **Shell** model.

However, a grid of models using four parameters becomes rapidly inefficient when trying to refine the grid. A grid with an appropriate binning of each parameter can have as many elements as the number of galaxies for which the grid was constructed for, and hence also becomes prohibitively expensive.

Therefore, we look for degeneracies in the escape fraction when using combinations of the input parameters from GALFORM, in order to reduce the number of parameters for the construction of the grid. The idea is to find a combination of parameters which, when kept fixed while varying its individual components, gives the same escape fraction.

The natural choice for this is to use the column density N_H as one parameter. Neufeld (1990) found that the escape fraction from a homogeneous, dusty slab is a function of the optical depth at the line centre τ_0 and the optical depth of absorption τ_a (see equation 5.98 in Chapter 5). Both quantities are, in turn, a function of the column density N_H. In the **Shell** model, $N_H \propto M_{\rm gas}/\langle R_{1/2}\rangle^2$, whereas in the **Wind** model $N_H \propto \dot{M}_{ej}/(\langle R_{1/2}\rangle V_{\rm exp})$. Although

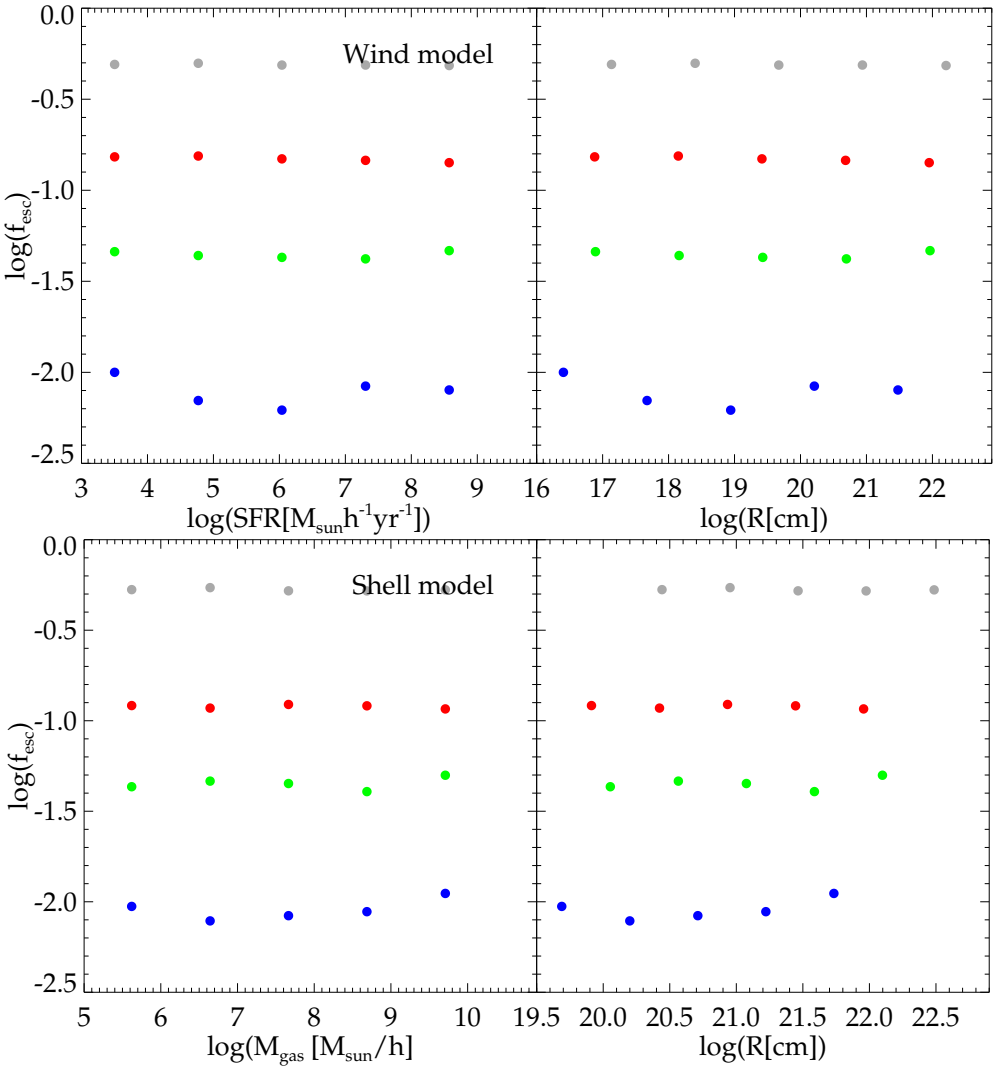

Figure 6.7: The Lyα escape fraction of configurations with different parameters but with a fixed combination of 2 parameters. The top panels show the effect of varying the mass ejection rate $\dot{M}_{\rm ej}$ (left) and the inner radius $R_{\rm inn}$ (right) but keeping the ratio $C_{\rm wind} = \dot{M}_{\rm ej}/R_{\rm inn}$ constant. The lower panels shows the same effect on the mass of the outflow $M_{\rm gas}$ and the radius $R_{\rm out}$ when fixing $C_{\rm shell} = M_{\rm gas}/R^2_{\rm out}$. Different colours correspond to different choices of $C_{\rm wind}, C_{\rm shell}$ known to give different values of the escape fraction.

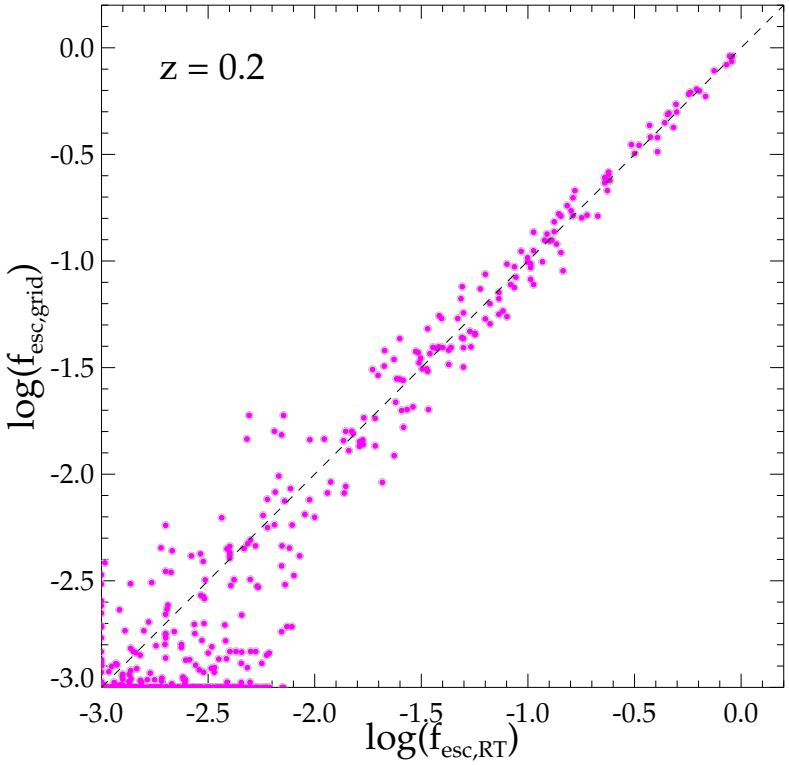

Figure 6.8: Comparison of the escape fraction obtained using a direct calculation ($f_{\rm esc,RT}$) or interpolating from a grid ($f_{\rm esc,grid}$). Points represent a subsample of ~ 1000 galaxies selected from GALFORM at $z = 0.2$.

promising, we find that we do not recover a constant escape fraction in the **Wind** model when the column density is kept fixed while varying its individual terms. The reason is that the expansion velocity plays a more complicated role when computing the escape fraction, with the escape fraction increasing rapidly with increasing velocity regardless of the other parameters of the medium.

Figure 6.5 illustrates a suitable combination of parameters for each model which keeps the escape fraction constant. In the **Wind** model, we find that the escape fraction is insensitive to variations in $\dot{M}_{\rm ej}$ and $R_{\rm inn}$ when the parameter $C_{\rm wind} \equiv \dot{M}_{\rm ej}/R_{\rm inn}$ is left constant. Likewise, in the **Shell** model, the escape fraction does not vary for different values of $M_{\rm gas}$ and $R_{\rm out}$ when $C_{\rm shell} \equiv M_{\rm gas}/R_{\rm out}^2$ is fixed.

Therefore, we construct three-dimensional grids for each outflow model. In the **Wind** model the parameters are $C_{\rm wind}, V_{\rm exp}$ and $Z_{\rm gas}$, whereas in the **Shell** model we use $C_{\rm shell}, V_{\rm exp}$ and $Z_{\rm gas}$. We choose to cover each parameter with a bin size appropriate to recover the

expected escape fraction with a reasonable accuracy when interpolating in the grid, but also ensuring that the number of grid elements to be computed is significantly smaller than the total number of galaxies in the sample.

We find that, when covering each parameter in logarithmic bins of 0.1 we get escape fractions that are accurate enough, and the number of elements of the grid we need to compute is usually a factor ~ 20 smaller than the total number of galaxies in the sample.

We fix the number of photons to run for each grid point to compute the escape fraction, since this will determine the speed at which each configuration will be completed. By studying the resulting luminosity function of galaxies (see next section), we find that running the code with a maximum number of photons $N_p = 1000$ gives results which have converged over the range of luminosities observed. This means that the minimum escape fraction we are able to compute is $f_{esc} = 10^{-3}$. Although there are configurations where the f_{esc} can be lower than this, they do not contribute significantly to the luminosity functions, as we discuss in the next section.

Figure 6.8 shows an example of the performance of the grid we use to compute the escape fraction in the **Shell** model using a sub-sample of galaxies from GALFORM at $z = 0.2$, chosen in a way to cover the entire range of intrinsic Lyα luminosities. The accuracy of the grid gets progressively worse with lower escape fractions, since these have intrinsically larger errors due to the constraint on the maximum number of photons used to compute f_{esc}. However, as discussed previously, we found that there is no need to reduce the size of the parameter bins or increase the number of photons used to reproduce accurately the luminosity functions.

6.6 Reproducing the observational properties of Lyα emitters

The above procedure allows us to efficiently assign an escape fraction to each galaxy in GALFORM. However, the outflow models also have a set of free parameters which need to be fixed before constructing the grid of configurations. In order to set the value of the free parameters $[f_M, f_V, f_R]$ for the **Shell** model, and $[f_V, f_R]$ for the **Wind** model, we attempt to fit the observed cumulative Lyα luminosity function (CLF) in the redshift range $0.2 < z < 6.6$.

To do this we generate a subsample of $\sim$ 1000 galaxies at different redshifts, and we run the radiative transfer code directly over each one. By doing this we can efficiently seek for the most suitable choice of parameters for each model, which is then used to construct the multidimensional grid described previously.

6.6.1 Tuning the models with the observed Lyα luminosity functions

In principle, we would like to find a single set of parameter values to reproduce the observed luminosity functions at different redshifts. Figure 6.9 shows a comparison of the predicted Lyα CLFs at several redshifts compared to observational measurements, choosing the parameter values to best fit the observed luminosity function at $z = 3.0$. Observational data is taken from Deharveng et al. (2008) and Cowie et al. (2010) at $z = 0.2$; Gronwall et al. (2007), Ouchi et al. (2008) and Rauch et al. (2008) at $z = 3.0$; Shimasaku et al. (2006), Ouchi et al. (2008) and Hu et al. (2010) at $z = 5.7$; and Kashikawa et al. (2006) and Hu et al. (2010) at $z = 6.6$.

Overall, the predicted CLFs at redshifts different from $z = 3.0$ do not reproduce the observational measurements. At $z = 0.2$, both models largely over-predict the abundance of Lyα emitters. At $z = 5.7$ and $z = 6.6$, the predicted CLF with both models is below the observational measurements. Figure 6.9 also shows that the fiducial model, that we used in Chapter 3, with a constant escape fraction of $f_{esc} = 0.02$ seems to reproduce well the observed CLFs at all redshifts except $z = 0.2$. However, it must be kept in mind that the scenario of a constant escape fraction, regardless of the galaxy properties, is unphysical, given the complexity of the physics determining the escape of Lyα photons. Therefore, the difficulty to fit the observed CLFs with a single choice of model parameters suggests that additional physical processes not included in our model might be playing a crucial role, like the escape from a multi-phase medium, the attenuation of the Lyα radiation by the IGM, or an erroneous calculation of the *intrinsic* Lyα luminosity of galaxies, which does not depends upon our Lyα radiative transfer modelling. Also, our outflow models are possibly too simplified, since we are assuming a spherical geometry with a central empty region and photons generated in the centre of it.

We can test whether `GALFORM` computes the correct intrinsic Lyα luminosity by studying

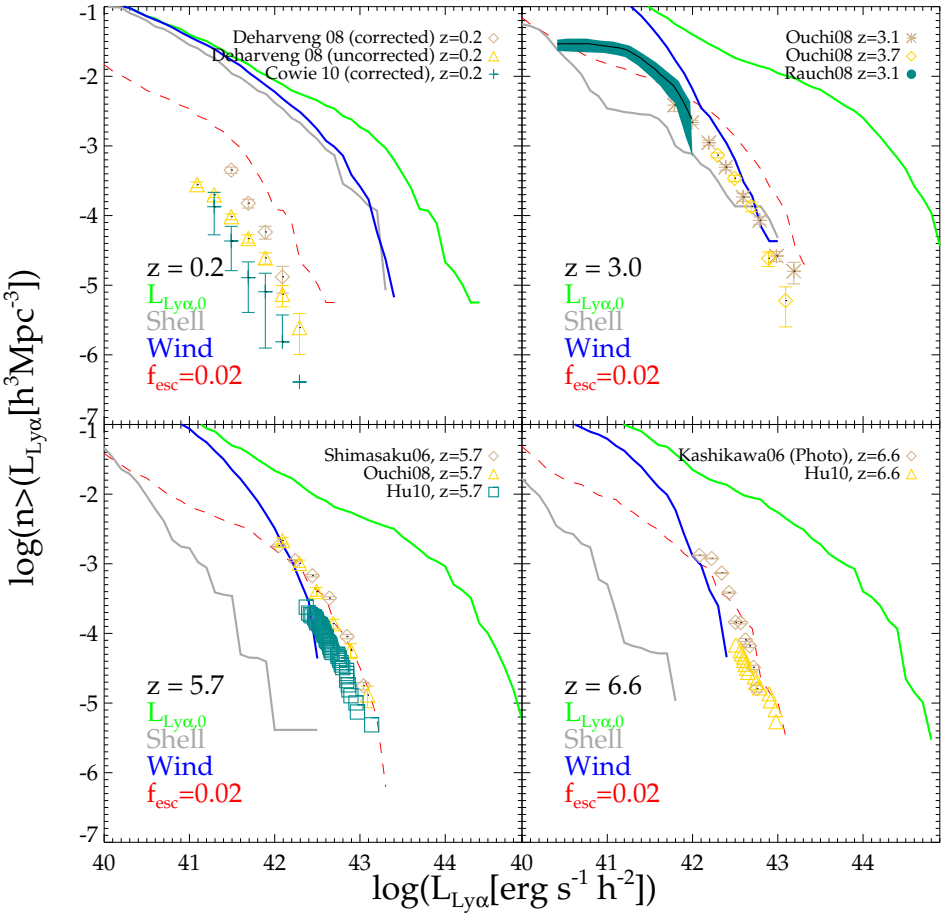

Figure 6.9: The cumulative luminosity function of Lyα emitters for redshifts $z = 0.2, 3.0, 5.7$ and 6.6. Symbols show observational estimates of the Lyα CLF. Green curves show the CLF obtained when using the intrinsic Lyα luminosity, with no attenuation. The blue and gray curves show the CLF obtained by using the **Wind** and **Shell** models, respectively, where the free parameters of each model were chosen to match the observed CLF at $z = 3.0$. The red dashed lines show the result of applying a constant escape fraction ($f_{esc} = 0.02$) in the CLF at every redshift.

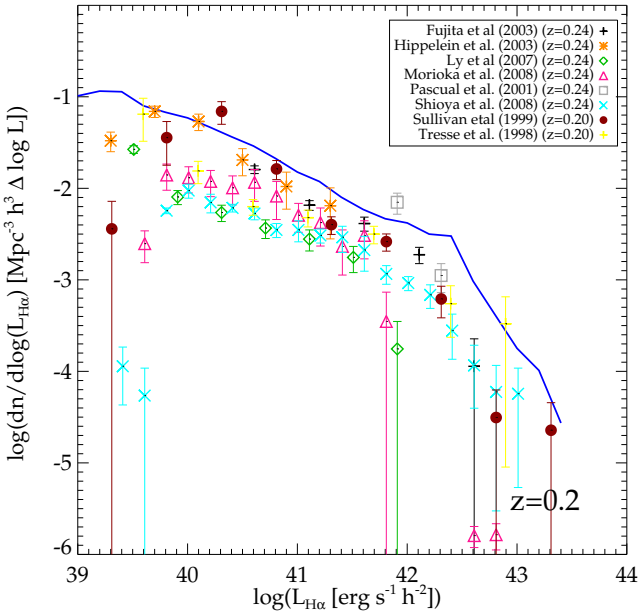

Figure 6.10: The unattenuated LF of Hα emitters at $z = 0.2$. Symbols show observational data from different authors as described in the legend. The blue curve show the Hα predicted LF of the Baugh et al. (2005) model.

the intrinsic (unattenuated) Hα luminosity function, since both emission-line luminosities are directly related to the rate of hydrogen recombinations, which are in turn linked to the production rate of Lyman continuum photons, so they differ only by their case B recombination emission coefficient. As discussed in Chapter 4, the Hα emission from galaxies is less sensitive to dust than Lyα since these photons do not undergo resonant scattering in the ISM, which makes their path lengths shorter than the typical path lengths Lyα photons experience. In addition, the small attenuation by the existing dust in the galaxy can be estimated observationally by computing the ratio of the intensity of two or more emission lines and comparing to what is expected from case B recombination (see, for example, Kennicutt, 1983, 1998b).

Figure 6.10 shows the resulting Hα luminosity function at $z = 0.2$ after correcting by extinction removing the attenuation of the line by dust, i.e. using the intrinsic Hα luminosity. Notice that this LF differs from the one shown in figure 4.1 of Chapter 4, where data attenuated by dust was shown instead. The observational estimates of the unattenuated Hα LF are taken from Fujita et al. (2003); Hippelein et al. (2003); Ly et al. (2007); Morioka et al. (2008); Pascual et al. (2001); Shioya et al. (2008); Sullivan et al. (2000) and Tresse and Maddox (1998). The predicted Hα LF from `GALFORM` is found to be shifted by roughly a fac-

	$z = 0.2$	$z = 3.0$	$z = 5.7$	$z = 6.6$
Shell model				
f_M	0.10	0.10	0.10	0.10
f_R	0.30	1.00	1.80	1.80
f_v	1.00	1.00	1.00	1.00
Wind model				
f_R	0.01	0.15	0.50	0.50
f_v	1.00	1.00	1.00	1.00

Table 6.1: Summary of the parameter values of the **Shell** and **Wind** models used to fit the Lyα cumulative luminosity function at different redshifts.

tor ~ 3 towards bright luminosities. This means that the intrinsic Lyα luminosity at $z = 0.2$ obtained with GALFORM could also be a factor ~ 3 overestimated. Nevertheless, the Lyα luminosities obtained at $z = 0.2$ are a factor ~ 10 brighter than what is needed to reproduce the observational results, meaning that the uncertainty in the intrinsic Lyα luminosity is not responsible for the offset found in the LFs.

In order to reproduce the observed abundances of Lyα emitters at $0.2 < z < 6.6$ we vary the free parameters of the models at every redshift studied to find the best combination of values. Table 6.1 shows the parameter values found for each redshift. For simplicity, we set $f_V = 1.0$ in both models. Also, we chose $f_M = 0.1$ in the **Shell** model for every redshift, so, in practise, only f_R varies as a function of redshift. Therefore, all the model predictions shown in the following make use of the parameter values shown in table 6.1.

Figure 6.11 shows the resulting CLFs when using the model parameters shown in table 6.1. When comparing the predicted CLFs from each model we notice a difference in the predicted slope from both models. This arises because the **Wind** model produces higher escape fractions than the **Shell** model at faint luminosities, whereas at the bright end both models give similar luminosity functions (since the free parameter values were chosen to match the observed CLFs).

Current observational samples have only measured the faint end of the luminosity func-

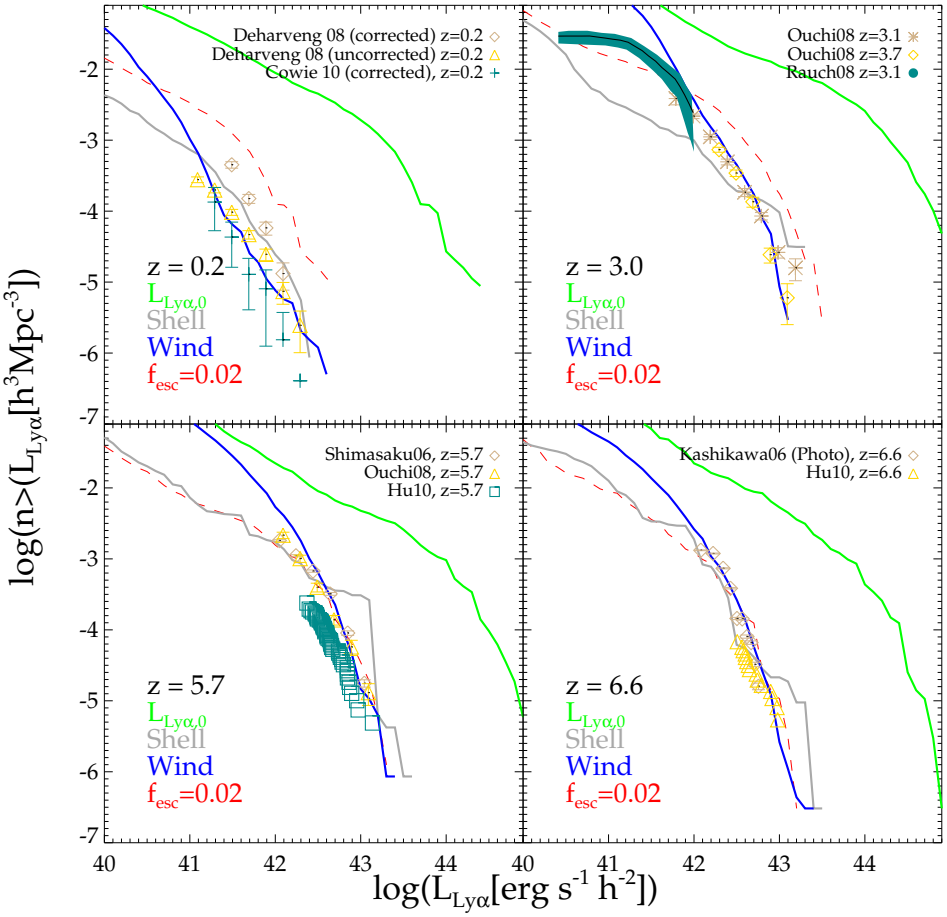

Figure 6.11: Same as figure 6.9, but here the free parameters of the models were chosen to match the observed CLF at each redshift (see Table 6.1).

tion at $z = 3.0$ (Rauch et al., 2008), so it is difficult to assess which model reproduces the faint end of the CLF the best at other redshifts. Both models seem to bracket the measurement of the faint end of the CLF at $z = 3.0$, although the CLF predicted by the **Wind** model is closer to the Rauch et al. (2008) data.

It is interesting to notice that the **Shell** model, with the choice of parameters shown in table 6.1, produces similar CLFs to those obtained with a constant escape fraction of $f_{esc} = 0.02$ over the luminosity range for redshifts $z = 5.7$ and $z = 6.6$. Nevertheless, as we will see next, the escape fractions of galaxies predicted by the **Shell** model are anything but constant. To avoid confusion, we emphasize that, in the following, all model predictions are calculated using the parameter values shown in table 6.1.

6.6.2 The Lyα escape fractions

As pointed out earlier, our outflow models are able to compute a minimum value for the escape fraction of 10^{-3} when one photon out of 1000 escapes. However, in a significant fraction of galaxies none of the photons escape from the outflows, as shown in figure 6.12. For these cases, our models can only place an upper limit to the value of the *real* escape fraction.

As expected, the fraction of galaxies with real escape fractions $f_{esc} < 10^{-3}$ increases with intrinsic Lyα luminosity, since, as shown in figure 6.3, $M_{gas}, \dot{M}_{ej}$ and Z_{gas} increase towards higher $L_{Ly\alpha,0}$, thus making the optical depth in the outflows thicker. Up to an intrinsic Lyα luminosity of $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) \sim 42$, the **Wind** model has a negligible fraction of galaxies with $f_{esc} < 10^{-3}$. The Shell model, on the other hand, shows a much higher fraction of galaxies with $f_{esc} < 10^{-3}$, with approximately half of the galaxies at the faintest intrinsic Lyα luminosity considered of $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) = 40$ having escape fractions below our numerical limit of 10^{-3}.

We assessed the effect of increasing the number of photons used to compute f_{esc} from 1000 to 10000, finding that only a negligible fraction of galaxies have escape fractions in the range 10^{-3} to 10^{-4}. Furthermore, the effect of increasing the number of photons used by a factor 10 is also negligible in the predicted Lyα luminosity functions, since the number density of galaxies with escape fractions below the numerical limit is small. Therefore, although our outflow models cannot compute a non-zero escape fraction for all galaxies in GALFORM, our statistical results are accurate enough and allows us to compare the distribution of escape fractions observed with what our models predict.

Observational estimates of the Lyα escape fraction are generally based on either inferring the SFRs from the intrinsic Lyα luminosity and comparing them to the UV continuum SFR, or on the line ratio between Lyα and other non-resonant hydrogen recombination line. The second method is regarded as more direct, since the first one depends on assumptions on the stellar evolution models (the choice of the IMF, the modelling of dust) and the calibrations of the SFR, whereas the second one only relies in the assumption that the extinction of the comparison line can be estimated from an extinction curve, and that the intrinsic line ratio corresponds to the ratio of the emission coefficients assuming case B recombination.

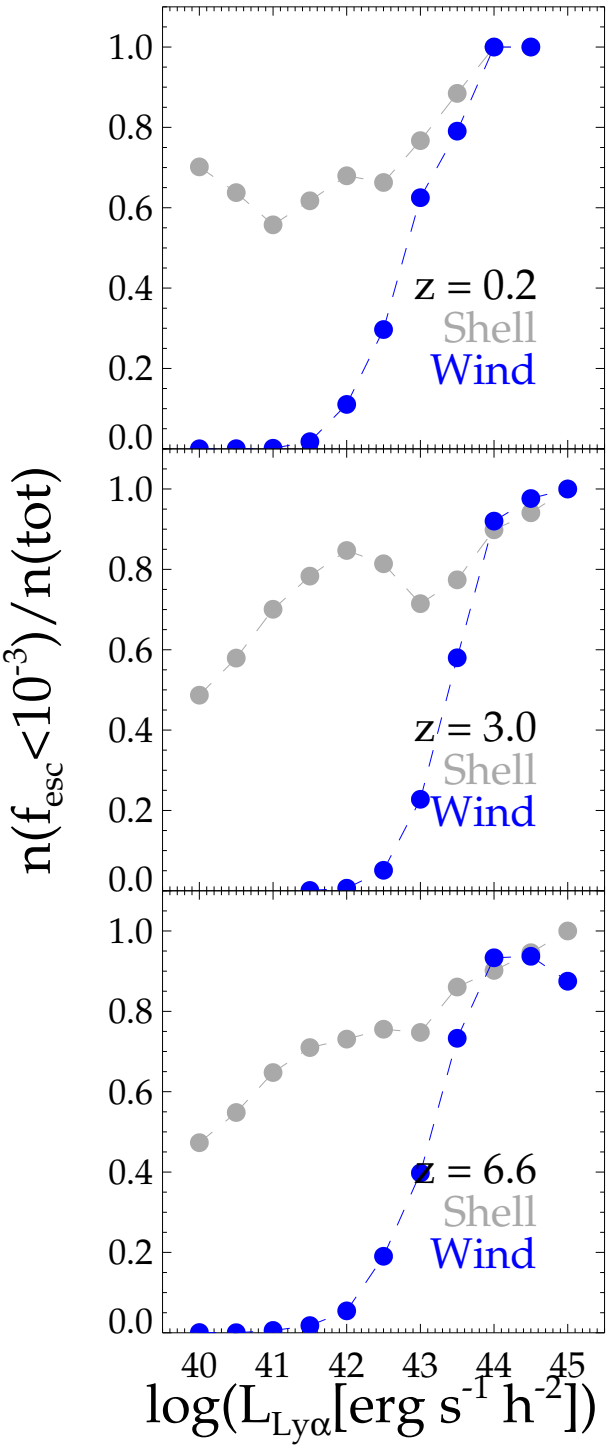

Figure 6.12: The fraction of galaxies with f_{esc} below 10^{-3} as a function of intrinsic Lyα luminosity for redshifts $z = 0.2$ (top), $z = 3.0$ (middle) and $z = 6.6$ (bottom). Grey and blue points show the results from the **Shell** and **Wind** models respectively.

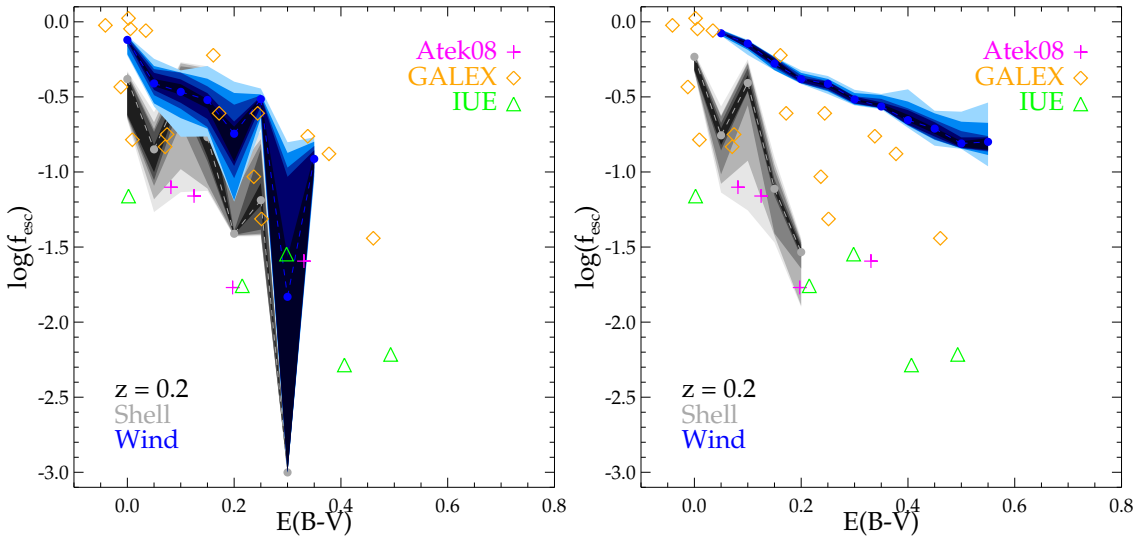

Figure 6.13: The escape fraction as a function of extinction at $z = 0.2$. Symbols show observational measurements. Blue and grey circles show the median of the predicted distribution for the **Wind** and **Shell** models respectively. The coloured regions show different percentiles around the median values so that 10% to 50% of the simulated data is shown from the darkest to the brightest colours. The left panel shows the predicted $E(B-V)$ using the extinction of the continuum computed by GALFORM to estimate the attenuation of the $H\alpha$ and $H\beta$ lines, whereas the right panel shows the predicted $E(B-V)$ using the extinction computed by using the column density and metallicity of the outflows.

For this reason we focus on the escape fractions measured using line ratios. These are often presented as a function of dust extinction $E(B-V)$ estimated from non-resonant recombination lines such as Hα and Hβ (Atek et al., 2008, 2009; Östlin et al., 2009; Hayes et al., 2010a).

The method usually consists of inferring $E(B-V)$ by using (Atek et al., 2008)

$$E(B-V) = \frac{2.5 \times \log(2.86/R_{\rm obs})}{k(\lambda_\alpha) - k(\lambda_\beta)}, \tag{6.26}$$

where $R_{\rm obs} = f_{H\alpha}/f_{H\beta}$ is the measured line ratio between $H\alpha$ and $H\beta$, and $k(\lambda_\alpha) = 2.63, k(\lambda_\beta) = 3.71$ are the values of the extinction at each corresponding wavelength from the Cardelli et al. (1989) extinction curve. The constant 2.86 corresponds to the assumed intrinsic line ratio under case B recombination for a medium at a temperature of $T = 10000[K]$. (Osterbrock, 1989)

GALFORM predicts the intrinsic and attenuated $H\alpha$ and $H\beta$ luminosities, so we can use these to compute $E(B-V)$ from equation (6.26). Moreover, the attenuation at each line is computed as the attenuation of the continuum at each corresponding wavelength using the radiative transfer model of Ferrara et al. (1999), taking into consideration the density profiles of the disk and bulge and a given inclination, as discussed in Chapter 2 of this thesis.

Alternatively, we can also calculate the attenuation of $H\alpha$ and $H\beta$ when passing through the outflow, since we know the column density and the metallicity of the outflow in each galaxy. The line ratio is then computed as

$$R_{\rm obs} = \frac{L_{H\alpha,0}}{L_{H\beta,0}} e^{[\tau_a(H\beta) - \tau_a(H\alpha)]}, \tag{6.27}$$

where $L_{H\alpha,0}$ and $L_{H\beta,0}$ are the intrinsic Hα and Hβ luminosities predicted by GALFORM respectively, and τ_a at each line can be computed by using equations (6.9) and (6.10).

Figure 6.13 shows the predicted relation between the Lyα escape fraction and the extinction $E(B-V)$ compared to observational data from Atek et al. (2008) and an analysis of UV spectroscopic data from the GALEX and IUE surveys by Atek et al. (2009).

The model predictions shown in figure 6.13 include only galaxies with $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) > 41.5$, in order to reproduce the selection of Lyα emitters in the GALEX survey (Deharveng et al., 2008), although it is worth pointing out that the observational points shown in figure 6.13 do not represent a statistical sample, and instead are measurements of individual sources.

Overall, both models show that with increasing extinction, the escape fraction of Lyα decreases, which is in qualitative agreement with the observations. The $E(B-V)$ values computed with the dust extinction model of GALFORM show a stronger anti-correlation which seems to agree better with the observational results than the $E(B-V)$ computed with the extinction due to the outflows. In this case, the **Wind** model predicts a flatter correlation between f_{esc} and $E(B-V)$ than the **Shell** model.

As we will see later, the difference on the predicted $E(B-V)$ value is directly related to the column density N_H of the outflow, and, in general, the **Wind** model predicts a larger column density for the outflows than the **Shell** model.

6.6.3 The Lyα UV luminosity function

Observationally, the lower amplitude of the bright end of the Lyα LF at $z = 6.6$ compared with the Lyα LF at $z = 5.7$ has been suggested as evidence that the ionization state of the IGM is different between these two redshifts (Kashikawa et al., 2006). This interpretation is supported by two arguments: first, observational data seems to be consistent with little or zero evolution in the Lyα LF between $z = 3.0$ to $z = 5.7$ (Shimasaku et al., 2006; Ouchi et al., 2008), so it is at least surprising that there is little evolution over a time period of over 1 Gyr, and then a significant evolution over ~ 0.2 Gyrs. The second argument is the measured lack of evolution of the UV LF of Lyα emitters from $z = 5.7$ to $z = 6.6$ (Kashikawa et al., 2006), which is assumed to be insensitive to the ionization state of the Universe.

Figure 6.14 shows the UV LF of galaxies selected as Lyα emitters at redshifts $z = 5.7$ and $z = 6.6$. Here, the UV magnitude corresponds to the z'-band of the Subaru Suprime-Cam (with mean wavelength $\lambda_{z'} \approx 9500$Å), which, in the rest frame at $z \approx 6$ gives a mean wavelength around 1300Å. Observational data at $z = 5.7$ comes from Shimasaku et al. (2006),

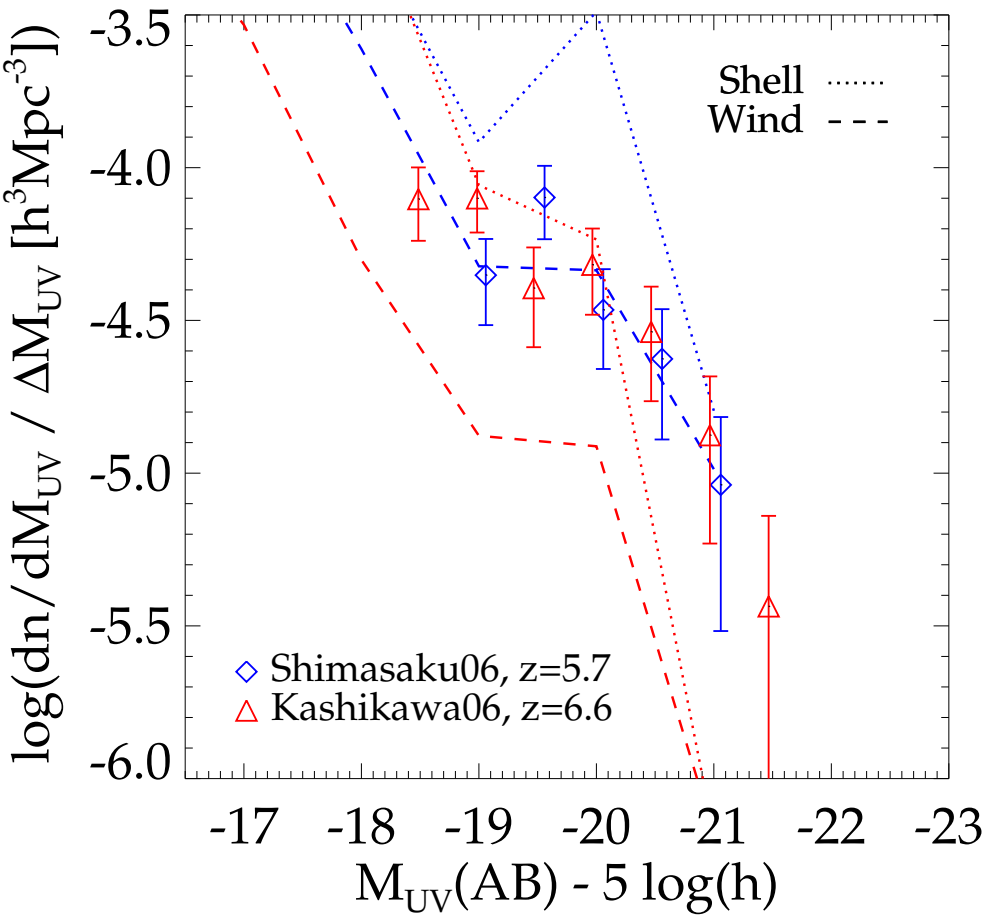

Figure 6.14: The UV luminosity function of Lyα emitters at $z = 5.7$ (in blue) and $z = 6.6$ (in red). Different symbols show observational measurements of the UV LF from a sample of Lyα emitters. The dotted lines show the predicted UV LF of the **Shell** model. The dashed line shows the predicted UV LF of the **Wind** model.

and at $z = 6.6$ from Kashikawa et al. (2006).

The obtained UV LFs with the **Shell** model seem to be consistent with the observational measurements at both redshifts, whereas the **Wind** model predicts a consistent UV LF at $z = 5.7$ only. The UV LF at $z = 6.6$ in the **Wind** model has a lower amplitude than the observed one, which is consistent with the offset in the Lyα CLF at $z = 6.6$ shown in figure 6.11.

It is interesting to notice that the **Shell** model seems to be more consistent with the lack of evolution in the UV LF at these two redshifts, whereas the **Wind** model predicts an offset of the order of 1 magnitude. The Baugh et al. (2005) version of GALFORM does not include a treatment of the attenuation of light by the IGM, although it does prevent the cooling of haloes with $V_{\rm circ} < 60[{\rm kms}^{-1}]$ at $z < z_{\rm reio}$, where the redshift of reionization is assumed to be $z_{\rm reio} = 10$. Therefore, our predictions show that current observational measurements are not sufficient to invoke a partial ionization state of the IGM at redshifts $z \sim 6$ to explain the Lyα and UV luminosity functions. Instead, the relative number densities between $z = 5.7$ and $z = 6.6$ are sensitive, in our case, to the choice of the outflow model.

6.6.4 The observed Lyα line profiles

Several observational measurements of the Lyα line profiles show characteristic features which suggest the presence of outflows in galaxies (e.g. Shapley et al., 2003; Kashikawa et al., 2006; Dawson et al., 2007; Hu et al., 2010; Kornei et al., 2010). Figure 6.15 shows the result of stacking the Lyα line profiles of a subsample of galaxies from GALFORM at different redshifts. The limiting luminosity used, $\log(L_{Ly\alpha}[{\rm erg\ s}^{-1}\ {\rm h}^{-2}]) > 42.0$, is set to reproduce a typical Lyα limit from a high redshift survey.

We find that the stacking of Lyα profiles reveals an asymmetric peak in both models at all redshifts studied. Furthermore, the detailed shape of the profile depends on the outflow model used. The **Shell** model produces narrow asymmetric profiles, with increasing FWHM for lower redshifts. It also produces a prominent double peaked profile at $z = 0.2$, evidencing the backscattering process discussed in section 6.4. A very small peak at the line centre, corresponding to photons escaping without experiencing any backscattering, is also evident

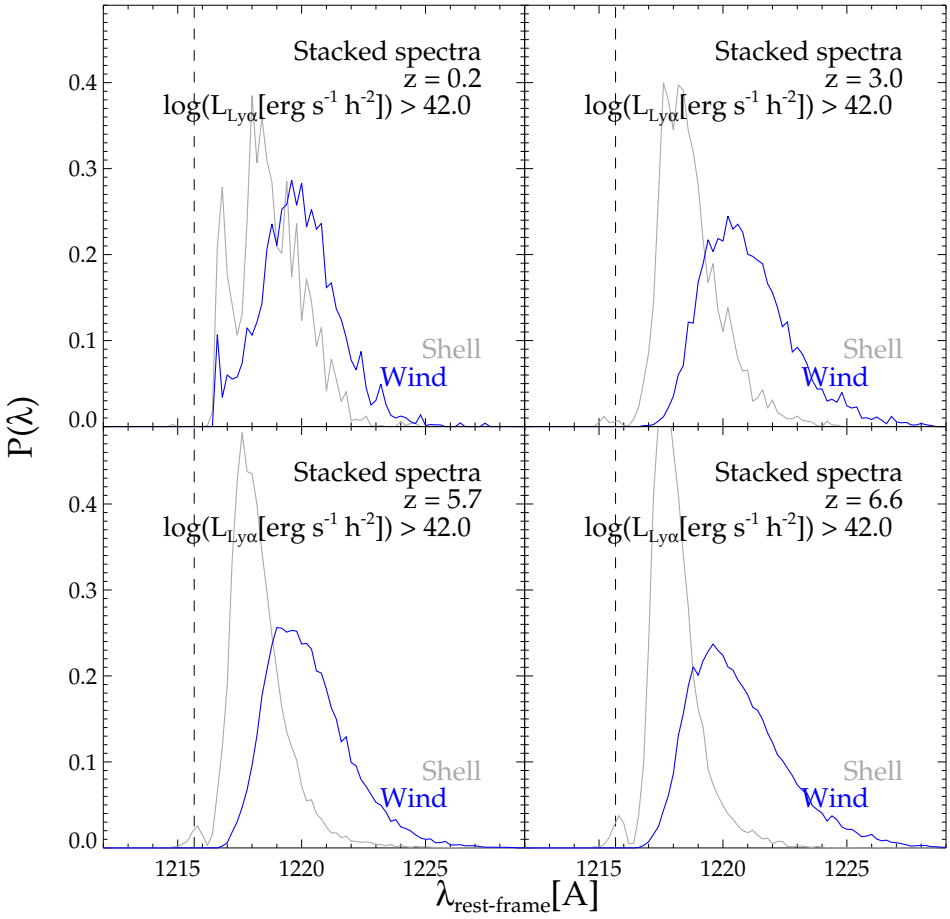

Figure 6.15: The stacked Lyα line profile at redshifts $0.2 < z < 6.6$ for both models. The line profiles of galaxies with $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) > 42.0$ and $EW_{\mathrm{rf}} > 20$ Å were stacked. The gray and blue line shows the obtained stacked spectra with the **Shell** and **Wind** models, respectively.

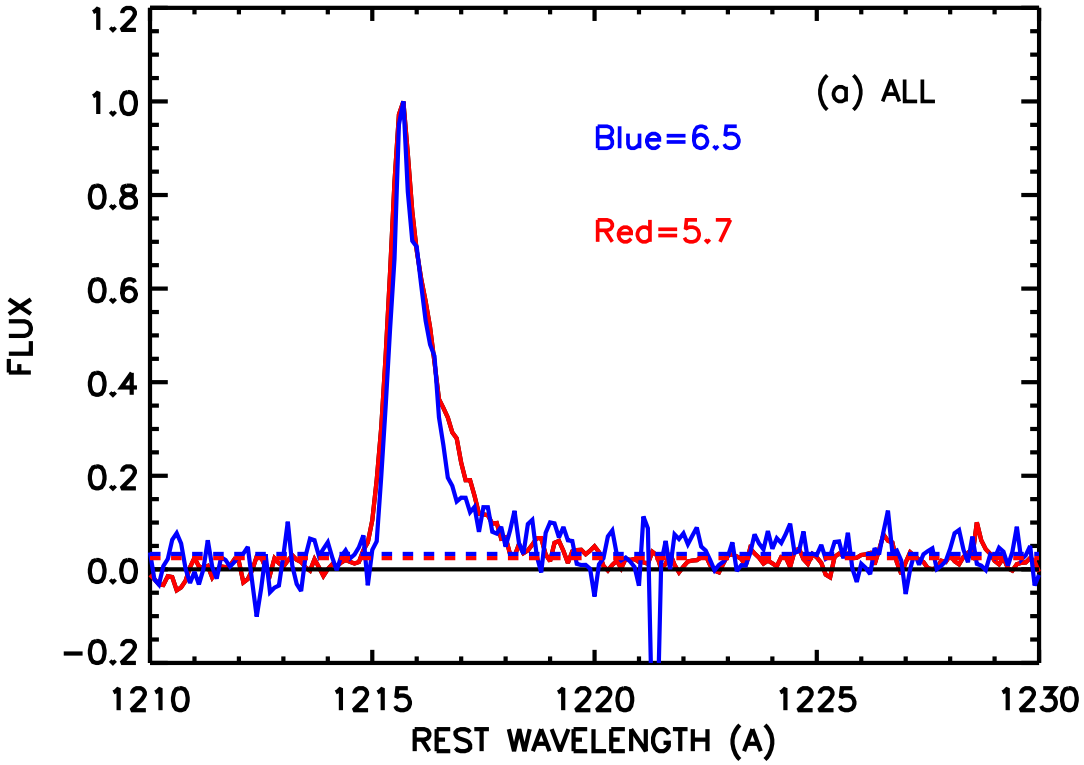

Figure 6.16: Stacked Lyα spectra of samples at $z = 5.7$ (red) and $z = 6.6$ (blue). *Figure taken from Hu et al. (2010)*.

at higher redshifts.

The **Wind** model, on the other hand, shows less evolution in the shape and width of the Lyα profiles over the redshift range studied. The most noticeable difference between the predicted profiles in both models is the difference in the position of the main peak. The peaks in the **Wind** model seems to be redshifted by ~ 3Å at all redshifts compared to the **Shell** model. In addition, at $z = 0.2$ the **Wind** model predicts a first backscattering peak that is much less prominent than in the **Shell** model. The FWHM of the line profiles are larger by a factor ~ 2 in the **Wind** model compared to the **Shell** model, except at $z = 0.2$, at which both models give comparable widths of the line profiles.

A qualitative comparison with available observational data seems to support the shape of the predictions of the **Shell** model. For example, figure 6.16 shows stacked Lyα profiles taken from Hu et al. (2010) at $z = 5.7$ and $z = 6.6$. The stacking is made over the good quality data, and consists of normalising the observed Lyα profiles to unity and then averag-

ing over the normalised spectra. The resulting stacked spectra agree qualitatively well with both models, but the width of the predicted stacked profiles with the **Shell** model seems to be in closer agreement to the observational data of Hu et al. (2010) than the **Wind** model. Nevertheless, it is not possible to rule out the predictions of any of the models on the basis of a qualitative analysis. A more detailed analysis, including the absorption of continuum emission and comparing to more observed Lyα profiles is needed, and it will be presented in a future work.

6.6.5 The Lyα Equivalent width distribution

The EW measures the strength of the line with respect to the continuum around it. We compute the EWs simply by taking the ratio of the predicted Lyα luminosity of galaxies and their continuum around the Lyα line computed by `GALFORM`, including attenuation by dust. Figure 6.17 shows a comparison of the EW distribution measured at different redshifts with the predictions from our outflow models. In order to present a fair comparison between the observational measurements and the model predictions, we impose the luminosity cut of the observed sample on the model predictions.

At $z = 0.2$, both outflow models are consistent with the EW distribution of the Cowie et al. (2010) sample. In detail, both models show little differences between them, and both decrease the abundance to higher values of EW extending to $EW_{\rm rf} \sim 300$Å, whereas the observational data shows a sharp break at around $EW_{\rm rf} \sim 150$Å.

At $z = 3.0$, we compare the model predictions with the samples at $z = 3.1$ and 3.7 from Ouchi et al. (2008), taking the Lyα limiting luminosity to be the same as the sample at $z = 3.1$. Both observational measurements seem to peak at around $EW_{\rm rf} \sim 100$Å and then decline until reaching a maximum value of $EW_{\rm rf} \sim (150, 250)$Å, for the $z = 3.1$ and $z = 3.7$ samples respectively. The EWs predicted from the models show a similar distribution. However, both models predict EWs reaching values up to $EW_{\rm rf} \sim 450$Å. The majority of the EWs predicted by both models are, as shown in figure 6.17, consistent with the values measured observationally.

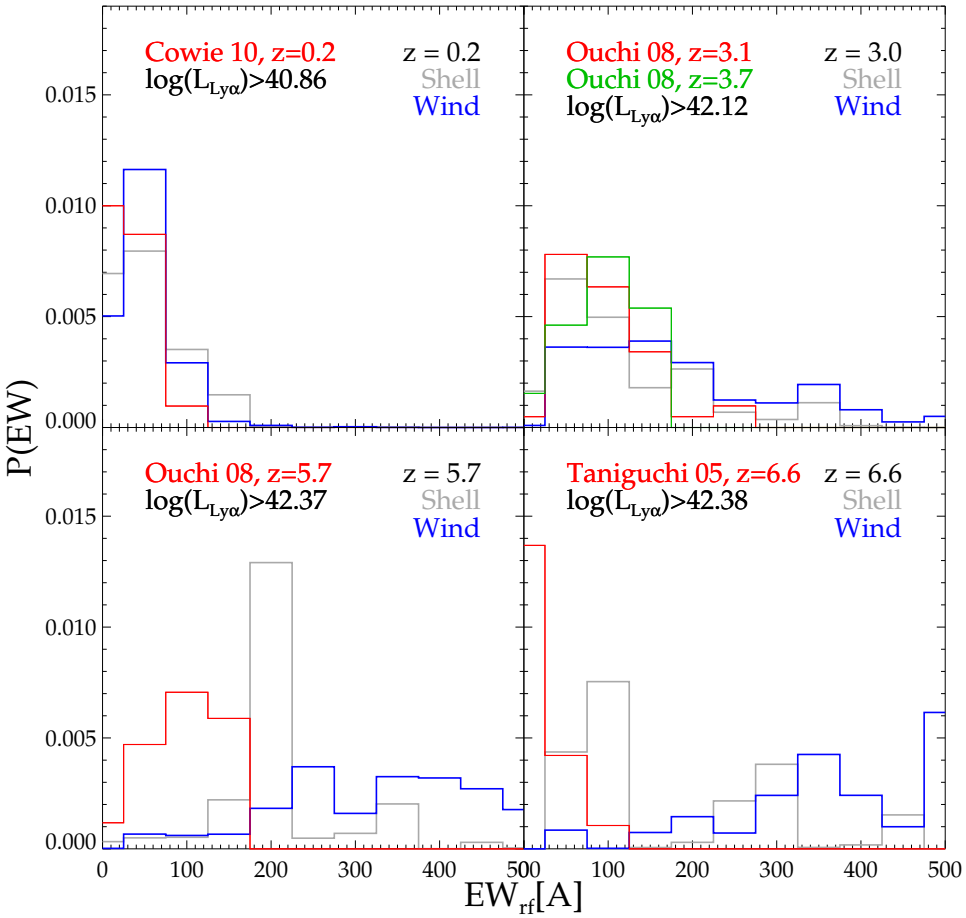

Figure 6.17: The equivalent width (rest-frame) distribution of Lyα emitters for $0.2 < z < 6.6$. Observational measurements are shown in red and green histograms. The **Wind** and **Shell** models histograms are shown in blue and gray respectively. Model galaxies selected to construct the EW distributions shown have a Lyα luminosity greater than the value shown in the labels, to reproduce the observational selection at each redshift.

The EWs from the outflow models at $z = 5.7$ and $z = 6.6$ are distributed in a completely different way to the observational samples of Ouchi et al. (2008) at $z = 5.7$ and Taniguchi et al. (2005) at $z = 6.6$. In both cases, the **Wind** model presents a flat distribution, without any well defined single peak. The **Shell** model, on the other hand, peaks strongly at around $EW_{\rm rf} \sim 200$Å for $z = 5.7$ and $EW_{\rm rf} \sim 100$Å at $z = 6.6$.

The disagreement in the EW distributions between the model predictions and the observations at $z > 5$ is difficult to explain from figure 6.17 alone. A possible explanation for this general disagreement is the predicted abundance of Lyα emitters at the bright end of the LF. The **Shell** model predicts more bright emitters than the observed data, and the **Wind** model does the opposite, as shown in figure 6.11. This difference could translate into the distribution of EWs shown in figure 6.17.

However, a more intrinsic difference is noticed when we study the relation between the EW and the Lyα luminosity, as shown in figure 6.18. This relation is basically determined by the dependence of the extinction of the continuum around the Lyα line for different Lyα luminosities.

Figure 6.18 confirms the good agreement between the model predictions and the observational data at $z = 0.2$, where both models seem to reproduce well the observational measurements. At $z = 3.0$, however, the observational data are found at luminosities much brighter than those predicted by the outflow models. The lack of bright Lyα emitters at $z = 3.0$ predicted by the outflow models occurs because galaxies with the highest intrinsic Lyα luminosities are found to have $f_{\rm esc} < 10^{-3}$, and thus they do not appear in the model predictions. The relation between the predicted escape fraction and the intrinsic Lyα luminosity is studied in more detail in the next section.

At $z = 5.7$ and $z = 6.6$, we find that both outflow models predict Lyα EWs higher than those measured. This is translated into the disagreement found in the EW distribution shown in figure 6.17 between the observational data and the outflow models predictions. In detail, the **Shell** model predict EWs that are only slightly above the measured values.

However the **Wind** model predicts EWs an order of magnitude or more above the measured EWs. The difference in the predicted values of the EWs is due to the escape fractions

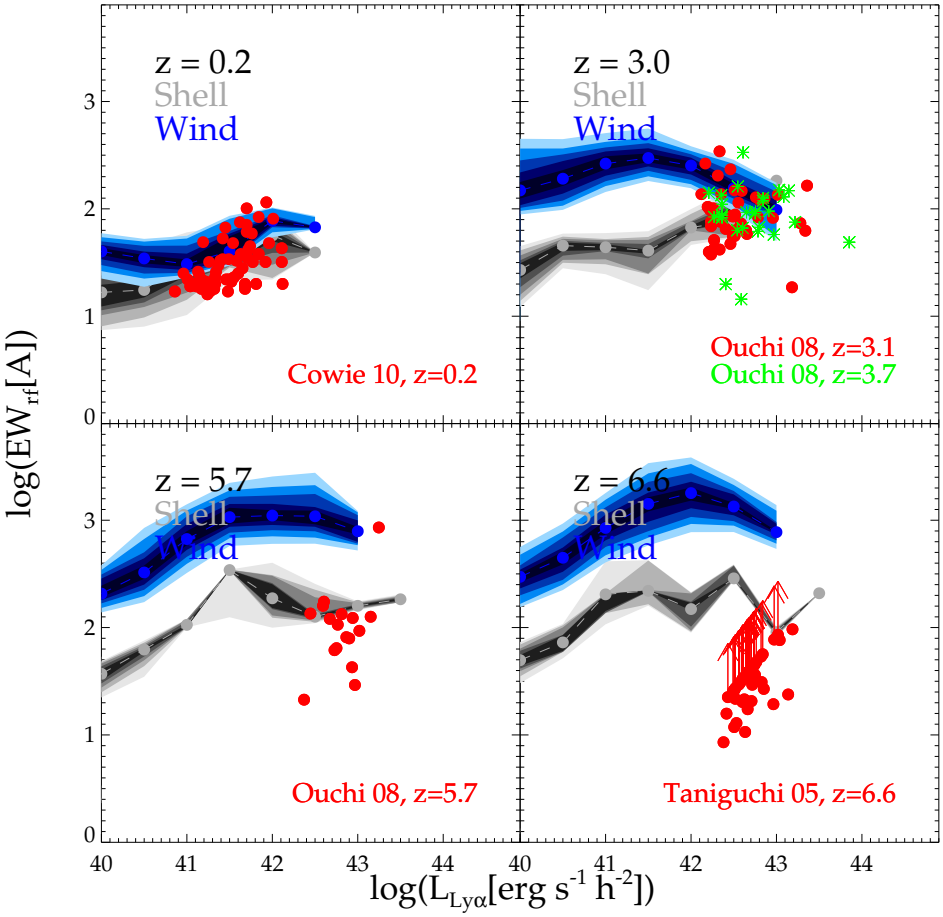

Figure 6.18: The rest-frame Lyα *EW* as a function of Lyα luminosity for different redshifts, as labelled on the plot. The same observational data shown in figure 6.17 is shown here with red circles and, in particular, with green asterisks for the data of Ouchi et al. (2008) at $z = 3.7$. The arrows show data with only lower limits on the measured *EW*, taken from Taniguchi et al. (2005) at $z = 6.6$. The model predictions for the **Shell** and **Wind** model are shown in grey and blue respectively. The coloured regions show the 45-55, 40-60, 35-65, 30-70 and 25-75 percentiles with the darkest to brightest colours, respectively.

predicted, which differ greatly at high luminosities.

Nevertheless, at $z = 6.6$, most of the EWs measured by Taniguchi et al. (2005) are lower limits only, so in reality the comparison between measured and predicted EWs could agree better at this redshift with improved observational measurements of EWs. Also, as discussed previously, the IGM is expected to play an important role attenuating the observed Lyα luminosities at these redshifts. This effect should also modify significantly the measured EWs.

6.7 The physical properties of Lyα emitters

In the following we study some intrinsic physical properties of Lyα emitters that cannot be observed directly, but they prove to be helpful to understand the physics that shapes the escape of Lyα photons.

The most fundamental prediction of our models is the Lyα escape fraction from galaxies, since all the properties studied in this chapter depend directly on f_{esc}. Therefore, we begin by studying the distribution of predicted f_{esc} from the two outflow models, shown in figure 6.19.

Perhaps not surprisingly, given the differences in the predictions from both models shown in the previous sections, the two models give completely different abundances of galaxies with a given escape fraction. Overall, it is surprising though that the number density of galaxies at a given f_{esc} does not seem to evolve dramatically throughout the redshift range $0.2 < z < 6.6$, considering that the properties of the outflows do evolve in redshift.

The **Wind** model predicts that the majority of galaxies have $f_{esc} \sim 1$. This is consistent with the shape of the predicted CLFs shown in figure 6.11, where low luminosity galaxies (the most abundant) showed Lyα luminosities very close to their intrinsic values, hence producing a very steep luminosity function.

The **Shell** model has a much flatter relation between the abundance of galaxies and their escape fractions. Galaxies with high escape fractions are slightly more abundant than galax-

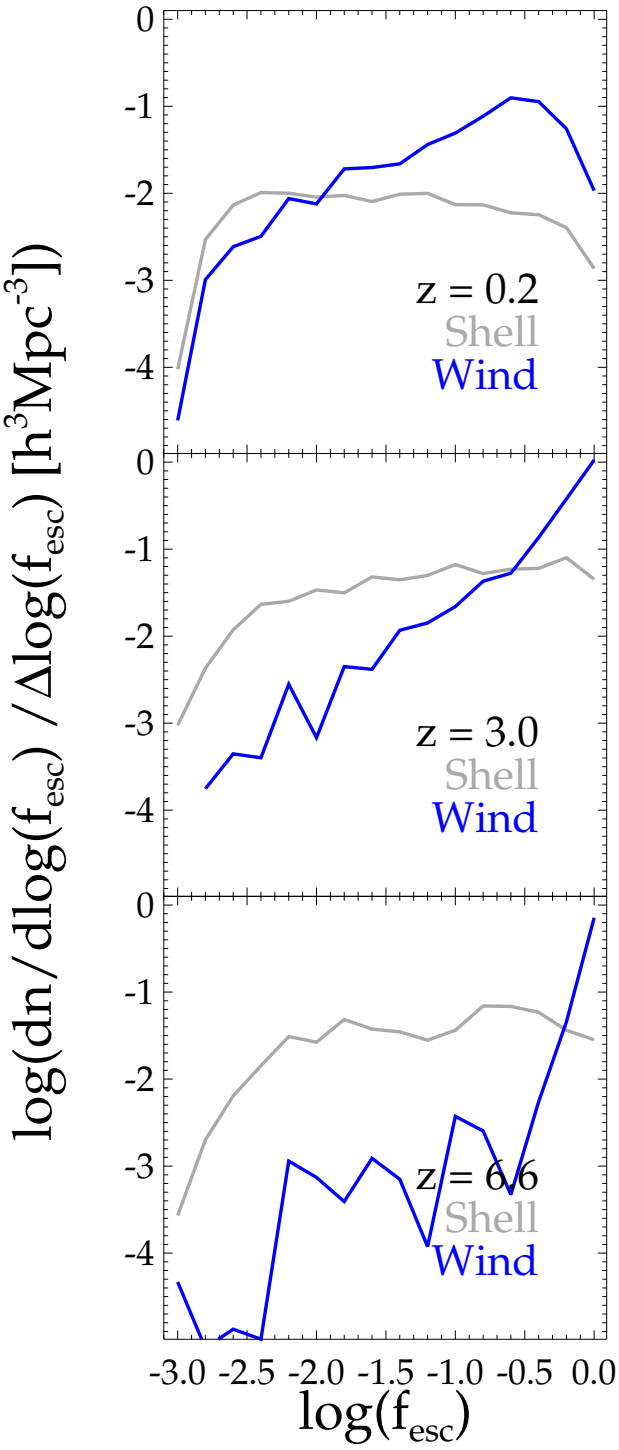

Figure 6.19: The f_{esc} distribution for the **Shell** (grey) and **Wind** (blue) models at $z = 0.2$ (top), $z = 3.0$ (middle) and $z = 6.6$ (bottom).

ies with low escape fractions. This correlation is erased for higher redshifts and instead the **Shell** model shows a weak preference towards higher escape fractions.

In order to understand the predicted f_{esc} of both models, and hence the Lyα luminosities, we now focus on the column density distributions, since as shown in figures 6.4,6.5 and 6.6, the value of the column density N_H plays an important role in shaping the Lyα line profile and the escape fraction f_{esc}, given its proportionality to the optical depth of scattering and absorption.

Figure 6.20 shows the predicted hydrogen column density as a function of Lyα luminosity. In the left panel we see how the predicted column densities correlate with the intrinsic Lyα luminosity at different redshifts. Overall in the **Wind** model the column densities increase towards higher $L_{Ly\alpha,0}$. This correlation is mainly driven by the relation between the mass ejection rate and the intrinsic Lyα luminosity. Thus, galaxies with intrinsic Lyα luminosities of $\log(L_{Ly\alpha}[\mathrm{erg\ s^{-1}\ h^{-2}}]) \sim 40$ are predicted to have outflows with column densities around $N_H \sim 10^{21}[\mathrm{cm^{-2}}]$, whereas galaxies with the brightest intrinsic Lyα luminosities have $N_H \sim 10^{23}[\mathrm{cm^{-2}}]$. This behaviour is common throughout the redshift range $0.2 < z < 6.6$, although in detail the correlation becomes noisier towards higher redshifts.

When studying the column density distribution in the **Wind** model as a function of the predicted Lyα luminosity (i.e. including the escape fraction computed for each galaxy) the relation changes significantly, becoming noisier at $z = 0.2$. The highest column densities are found in fainter galaxies (in Lyα luminosity), basically because when the medium is very optically thick then the escape fraction is very small, thus displacing galaxies with a high intrinsic Lyα luminosity to fainter luminosity bins. However, at higher redshifts the most luminous galaxies still host the outflows with the highest column densities, although now the median hydrogen column density for the brightest galaxies decreases to $N_H \sim 10^{22}[\mathrm{cm^{-2}}]$.

The **Shell** model, on the other hand, predicts a nearly constant column density slightly below $N_H \sim 10^{22}[\mathrm{cm^{-2}}]$ regardless of the intrinsic Lyα luminosity at $z = 0.2$. The higher cold gas mass of a bright Lyα emitters is compensated by the slight increase in their radius, as shown in figure 6.3, since the column densities predicted by the **Shell** model are more

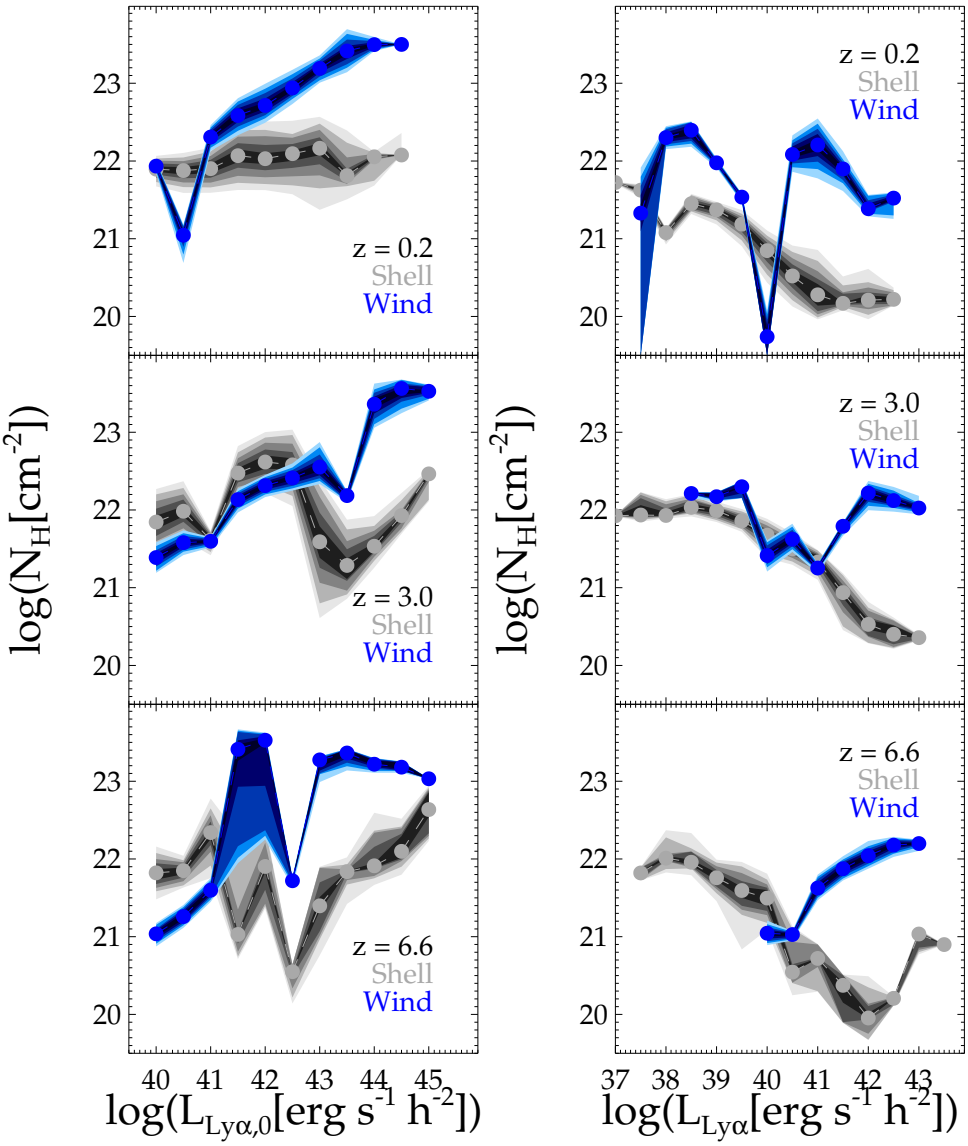

Figure 6.20: The column density of hydrogen for galaxies with different Lyα luminosities at redshifts $z = 0.2$ (top), $z = 3.0$ (middle) and $z = 6.6$ (bottom). The left column shows the column density as a function of intrinsic Lyα luminosity, whereas the right column shows the column densities as a function of the Lyα luminosity predicted by the outflow models. The predictions of the **Wind** model are shown in blue, and the **Shell** model is shown in grey. The circles show the median of the column densities for different luminosity bins. The coloured regions show different percentiles around the median covering from 10% to 50% of the data from the darkest to the brightest colours.

sensitive to the radius of galaxies than the **Wind** model. At higher redshifts the column densities fluctuate in a rather noisy fashion roughly between $N_H \sim 10^{21} - 10^{23}[\mathrm{cm}^{-2}]$. When using the predicted Lyα luminosity instead of the intrinsic, then the **Shell** model predicts a consistent anti-correlation between the hydrogen column densities and the Lyα luminosity at all redshifts studied, as shown in figure 6.20.

It is clear from figure 6.20 that Lyα emitter properties cannot be characterised only by the column density of the medium where photons travel alone. Instead of looking at the dependence of the escape fraction on each property of the outflow, we turn our attention to the dependence of the escape fraction on the intrinsic Lyα luminosity $L_{Ly\alpha,0}$, as shown in figure 6.21.

The intrinsic Lyα luminosity is a good tracer of the physical processes relevant to the escape of Lyα photons, since it is proportional to the star formation rate of galaxies, due to the fact that active star forming galaxies produce higher rates of Lyman continuum photons than quiescent galaxies. Hence, $L_{Ly\alpha,0}$ scales linearly with the mass ejection rate from supernova outflows, as shown in figure 6.3. The cold gas mass was also shown to scale linearly with $L_{Ly\alpha,0}$ for both quiescent and starburst galaxies.

Overall, figure 6.21 shows that both outflow models predict a decline in f_{esc} with increasing $L_{Ly\alpha,0}$. At $z = 0.2$ the **Shell** model predictions do not appear on the plot since this relation is dominated by galaxies which have $f_{esc} < 10^{-3}$ (see also figure 6.12).

The **Wind** model, on the other hand, predicts escape fractions near unity for low intrinsic Lyα luminosities. These galaxies are responsible for the steep slope of the predicted cumulative luminosity function (see figure 6.11), since their net luminosity is very close to the intrinsic one. For higher luminosities there is a sharp decrease in the predicted f_{esc} reaching the lowest possible escape fraction for the brightest galaxies. The **Shell** model also presents a sharp transition between higher and lower escape fractions for high redshifts. However, at a given intrinsic Lyα luminosity the escape fraction is generally lower in the **Shell** model than in the **Wind** model.

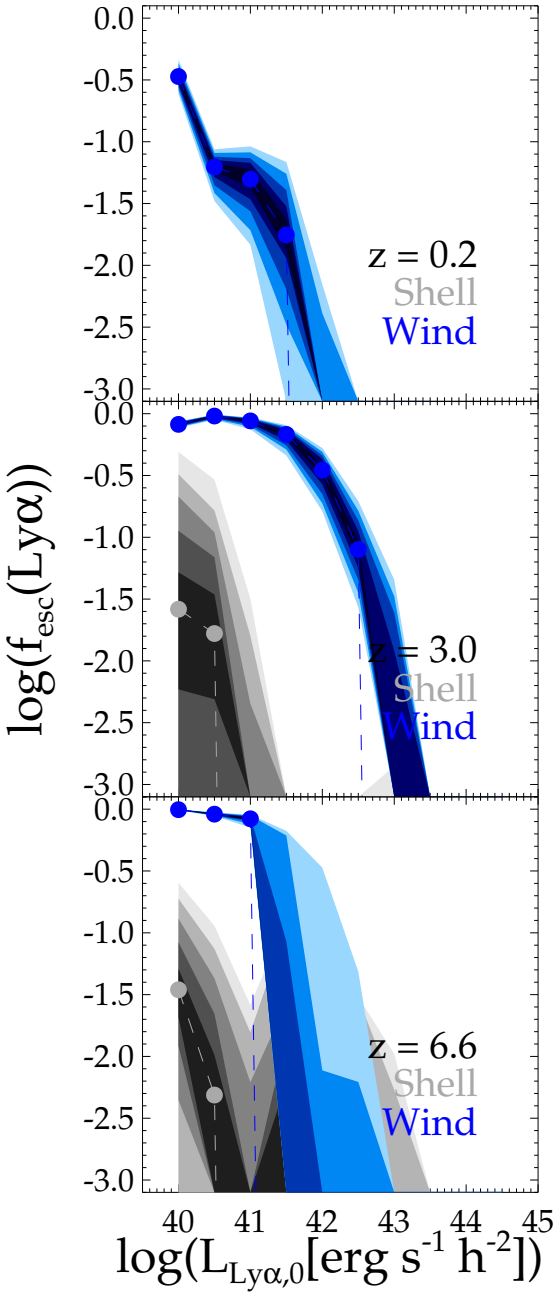

Figure 6.21: The escape fraction as a function of the intrinsic Lyα luminosity at $z = 0.2$ (top), $z = 3.0$ (middle) and $z = 6.6$ (bottom). The predictions of the **Wind** model are shown in blue, and the **Shell** model is shown in grey. The circles show the median of the escape fraction at different luminosity bins. The coloured regions show different percentiles around the median covering from 10% to 50% of the data from the darkest to the brightest colours.

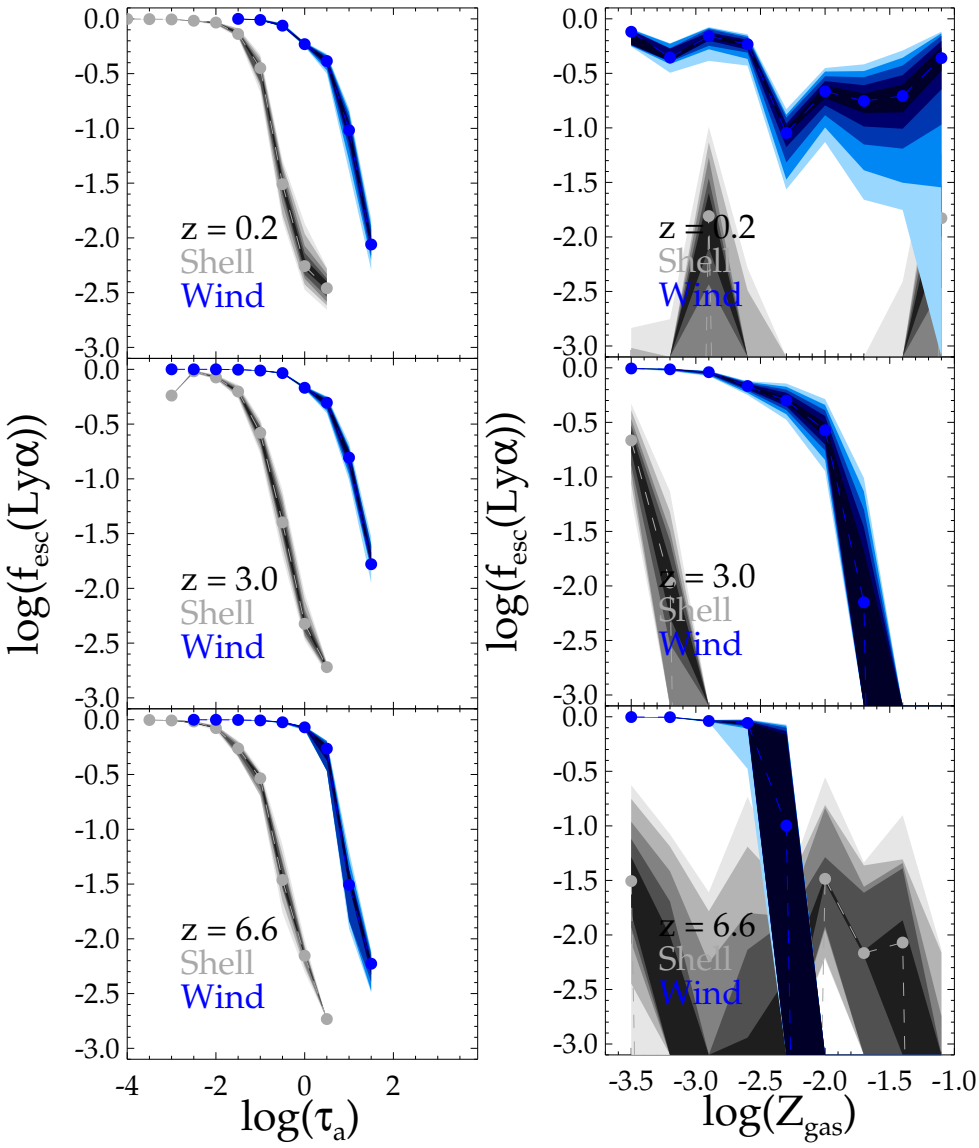

Figure 6.22: The escape fraction as a function of the optical depth of absorption τ_a (left column) and the metallicity of the gas in the outflow $Z_{\rm gas}$ (right column) at $z = 0.2$ (top), $z = 3.0$ (middle) and $z = 6.6$ (bottom). The coloured regions and circles represent the results from the **Shell** and **Wind** models, in the same way as in figure 6.21.

6.7.1 The role of dust

Another important factor in the modelling of the escape fraction of Lyα emitters is the presence of dust in the outflows. It is generally accepted that Lyα emitters are sensitive to even small amounts of dust. The effect of dust on the escape fraction has only been inferred from the extinction $E(B-V)$, as shown in figure 6.13 for $z = 0.2$.

Here we study a more fundamental dependence of f_{esc} on the dust content of the galaxy. Figure 6.22 shows a tight correlation between the escape fraction and the optical depth of absorption τ_a (see equation 6.10), and the more complex relation between the escape fraction and the metallicity of the gas in the outflow Z_{gas}.

As expected, the escape fraction anti-correlates with the optical depth of absorption. Galaxies with low values of τ_a have low amounts of dust in the outflow, and, thus most of the photons manage to escape regardless of the other physical properties of the medium they travel through. However, the point at which the escape fraction turns to low values is different for the two models: the **Shell** model is more sensitive to dust than the **Wind** model, for the same value of τ_a.

This conclusion is consistent with what we found in the comparison of f_{esc} with $E(B-V)$ (figure 6.13), and is a direct consequence of the larger number of scatterings photons must undergo in the **Shell** model compared to the **Wind** model, as we found in Section 6.4 (see figures 6.4, 6.5 and 6.6). It is important to emphasize that this is an intrinsic property of the outflow models, since configurations with the same physical properties are found to have different escape fractions depending on the outflow model we use.

Since the metallicity is an indicator of the the dust content of galaxies, perhaps naively we could expect the metallicity of the gas Z_{gas} to correlate somehow with the Lyα escape fraction. However, as figure 6.22 shows, the relation between f_{esc} and Z_{gas} does not resemble the one found for τ_a, except perhaps at $z = 3.0$. At $z = 0.2$, for example, the escape fraction is found to be nearly unity for both low and high metallicity outflows in the **Wind** model. Likewise, in the **Shell** model the escape fraction fluctuates between lower and higher values of f_{esc} over the whole metallicity range. At $z = 6.6$ both models predict an anti-correlation between f_{esc} and Z_{gas}, although the **Shell** model predicts a region of high escape fractions for intermediate metallicities. This is not really surprising since the amount of dust is given by a combination of the metallicity of galaxies and their column densities, which were found to vary significantly for different luminosities, as shown in figure 6.20.

6.7.2 Can Lyα emitters be used to trace the star formation rate?

Having in mind all the physical processes contributing to the way Lyα photons escape from galaxies, it is interesting to investigate whether the Lyα luminosity can be considered as a tracer of the star formation rate. Since Lyα is a hydrogen recombination line directly related to the production rate of Lyman continuum photons, we would expect the Lyα luminosity to correlate with the star formation rate. Figure 6.23 shows this relation in different redshifts.

As expected, the intrinsic Lyα luminosity shows a tight correlation with the SFR at all redshifts. The Lyα luminosities obtained when using the outflow models are, however, correlated to the SFR in a different way. For example, at $z = 0.2$, the **Wind** model predicts that the Lyα luminosity of galaxies has a very weak correlation with the SFR.

The dotted line in figure 6.23 shows where the median of the intrinsic Lyα luminosity would be placed if adding a constant escape fraction of $f_{esc} = 10^{-3}$ to it (which corresponds to the resolution limit of our outflow models), so it represents an upper limit under which the median and percentiles shown are dominated by galaxies with $f_{esc} < 10^{-3}$. As figure 6.23 shows, the samples on bins with high SFRs are dominated by galaxies with $f_{esc} < 10^{-3}$.

Since the **Shell** model predictions are dominated by galaxies with $f_{esc} < 10^{-3}$, as shown in figure 6.12, it is not possible to establish a correlation between the Lyα luminosity and the SFR, except at the lower bins of SFR. The **Wind** model, on the other hand, exhibits a tight correlation in the regime where the escape fractions are high, and thus the predicted Lyα luminosities are very close to the intrinsic ones. However, regardless of the redshift, the median luminosity falls rapidly below our resolution limit for galaxies with SFRs $\geq 1[M_{\odot} yr^{-1} h^{-1}]$.

In summary, whether there is a correlation between the observed Lyα luminosity and the star formation rate of galaxies or not depends entirely on the model assumptions, which is why it is not possible to determine a one-to-one correlation between both galaxy properties. Hence, our results stress the need to compute the escape fractions of Lyα emitters if the Lyα luminosity is to be thought of as a reliable indicator of the star formation rate.

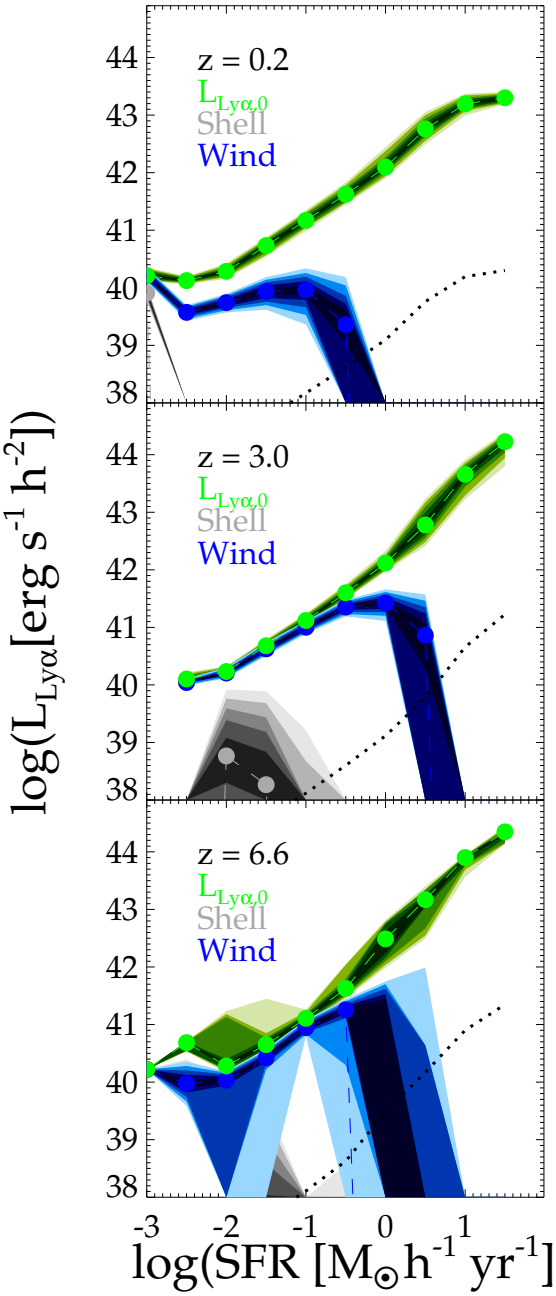

Figure 6.23: The Lyα luminosity as a function of the SFR for redshifts $z = 0.2$ (top), $z = 3.0$ (middle) and $z = 6.6$ (bottom). The green colour represents the intrinsic Lyα luminosity, whereas the grey and blue are the **Shell** and **Wind** models predictions respectively. The coloured regions show different percentiles around the median covering from 10% to 50% of the data from the darkest to the brightest colours. The dotted black line corresponds to a factor 10^{-3} of the median of the intrinsic Lyα luminosity, to illustrate the lowest luminosity the models can predict at each SFR bin.

6.8 Summary and conclusions

In this chapter we couple the galaxy properties obtained from the Baugh et al. (2005) version of the semianalytical model GALFORM to a Monte Carlo radiative transfer model of Lyα photons to study the properties of Lyα emitters in a cosmological context.

Based on the observational evidence of galactic outflows shaping the asymmetric Lyα line profiles, we developed two different outflow models, each one defined using the predicted properties of galaxies in GALFORM in a slightly different way. Our **Shell** model, which consists in a spherical expanding thin shell, has a column density N_H proportional to the cold gas mass in the ISM of galaxies. Our **Wind** model, on the other hand, consists of a spherical expanding wind with decreasing number density. The column density in the **Wind** model is related to the mass ejection rate from supernovae, computed by GALFORM as a mode of feedback of galaxies.

The different geometry and detailed number densities inside the outflows for each model produce significantly different Lyα profiles and escape fractions, as shown in figures 6.4, 6.5 and 6.6. Hence, the global properties of Lyα emitters depend strongly on the assumed model for the escape of Lyα photons.

We assessed the effect of the UV background in reducing the number density of neutral hydrogen in the **Wind** model, and found that although the UV background can have a significant effect in shaping the number density profiles, the overall Lyα luminosities do not suffer a dramatic difference, and thus, quantities like the luminosity function are unaltered when activating the UV background.

In order to compute the Lyα escape fraction for each galaxy in a GALFORM output we construct multidimensional grids covering the full range of parameter space in GALFORM galaxies. In this way we obtain f_{esc} by interpolating on the grid, thus obtaining the escape fractions in a much more efficient way than computing f_{esc} for each galaxy individually. Furthermore, to optimise the number of configurations needed to construct the grids, we reduce the number of parameters used by noticing that the quantities $C_{\rm wind} = \dot{M}_{\rm ej}/R_{\rm inn}$ and $C_{\rm shell} = M_{\rm gas}/R^2_{\rm out}$ can describe the escape fractions in the **Wind** and **Shell** models respectively without the need to specify $\dot{M}_{\rm ej}, R_{\rm inn}$ and $M_{\rm gas}, R_{\rm out}$ individually. The obtained escape fractions from the grids

are sufficiently accurate to allow a robust comparison with the observational data.

In order to choose the free parameters of each model, we attempt to find the best values to fit the observed cumulative luminosity function of Lyα emitters over the redshift range $0.2 < z < 6.6$. Perhaps disappointingly, we found that a single choice of parameters cannot reproduce the observed CLFs. The need to tune the free parameters of both models at each redshift studied is a strong suggestion that there might be other physical processes determining the escape of Lyα photons. Perhaps the assumption of Lyα photons escaping through an outflow of material is not correct at all, or instead other physical conditions are also playing an important role, like the presence of a multi-phase medium (our model assumes that gas and dust are distributed in the same way inside the outflow), the effect of a partially ionized IGM, or perhaps a more sophisticated outflow model is necessary, including a non-spherical geometry or velocity gradients inside the outflow.

When tuning the parameters to fit the observed CLFs, we notice that the value of f_R in both models increases for higher redshifts. The luminosity-weighted radius of galaxies, however, decreases rapidly for higher redshifts. This anti-correlation could suggest that there is a preferential range of sizes for the outflows over which the escape fractions have the value required to reproduce the observed abundance of Lyα emitters. Roughly speaking, when taking into account the value of f_R for each outflow model, we find that the inner radius of the outflows in the **Wind** model is of the order of $\sim$ 100pc for all redshifts, whereas in the **Shell** model the outer radius is found at a scale of $\sim$ 500pc. We plan to investigate this further in a future work.

Despite the above, both models show good agreement in the relation between the escape fractions and the extinction, measured as $E(B-V)$. The equivalent width distributions are consistent with the ones observed for $z = 0.2$ or $z = 3.0$, although both models fail to reproduce the observed EWs at $z \sim 6$.

The Lyα line profiles of both models, although different, are consistent with what has been observed by several authors. A more detailed comparison of stacked spectra, including the absorption of the continuum around the Lyα line, could help understanding better which

one of our models reproduce the Lyα line profiles better, and predict which physical properties can be inferred from the observed profiles. This is a natural extension of the present work, and will be tackled in the future.

Perhaps the main differences in the predictions for Lyα emitters from the **Shell** and **Wind** models is due to the abundance of galaxies with a given escape fraction. The **Wind** model predicts that most of the galaxies have escape fractions close to 1, whereas in the **Shell** model there is no preferred escape fraction for galaxies. The contrast between these two behaviours is translated into the dependence of Lyα luminosities and the physical properties of galaxies.

For example, the column densities in the outflows for a given Lyα luminosity are found to be different between the two models. Furthermore, even with the same column density, it was shown that each model predicts a different Lyα profile and escape fraction, since the number of scatterings differs between the two models.

We found that the metallicity does not correlate with the escape fractions in a simple way, since the amount of dust in the outflows depend also in the column density of hydrogen in our models. However, even at the same optical depth of absorption, both models predict a different escape fraction, with the **Shell** model being more sensitive to dust than the **Wind** model.

Finally, we showed that the observed luminosity cannot be easily interpreted in terms of the star formation rate of galaxies. Depending on the range of SFRs and redshift, the Lyα luminosity could be anti-correlated or positively correlated to the SFR, varying according to the way the escape fraction is related to the intrinsic Lyα luminosity.

The models presented in this chapter for the emission of Lyα in galaxies represents our first attempt towards a detailed understanding of the physical properties of these galaxies. Despite the further investigation that is clearly needed to refine the model predictions, we have shown the potential that the combination of a Monte Carlo radiative transfer model of Lyα photons and a semianalytical galaxy formation model has to understand the properties of these high redshift galaxies. With the advent of large observational campaigns in the forth-

coming years focusing in detecting Lyα emitters at high redshifts, a physical understanding of these galaxies will enable us to improve our knowledge of galaxy formation and evolution, particularly in the high redshift Universe.

Chapter 7 | *Conclusions*

Throughout this thesis we have studied the nature of emission-line galaxies from a cosmological perspective, with two goals in mind: first, to understand what can emission line galaxies tell us about galaxy formation and evolution across cosmic time; and second, to assess emission line galaxies as a cosmological tool to probe dark energy.

Many current and forthcoming observational surveys are designed to trace the large scale structure of the Universe with emission-line galaxies (e.g. Nilsson et al., 2007a; Hill et al., 2008; Cimatti et al., 2009; Blake et al., 2009). As reviewed in Chapter 1, a correct interpretation of the outcome of those surveys can only be achieved with a detailed understanding of the properties of emission-line galaxies.

To accomplish both tasks we make use of the GALFORM semianalytical model of galaxy formation, which is described in detail in Chapter 2. Semianalytical models are ideal for our purposes due to their statistical robustness (since samples with a large number of galaxies can be generated at any desired redshift) and the efficient parametrisation of the physical processes governing galaxy formation and evolution, which allows us to make predictions for emission-line galaxies in cosmological scales.

We start in Chapter 3 by studying the clustering of Lyα emitters. We tackle the uncertainty on the escape fraction of Lyα photons by assuming that is constant at all redshifts and has a value equal to 0.02. This oversimplification reproduces, surprisingly, the abundance and clustering of this galaxy population remarkably well when comparing to observational measurements. The clustering properties from our model predict a weak dependence of the clustering strength on Lyα luminosity, except for the brightest galaxies. At the present day, typical Lyα emitters are less clustered than the dark matter. However, for higher redshifts, the effective bias of Lyα emitters is found to evolve strongly with redshift, although weakly

with luminosity.

Furthermore, many current and future Lyα surveys attempt to measure and characterise the clustering of Lyα emitters using samples of galaxies that are not large enough in a cosmological sense, i.e. they suffer from a significant cosmic variance. Hence, we use our model to predict the expected variance in a selection of current and forthcoming surveys by constructing mock catalogues.

We expect our predictions to encourage the design of surveys of emission line galaxies covering larger solid angles when clustering measurements are proposed to be undertaken, especially for high redshift galaxies.

In Chapter 4 we then focus our attention on Hα emitters and attempt to assess their performance at probing dark energy from a large redshift survey. Basically, there are two possible configurations being considered in the design of space missions to survey galaxies in the whole extragalactic sky. The first one makes use of slitless spectroscopy targetting Hα emission, whereas the second one relies on slit spectroscopy of a magnitude selected sample of galaxies in the H-band.

We find Hα emitters to be weakly clustered compared to H-band galaxies, mainly because the former do not trace the highest density peaks of the dark matter distribution, since the cores of galaxy clusters are dominated by massive and passive galaxies, and Hα emission traces active star forming galaxies instead. The H-band, on the other hand, is found to trace the massive structures. Nevertheless, for this reason Hα emitters are less affected by random galaxy motions giving rise to fingers of god in redshift space, thus they are potentially better for estimating cosmological redshift space distortions.

An analysis of the variance in the power spectrum for typical galaxy selections show that an Hα emitter survey can be competitive with an H-band selected survey provided that Hα emitters are selected down to faint enough fluxes. We predict the equation of state parameter of dark energy w_{DE} to be measured with an error $< 1\%$ for such cases.

In contrast to the Hα emission, Lyα is a resonant line with a large cross section of scattering, which means that photons typically experience much larger path lengths before escaping

from a galaxy. This makes Lyα photons very sensitive to the presence of dust, in a way that cannot be described by a simple extinction curve.

On the other hand, the increasing relevance of Lyα emitters as tracers of very high redshift galaxies and the many cosmological applications they can be used for motivates our interest in developing a more physical model of the luminosities of Lyα emitters than the one shown in Chapter 3.

The only robust way to understand the physics of the escape of Lyα photons is through a detail modelling of the radiative transfer photons experience when crossing an HI region. Therefore, we start by developing a Monte Carlo radiative transfer code of Lyα photons in Chapter 5. In summary, photons are generated and followed individually in a Monte Carlo algorithm through each interaction with either a hydrogen atom or a dust grain. Dust grains have the ability to scatter or absorb the photon, whereas hydrogen atoms scatter them and can change their frequency. The code follows all the interactions the photon experiences until it either escapes or is absorbed by dust. The whole process is repeated several times until we find convergence of the properties we are interested in. This usually requires from 10^3 to 10^5 photons. The output of the code is the frequency distribution of the photons that managed to escape from a given HI region, defined arbitrarily by the user. Our code is similar in fashion to other codes found in the literature, which simplifies the task of testing it.

Finally, in Chapter 6 we couple this Monte Carlo radiative transfer code to the galaxies predicted by `GALFORM`. To achieve this we investigated two different outflow models for the escape of Lyα photons from galaxies.

We find that the predicted escape fraction of galaxies can vary greatly. However, the value of f_{esc} predicted depends strongly on the parameters and geometry of the outflow model assumed.

The comparison with observational data forces us to tune the free parameters in our outflow models to different values for different redshifts to reproduce the observed luminosity functions of Lyα emitters at $0.2 < z < 6.6$. We believe this suggests that the escape of Lyα photons is in reality more complicated than what we are modelling. Nevertheless, we carry on and study the predictions from our models, i.e. the dependence of the escape fraction

with the intrinsic Lyα luminosity, metallicity, extinction, etc.

Furthermore, we find that our outflow models predict consistent line profiles with those found in the literature for several redshifts. The detailed shape is, however, model dependent, which opens the possibility of favouring one model over the other with a detailed observational comparison.

Finally, we predict that the observed Lyα emission does not correlate in a simple way with the star formation rate of galaxies, due to the complicated dependence of the escape fraction with the physical properties of the outflows.

7.1 Future directions

Many of the topics we investigated throughout this thesis require further development. In particular, we plan to address the following extensions of the work shown here in the future:

- The clustering of Lyα emitters is becoming a popular probe of the large scale structure of the Universe at high redshifts. With the advent of large Lyα surveys such as HETDEX (Blanc et al., 2007; Hill et al., 2008), it will be interesting to assess the accuracy of Lyα clustering predictions. So far, as shown in Chapter 3, we have focused only on general clustering properties and their performance for small narrow band surveys. Therefore, we plan to extend the current work to study, for example, the accuracy in measuring the BAO scale with Lyα emitters. Further improvement over the work shown in Chapter 3 can be achieved by using the more detailed model for Lyα emitters we introduced in Chapter 6.

- Hα emitters can be regarded as *clean* tracers of the star formation history in the Universe, since their luminosity scales directly with the production rate of Lyman continuum photons, once dust attenuation is taken into account. The model presented in Chapter 4 represents our first attempt to model this galaxy population. A natural extension of the current work involves a careful treatment of the extinction of Hα. We found evidence that the Baugh et al. (2005) overestimates the intrinsic Lyman continuum luminosities, which in turn affects all our predictions. In addition, the production

rate of Lyman continuum photons can be directly inferred from the unattenuated Hα luminosity function. A non-zero escape fraction of Lyman continuum photons plus a calculation of their intrinsic extinction when crossing HII clouds before reaching the ISM could, for example, explain the offset between the predicted unattenuated Hα LF and the observed one.

A model reproducing the observed abundances of Hα emitters could potentially characterise accurately the star formation history of the Universe, allowing us to get a direct insight into the physical processes which determine the star formation of galaxies, and hence, galaxy formation and evolution itself.

- Our most immediate future work planned is to investigate further the predictions of the properties of Lyα emitters shown in Chapter 6. Several areas which require further research include:

  - There is a possible preferred size of the outflows suggested by the evolution of the parameter f_R with redshift in both models. Outflows over which Lyα photons escape could well happen locally in a galaxy, and thus they may not be related to the size of the galactic disk or bulge at all.

  - The shape of the predicted Lyα line profiles could potentially help to favour one outflow model over the other when comparing in detail with observational data. In addition, quantitative measures of the properties of the Lyα line profiles, like the skewness of the line, the FWHM, and the offset of the (single or multiple) peaks from the line centre, can be assessed and used to interpret current observations in terms of physical properties.

  - When studying galaxies at $z > 6$, it is generally believed that a partially ionized IGM plays an important role at attenuating and shaping the observed Lyα line (Dijkstra et al., 2007; Dayal et al., 2010b; Laursen et al., 2010). Therefore, a complete physical model of Lyα emitters must include a model for the attenuation of the emission by the neutral fraction of the IGM. It is expected, however, that the Lyα properties would not be greatly affected for galaxies with $z < 6$.

In summary, there is still plenty of research to be done in developing a complete understanding of galaxy formation traced by emission-line galaxies. The theoretical interpretation of the outcome of current surveys is still limited and sparse, but the increasing relevance of the subject urges the need to develop further galaxy formation theories to interpret correctly the current and forthcoming observational data; and, most importantly, to improve our understanding of galaxy formation and evolution.

Bibliography

G. O. Abell. ApJS, 3:211–+, May 1958.

T. F. Adams. ApJ, 168:575–+, September 1971.

T. F. Adams. ApJ, 174:439–+, June 1972.

S.-H. Ahn. *Journal of Korean Astronomical Society*, 36:145–148, September 2003.

S.-H. Ahn. ApJ, 601:L25–L28, January 2004.

M. Ajiki, Y. Taniguchi, S. S. Fujita, Y. Shioya, T. Nagao, T. Murayama, S. Yamada, K. Umeda, and Y. Komiyama. AJ, 126:2091–2107, November 2003.

A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu, J. Huth, M. Kamionkowski, E. W. Kolb, L. Knox, J. C. Mather, S. Staggs, and N. B. Suntzeff. *ArXiv Astrophysics e-prints*, September 2006.

R. E. Angulo, C. M. Baugh, C. S. Frenk, and C. G. Lacey. MNRAS, 383:755–776, January 2008a.

R. E. Angulo, C. M. Baugh, and C. G. Lacey. MNRAS, 387:921–932, June 2008b.

H. Atek, D. Kunth, M. Hayes, G. Östlin, and J. M. Mas-Hesse. A&A, 488:491–509, September 2008.

H. Atek, D. Kunth, D. Schaerer, M. Hayes, J. M. Deharveng, G. Östlin, and J. M. Mas-Hesse. A&A, 506:L1–L4, November 2009.

L. W. Avery and L. L. House. ApJ, 152:493–+, May 1968.

L. A. Barnes and M. G. Haehnelt. MNRAS, 403:870–885, April 2010.

E. J. Barton, R. Davé, J.-D. T. Smith, C. Papovich, L. Hernquist, and V. Springel. ApJ, 604:L1–L4, March 2004.

C. M. Baugh. *Reports on Progress in Physics*, 69:3101–3156, December 2006.

C. M. Baugh. *Royal Society of London Philosophical Transactions Series A*, 366:4381–4392, December 2008.

C. M. Baugh, J. P. Gardner, C. S. Frenk, and R. M. Sharples. MNRAS, 283:L15–L19, November 1996.

C. M. Baugh, A. J. Benson, S. Cole, C. S. Frenk, and C. G. Lacey. MNRAS, 305:L21–L25, May 1999.

C. M. Baugh, C. G. Lacey, C. S. Frenk, G. L. Granato, L. Silva, A. Bressan, A. J. Benson, and S. Cole. MNRAS, 356: 1191–1200, January 2005.

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

A. J. Benson and R. Bower. *ArXiv e-prints*, April 2010.

A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh, and C. G. Lacey. MNRAS, 311:793–808, February 2000.

A. J. Benson, F. R. Pearce, C. S. Frenk, C. M. Baugh, and A. Jenkins. MNRAS, 320:261–280, January 2001.

A. J. Benson, C. G. Lacey, C. M. Baugh, S. Cole, and C. S. Frenk. MNRAS, 333:156–176, June 2002.

A. J. Benson, R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole. ApJ, 599:38–49, December 2003.

A. A. Berlind and D. H. Weinberg. ApJ, 575:587–616, August 2002.

A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, A. Jenkins, N. Katz, and C. G. Lacey. ApJ, 593:1–25, August 2003.

C. Blake, R. J. Jurek, S. Brough, M. Colless, W. Couch, S. Croom, T. Davis, M. J. Drinkwater, D. Forbes, K. Glazebrook, B. Madore, C. Martin, K. Pimbblet, G. B. Poole, M. Pracy, R. Sharp, T. Small, and D. Woods. MNRAS, 395:240–254, May 2009.

G. A. Blanc, J. Adams, K. Gebhardt, G. Hill, P. MacQueen, N. Drory, R. Koehler, J. Murphy, A. Fry, and HETDEX Collaboration. In *Bulletin of the American Astronomical Society*, volume 38 of *Bulletin of the American Astronomical Society*, pages 822–+, December 2007.

G. R. Blumenthal, S. M. Faber, R. Flores, and J. R. Primack. ApJ, 301:27–34, February 1986.

J. S. Bolton and M. G. Haehnelt. MNRAS, 382:325–341, November 2007.

J. R. Bond and G. Efstathiou. ApJ, 285:L45–L48, October 1984.

J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. ApJ, 379:440–460, October 1991.

R. G. Bower. MNRAS, 248:332–352, January 1991.

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. MNRAS, 370:645–655, August 2006.

G. Bruzual and S. Charlot. MNRAS, 344:1000–1028, October 2003.

R. R. Caldwell and M. Kamionkowski. *Annual Review of Nuclear and Particle Science*, 59:397–429, November 2009.

R. Cannon, M. Drinkwater, A. Edge, D. Eisenstein, R. Nichol, P. Outram, K. Pimbblet, R. de Propris, I. Roseboom, D. Wake, P. Allen, J. Bland-Hawthorn, T. Bridges, D. Carson, K. Chiu, M. Colless, W. Couch, S. Croom, S. Driver, S. Fine, P. Hewett, J. Loveday, N. Ross, E. M. Sadler, T. Shanks, R. Sharp, J. A. Smith, C. Stoughton, P. Weilbacher, R. J. Brunner, A. Meiksin, and D. P. Schneider. MNRAS, 372:425–442, October 2006.

J. A. Cardelli, G. C. Clayton, and J. S. Mathis. ApJ, 345:245–256, October 1989.

R. Cen. ApJS, 78:341–364, February 1992.

H.-W. Chen, P. J. McCarthy, R. O. Marzke, J. Wilson, R. G. Carlberg, A. E. Firth, S. E. Persson, C. N. Sabbey, J. R. Lewis, R. G. McMahon, O. Lahav, R. S. Ellis, P. Martini, R. G. Abraham, A. Oemler, D. C. Murphy, R. S. Somerville, M. G. Beckett, and C. D. Mackay. ApJ, 570:54–74, May 2002.

A. Cimatti, M. Robberto, C. Baugh, S. V. W. Beckwith, R. Content, E. Daddi, G. De Lucia, B. Garilli, L. Guzzo, G. Kauffmann, M. Lehnert, D. Maccagni, A. Martínez-Sansigre, F. Pasian, I. N. Reid, P. Rosati, R. Salvaterra, M. Stiavelli, Y. Wang, M. Zapatero Osorio, M. Balcells, M. Bersanelli, F. Bertoldi, J. Blaizot, D. Bottini, R. Bower, A. Bulgarelli, A. Burgasser, C. Burigana, R. C. Butler, S. Casertano, B. Ciardi, M. Cirasuolo, M. Clampin, S. Cole, A. Comastri, S. Cristiani, J.-G. Cuby, F. Cuttaia, A. de Rosa, A. D. Sanchez, M. di Capua, J. Dunlop, X. Fan, A. Ferrara, F. Finelli, A. Franceschini, M. Franx, P. Franzetti, C. Frenk, J. P. Gardner, F. Gianotti, R. Grange, C. Gruppioni, A. Gruppuso, F. Hammer, L. Hillenbrand, A. Jacobsen, M. Jarvis, R. Kennicutt, R. Kimble, M. Kriek, J. Kurk, J.-P. Kneib, O. Le Fevre, D. Macchetto, J. MacKenty, P. Madau, M. Magliocchetti, D. Maino, N. Mandolesi, N. Masetti, R. McLure, A. Mennella, M. Meyer, M. Mignoli, B. Mobasher, E. Molinari, G. Morgante, S. Morris, L. Nicastro, E. Oliva, P. Padovani, E. Palazzi, F. Paresce, A. Perez Garrido, E. Pian, L. Popa, M. Postman, L. Pozzetti, J. Rayner, R. Rebolo, A. Renzini, H. Röttgering, E. Schinnerer, M. Scodeggio, M. Saisse, T. Shanks, A. Shapley, R. Sharples, H. Shea, J. Silk, I. Smail, P. Spanó, J. Steinacker, L. Stringhetti, A. Szalay, L. Tresse, M. Trifoglio, M. Urry, L. Valenziano, F. Villa, I. Villo Perez, F. Walter, M. Ward, R. White, S. White, E. Wright, R. Wyse, G. Zamorani, A. Zacchei, W. W. Zeilinger, and F. Zerbi. *Experimental Astronomy*, 23:39–66, March 2009.

M. Cirasuolo, R. J. McLure, J. S. Dunlop, O. Almaini, S. Foucaud, and C. Simpson. MNRAS, 401:1166–1176, January 2010.

S. Cole and N. Kaiser. MNRAS, 237:1127–1146, April 1989.

S. Cole and C. Lacey. MNRAS, 281:716–+, July 1996.

S. Cole, K. B. Fisher, and D. H. Weinberg. MNRAS, 267:785–+, April 1994.

S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk. MNRAS, 319:168–204, November 2000.

M. Colless, B. A. Peterson, C. Jackson, J. A. Peacock, S. Cole, P. Norberg, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins, W. Couch, N. Cross, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, W. Sutherland, and K. Taylor. *ArXiv Astrophysics e-prints*, June 2003.

A. Cooray and R. Sheth. Phys. Rep., 372:1–129, December 2002.

L. L. Cowie and E. M. Hu. AJ, 115:1319–1328, April 1998.

L. L. Cowie, A. J. Barger, and E. M. Hu. ApJ, 711:928–958, March 2010.

D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. MNRAS, 365:11–28, January 2006.

E. Daddi, A. Cimatti, L. Pozzetti, H. Hoekstra, H. J. A. Röttgering, A. Renzini, G. Zamorani, and F. Mannucci. A&A, 361:535–549, September 2000.

M. Davis and P. J. E. Peebles. ApJ, 267:465–482, April 1983.

M. Davis, J. Huchra, D. W. Latham, and J. Tonry. ApJ, 253:423–445, February 1982.

S. Dawson, J. E. Rhoads, S. Malhotra, D. Stern, J. Wang, A. Dey, H. Spinrad, and B. T. Jannuzi. ApJ, 671: 1227–1240, December 2007.

P. Dayal, A. Ferrara, and A. Saro. MNRAS, 402:1449–1457, March 2010a.

P. Dayal, A. Maselli, and A. Ferrara. *ArXiv e-prints*, February 2010b.

J.-M. Deharveng, T. Small, T. A. Barlow, C. Péroux, B. Milliard, P. G. Friedman, D. C. Martin, P. Morrissey, D. Schiminovich, K. Forster, M. Seibert, T. K. Wyder, L. Bianchi, J. Donas, T. M. Heckman, Y.-W. Lee, B. F. Madore, S. G. Neff, R. M. Rich, A. S. Szalay, B. Y. Welsh, and S. K. Yi. ApJ, 680:1072–1082, June 2008.

A. Dekel and J. Silk. ApJ, 303:39–55, April 1986.

M. Dijkstra and A. Loeb. MNRAS, 400:1109–1120, December 2009.

M. Dijkstra, Z. Haiman, and M. Spaans. ApJ, 649:14–36, September 2006.

M. Dijkstra, A. Lidz, and J. S. B. Wyithe. MNRAS, 377:1175–1186, May 2007.

A. Dotter, B. Chaboyer, J. W. Ferguson, H.-c. Lee, G. Worthey, D. Jevremović, and E. Baron. ApJ, 666:403–412, September 2007.

J. Dunkley, E. Komatsu, M. R. Nolta, D. N. Spergel, D. Larson, G. Hinshaw, L. Page, C. L. Bennett, B. Gold, N. Jarosik, J. L. Weiland, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. ApJS, 180:306–329, February 2009.

G. Efstathiou, G. Lake, and J. Negroponte. MNRAS, 199:1069–1088, June 1982.

G. Efstathiou, W. J. Sutherland, and S. J. Maddox. Nature, 348:705–707, December 1990.

G. Efstathiou, S. Moody, J. A. Peacock, W. J. Percival, C. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. de Propris, S. P. Driver, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, P. Norberg, B. A. Peterson, W. Sutherland, and K. Taylor. MNRAS, 330:L29–L35, February 2002.

D. J. Eisenstein and W. Hu. ApJ, 496:605–+, March 1998.

D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York. ApJ, 633:560–574, November 2005.

V. R. Eke, J. F. Navarro, and C. S. Frenk. ApJ, 503:569–+, August 1998.

R. Ellis, M. R. Santos, J.-P. Kneib, and K. Kuijken. ApJ, 560:L119–L122, October 2001.

N. Fanidakis, C. M. Baugh, A. J. Benson, R. G. Bower, S. Cole, C. Done, and C. S. Frenk. *ArXiv e-prints*, November 2009.

M. A. Fardal, N. Katz, J. P. Gardner, L. Hernquist, D. H. Weinberg, and R. Davé. ApJ, 562:605–617, December 2001.

A. Ferrara, S. Bianchi, A. Cimatti, and C. Giovanardi. ApJS, 123:437–445, August 1999.

G. B. Field. ApJ, 129:551–+, May 1959.

A. E. Firth, R. S. Somerville, R. G. McMahon, O. Lahav, R. S. Ellis, C. N. Sabbey, P. J. McCarthy, H.-W. Chen, R. O. Marzke, J. Wilson, R. G. Abraham, M. G. Beckett, R. G. Carlberg, J. R. Lewis, C. D. Mackay, D. C. Murphy, A. E. Oemler, and S. E. Persson. MNRAS, 332:617–646, May 2002.

E. L. Fitzpatrick and D. Massa. ApJ, 663:320–341, July 2007.

H. Francke. *New Astronomical Reviews*, 53:47–49, June 2009.

W. J. Frith, N. Metcalfe, and T. Shanks. MNRAS, 371:1601–1609, October 2006.

J. N. Fry and P. J. E. Peebles. ApJ, 221:19–33, April 1978.

J. N. Fry and P. J. E. Peebles. ApJ, 238:785–792, June 1980.

S. S. Fujita, M. Ajiki, Y. Shioya, T. Nagao, T. Murayama, Y. Taniguchi, K. Umeda, S. Yamada, M. Yagi, S. Okamura, and Y. Komiyama. ApJ, 586:L115–L118, April 2003.

S. R. Furlanetto, J. Schaye, V. Springel, and L. Hernquist. ApJ, 622:7–27, March 2005.

J. Gallego, J. Zamorano, A. Aragon-Salamanca, and M. Rego. ApJ, 455:L1+, December 1995.

L. Gao, V. Springel, and S. D. M. White. MNRAS, 363:L66–L70, October 2005.

E. Gawiser, P. G. van Dokkum, D. Herrera, J. Maza, F. J. Castander, L. Infante, P. Lira, R. Quadri, R. Toner, E. Treister, C. M. Urry, M. Altmann, R. Assef, D. Christlein, P. S. Coppi, M. F. Durán, M. Franx, G. Galaz, L. Huerta, C. Liu, S. López, R. Méndez, D. C. Moore, M. Rubio, M. T. Ruiz, S. Toft, and S. K. Yi. ApJS, 162: 1–19, January 2006.

E. Gawiser, H. Francke, K. Lai, K. Schawinski, C. Gronwall, R. Ciardullo, R. Quadri, A. Orsi, L. F. Barrientos, G. A. Blanc, G. Fazio, J. J. Feldmeier, J.-s. Huang, L. Infante, P. Lira, N. Padilla, E. N. Taylor, E. Treister, C. M. Urry, P. G. van Dokkum, and S. N. Virani. ApJ, 671:278–284, December 2007.

J. E. Geach, I. Smail, P. N. Best, J. Kurk, M. Casali, R. J. Ivison, and K. Coppin. MNRAS, 388:1473–1486, August 2008.

J. E. Geach, A. Cimatti, W. Percival, Y. Wang, L. Guzzo, G. Zamorani, P. Rosati, L. Pozzetti, A. Orsi, C. M. Baugh, C. G. Lacey, B. Garilli, P. Franzetti, J. R. Walsh, and M. Kümmel. MNRAS, 402:1330–1338, February 2010.

S. S. Gershtein and Y. B. Zel'Dovich. *Soviet Journal of Experimental and Theoretical Physics Letters*, 4:120–+, September 1966.

M. Giavalisco, A. Koratkar, and D. Calzetti. ApJ, 466:831–+, August 1996.

O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai. ApJ, 616:16–26, November 2004.

K. M. Gorski, G. Hinshaw, A. J. Banday, C. L. Bennett, E. L. Wright, A. Kogut, G. F. Smoot, and P. Lubin. ApJ, 430: L89–L92, August 1994.

G. L. Granato, C. G. Lacey, L. Silva, A. Bressan, C. M. Baugh, S. Cole, and C. S. Frenk. ApJ, 542:710–730, October 2000.

C. Gronwall, R. Ciardullo, T. Hickey, E. Gawiser, J. J. Feldmeier, P. G. van Dokkum, C. M. Urry, D. Herrera, B. D. Lehmer, L. Infante, A. Orsi, D. Marchesini, G. A. Blanc, H. Francke, P. Lira, and E. Treister. ApJ, 667:79–91, September 2007.

E. J. Groth and P. J. E. Peebles. ApJ, 217:385–405, October 1977.

A. H. Guth. Phys. Rev. D, 23:347–356, January 1981.

L. Guzzo, M. Pierleoni, B. Meneux, E. Branchini, O. Le Fèvre, C. Marinoni, B. Garilli, J. Blaizot, G. De Lucia, A. Pollo, H. J. McCracken, D. Bottini, V. Le Brun, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, L. Tresse, G. Vettolani, A. Zanichelli, C. Adami, S. Arnouts, S. Bardelli, M. Bolzonella, A. Bongiorno, A. Cappi, S. Charlot, P. Ciliegi, T. Contini, O. Cucciati, S. de la Torre, K. Dolag, S. Foucaud, P. Franzetti, I. Gavignaud, O. Ilbert, A. Iovino, F. Lamareille, B. Marano, A. Mazure, P. Memeo, R. Merighi, L. Moscardini, S. Paltani, R. Pellò, E. Perez-Montero, L. Pozzetti, M. Radovich, D. Vergani, G. Zamorani, and E. Zucca. Nature, 451:541–544, January 2008.

F. Haardt and P. Madau. In D. M. Neumann & J. T. V. Tran, editor, *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, 2001.

F. Haardt and P. Madau. ApJ, 461:20–+, April 1996.

Z. Haiman and M. Spaans. ApJ, 518:138–144, June 1999.

T. Hamana, M. Ouchi, K. Shimasaku, I. Kayo, and Y. Suto. MNRAS, 347:813–823, January 2004.

M. Hansen and S. P. Oh. MNRAS, 367:979–1002, April 2006.

H. Harari. *Physics Letters B*, 216:413–418, January 1989.

G. Harker, S. Cole, and A. Jenkins. MNRAS, 382:1503–1515, December 2007.

J. P. Harrington. MNRAS, 162:43–+, 1973.

L. W. Hartmann, J. P. Huchra, and M. J. Geller. ApJ, 287:487–491, December 1984.

L. W. Hartmann, J. P. Huchra, M. J. Geller, P. O'Brien, and R. Wilson. ApJ, 326:101–109, March 1988.

S. Hatton, J. E. G. Devriendt, S. Ninin, F. R. Bouchet, B. Guiderdoni, and D. Vibert. MNRAS, 343:75–106, July 2003.

E. Hawkins, S. Maddox, S. Cole, O. Lahav, D. S. Madgwick, P. Norberg, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, B. Jones, I. Lewis, S. Lumsden, W. Percival, B. A. Peterson, W. Sutherland, and K. Taylor. MNRAS, 346:78–96, November 2003.

T. Hayashino, Y. Matsuda, H. Tamura, R. Yamauchi, T. Yamada, M. Ajiki, S. S. Fujita, T. Murayama, T. Nagao, K. Ohta, S. Okamura, M. Ouchi, K. Shimasaku, Y. Shioya, and Y. Taniguchi. AJ, 128:2073–2079, November 2004.

M. Hayes, G. Östlin, D. Schaerer, J. M. Mas-Hesse, C. Leitherer, H. Atek, D. Kunth, A. Verhamme, S. de Barros, and J. Melinder. Nature, 464:562–565, March 2010a.

M. Hayes, D. Schaerer, and G. Östlin. A&A, 509:L5+, January 2010b.

J. C. Helly, S. Cole, C. S. Frenk, C. M. Baugh, A. Benson, and C. Lacey. MNRAS, 338:903–912, February 2003.

L. G. Henyey and J. L. Greenstein. ApJ, 93:70–83, January 1941.

L. Hernquist and J. C. Mihos. ApJ, 448:41–+, July 1995.

P. Hibon, J.-G. Cuby, J. Willis, B. Clément, C. Lidman, S. Arnouts, J.-P. Kneib, C. J. Willott, C. Marmo, and H. McCracken. A&A, 515:A97+, June 2010.

G. J. Hill, K. Gebhardt, E. Komatsu, N. Drory, P. J. MacQueen, J. Adams, G. A. Blanc, R. Koehler, M. Rafal, M. M. Roth, A. Kelz, C. Gronwall, R. Ciardullo, and D. P. Schneider. In T. Kodama, T. Yamada, & K. Aoki, editor, *Astronomical Society of the Pacific Conference Series*, volume 399 of *Astronomical Society of the Pacific Conference Series*, pages 115–+, October 2008.

H. Hippelein, C. Maier, K. Meisenheimer, C. Wolf, J. W. Fried, B. von Kuhlmann, M. Kümmel, S. Phleps, and H.-J. Röser. A&A, 402:65–78, April 2003.

F. Hjerting. ApJ, 88:508–+, November 1938.

M. Hoeft, G. Yepes, S. Gottlöber, and V. Springel. MNRAS, 371:401–414, September 2006.

A. M. Hopkins, A. J. Connolly, and A. S. Szalay. AJ, 120:2843–2850, December 2000.

A. Horton, I. Parry, J. Bland-Hawthorn, S. Cianci, D. King, R. McMahon, and S. Medlen. In A. F. M. Moorwood & M. Iye, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5492 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 1022–1032, September 2004.

E. M. Hu, L. L. Cowie, and R. G. McMahon. ApJ, 502:L99+, August 1998.

E. M. Hu, L. L. Cowie, R. G. McMahon, P. Capak, F. Iwamuro, J.-P. Kneib, T. Maihara, and K. Motohara. ApJ, 568: L75–L79, April 2002.

E. M. Hu, L. L. Cowie, P. Capak, R. G. McMahon, T. Hayashino, and Y. Komiyama. AJ, 127:563–575, February 2004.

E. M. Hu, L. L. Cowie, A. J. Barger, P. Capak, Y. Kakazu, and L. Trouille. *in press*, 2010.

E. Hubble. ApJ, 79:8–+, January 1934.

D. H. Hughes, S. Serjeant, J. Dunlop, M. Rowan-Robinson, A. Blain, R. G. Mann, R. Ivison, J. Peacock, A. Efstathiou, W. Gear, S. Oliver, A. Lawrence, M. Longair, P. Goldschmidt, and T. Jenness. Nature, 394:241–247, July 1998.

D. G. Hummer. MNRAS, 125:21–37, 1962.

M. Iye, K. Ota, N. Kashikawa, H. Furusawa, T. Hashimoto, T. Hattori, Y. Matsuda, T. Morokuma, M. Ouchi, and K. Shimasaku. Nature, 443:186–188, September 2006.

P. A. James, N. S. Shane, J. H. Knapen, J. Etherton, and S. M. Percival. A&A, 429:851–867, January 2005.

D. H. Jones and J. Bland-Hawthorn. ApJ, 550:593–611, April 2001.

N. Kaiser. In *The Advanced Maui Optical and Space Surveillance Technologies Conference*, 2006.

N. Kaiser. ApJ, 284:L9–L12, September 1984.

N. Kaiser. MNRAS, 227:1–21, July 1987.

N. Kashikawa, K. Shimasaku, M. A. Malkan, M. Doi, Y. Matsuda, M. Ouchi, Y. Taniguchi, C. Ly, T. Nagao, M. Iye, K. Motohara, T. Murayama, K. Murozono, K. Nariai, K. Ohta, S. Okamura, T. Sasaki, Y. Shioya, and M. Umemura. ApJ, 648:7–22, September 2006.

S. Kaspi, P. S. Smith, H. Netzer, D. Maoz, B. T. Jannuzi, and U. Giveon. ApJ, 533:631–649, April 2000.

G. Kauffmann, J. M. Colberg, A. Diaferio, and S. D. M. White. MNRAS, 303:188–206, February 1999.

R. C. Kennicutt, Jr. ApJ, 272:54–67, September 1983.

R. C. Kennicutt, Jr. ApJ, 498:541–+, May 1998a.

R. C. Kennicutt, Jr. ARA&A, 36:189–232, 1998b.

D. Kereš, N. Katz, D. H. Weinberg, and R. Davé. MNRAS, 363:2–28, October 2005.

H.-S. Kim, C. M. Baugh, S. Cole, C. S. Frenk, and A. J. Benson. MNRAS, 400:1527–1540, December 2009.

M. A. R. Kobayashi, T. Totani, and M. Nagashima. ApJ, 670:919–927, December 2007.

M. A. R. Kobayashi, T. Totani, and M. Nagashima. ApJ, 708:1119–1134, January 2010.

A. Kogut, D. N. Spergel, C. Barnes, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright. ApJS, 148:161–173, September 2003.

K. A. Kornei, A. E. Shapley, D. K. Erb, C. C. Steidel, N. A. Reddy, M. Pettini, and M. Bogosavljević. ApJ, 711: 693–710, March 2010.

K. Kovač, R. S. Somerville, J. E. Rhoads, S. Malhotra, and J. Wang. ApJ, 668:15–22, October 2007.

L. M. Krauss. *ArXiv High Energy Physics - Phenomenology e-prints*, February 2007.

R.-P. Kudritzki, R. H. Méndez, J. J. Feldmeier, R. Ciardullo, G. H. Jacoby, K. C. Freeman, M. Arnaboldi, M. Capaccioli, O. Gerhard, and H. C. Ford. ApJ, 536:19–30, June 2000.

D. Kunth, J. M. Mas-Hesse, E. Terlevich, R. Terlevich, J. Lequeux, and S. M. Fall. A&A, 334:11–20, June 1998.

C. Lacey and S. Cole. MNRAS, 262:627–649, June 1993.

C. G. Lacey, C. M. Baugh, C. S. Frenk, L. Silva, G. L. Granato, and A. Bressan. MNRAS, 385:1155–1178, April 2008.

O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees. MNRAS, 251:128–136, July 1991.

S. D. Landy and A. S. Szalay. ApJ, 412:64–71, July 1993.

P. Laursen and J. Sommer-Larsen. ApJ, 657:L69–L72, March 2007.

P. Laursen, A. O. Razoumov, and J. Sommer-Larsen. ApJ, 696:853–869, May 2009a.

P. Laursen, J. Sommer-Larsen, and A. C. Andersen. ApJ, 704:1640–1656, October 2009b.

P. Laursen, J. Sommer-Larsen, and A. O. Razoumov. *ArXiv e-prints*, September 2010.

M. Le Delliou, C. Lacey, C. M. Baugh, B. Guiderdoni, R. Bacon, H. Courtois, T. Sousbie, and S. L. Morris. MNRAS, 357:L11–L15, February 2005.

M. Le Delliou, C. G. Lacey, C. M. Baugh, and S. L. Morris. MNRAS, 365:712–726, January 2006.

H.-c. Lee, G. Worthey, A. Dotter, B. Chaboyer, D. Jevremović, E. Baron, M. M. Briley, J. W. Ferguson, P. Coelho, and S. C. Trager. ApJ, 694:902–923, April 2009.

J.-S. Lee. ApJ, 192:465–474, September 1974.

M. D. Lehnert and T. M. Heckman. ApJ, 462:651–+, May 1996.

D. N. Limber. ApJ, 117:134–+, January 1953.

A. Loeb and G. B. Rybicki. ApJ, 524:527–535, October 1999.

A. Lue, R. Scoccimarro, and G. Starkman. Phys. Rev. D, 69(4):044005–+, February 2004.

C. Ly, M. A. Malkan, N. Kashikawa, K. Shimasaku, M. Doi, T. Nagao, M. Iye, T. Kodama, T. Morokuma, and K. Motohara. ApJ, 657:738–759, March 2007.

J. Magorrian, S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S. M. Faber, K. Gebhardt, R. Green, C. Grillmair, J. Kormendy, and T. Lauer. AJ, 115:2285–2305, June 1998.

C. Maier, K. Meisenheimer, E. Thommes, H. Hippelein, H. J. Röser, J. Fried, B. von Kuhlmann, S. Phleps, and C. Wolf. A&A, 402:79–85, April 2003.

R. K. Malbon, C. M. Baugh, C. S. Frenk, and C. G. Lacey. MNRAS, 382:1394–1414, December 2007.

R. Mandelbaum, A. Tasitsiomi, U. Seljak, A. V. Kravtsov, and R. H. Wechsler. MNRAS, 362:1451–1462, October 2005.

C. Maraston. MNRAS, 362:799–825, September 2005.

A. T. Marlowe, T. M. Heckman, R. F. G. Wyse, and R. Schommer. ApJ, 438:563–589, January 1995.

P. Martini. AJ, 121:598–610, February 2001.

J. M. Mas-Hesse, D. Kunth, G. Tenorio-Tagle, C. Leitherer, R. J. Terlevich, and E. Terlevich. ApJ, 598:858–877, December 2003.

D. Massa and E. L. Fitzpatrick. ApJS, 60:305–321, January 1986.

P. J. McCarthy, L. Yan, W. Freudling, H. I. Teplitz, E. M. Malumuth, R. J. Weymann, M. A. Malkan, R. A. E. Fosbury, J. P. Gardner, L. J. Storrie-Lombardi, R. I. Thompson, R. E. Williams, and S. R. Heap. ApJ, 520:548–563, August 1999.

D. L. Meier and R. Terlevich. ApJ, 246:L109–L113, June 1981.

A. A. Meiksin. *Reviews of Modern Physics*, 81:1405–1469, October 2009.

N. Metcalfe, T. Shanks, P. M. Weilbacher, H. J. McCracken, R. Fong, and D. Thompson. MNRAS, 370:1257–1273, August 2006.

J. C. Mihos and L. Hernquist. ApJ, 437:L47–L50, December 1994.

D. Milisavljevic, R. A. Fesen, C. L. Gerardy, R. P. Kirshner, and P. Challis. ApJ, 709:1343–1355, February 2010.

H. J. Mo and S. D. M. White. MNRAS, 282:347–361, September 1996.

J. J. Monaghan. ARA&A, 30:543–574, 1992.

T. Morioka, A. Nakajima, Y. Taniguchi, Y. Shioya, T. Murayama, and S. S. Sasaki. PASJ, 60:1219–, December 2008.

N. J. Mostek, C. Bebek, A. Dey, A. Ealet, S. Ho, J. Kneib, M. Lampton, M. Levi, K. Olsen, N. Padmanabhan, J. Pobar, E. Prieto, D. J. Schlegel, M. Sholl, A. Slozar, M. White, and BigBOSS collaboration. In *Bulletin of the American Astronomical Society*, volume 41 of *Bulletin of the American Astronomical Society*, pages 518–+, January 2010.

E. Moy, P. Barmby, D. Rigopoulou, J.-S. Huang, S. P. Willner, and G. G. Fazio. A&A, 403:493–499, May 2003.

K. Nagamine, R. Cen, S. R. Furlanetto, L. Hernquist, C. Night, J. P. Ostriker, and M. Ouchi. *New Astronomy Review*, 50:29–34, March 2006.

K. Nagamine, M. Ouchi, V. Springel, and L. Hernquist. *ArXiv e-prints*, February 2008.

M. Nagashima, C. G. Lacey, C. M. Baugh, C. S. Frenk, and S. Cole. MNRAS, 358:1247–1266, April 2005a.

M. Nagashima, C. G. Lacey, T. Okamoto, C. M. Baugh, C. S. Frenk, and S. Cole. MNRAS, 363:L31–L35, October 2005b.

A. Nakajima, Y. Shioya, T. Nagao, T. Saito, T. Murayam, S. S. Sasaki, A. Yokouchi, and Y. Taniguchi. PASJ, 60: 1249–, December 2008.

J. F. Navarro, C. S. Frenk, and S. D. M. White. MNRAS, 275:720–740, August 1995.

J. F. Navarro, C. S. Frenk, and S. D. M. White. ApJ, 462:563–+, May 1996.

J. F. Navarro, C. S. Frenk, and S. D. M. White. ApJ, 490:493–+, December 1997.

D. A. Neufeld. ApJ, 350:216–241, February 1990.

D. A. Neufeld. ApJ, 370:L85–L88, April 1991.

R. Nichol. *ArXiv Astrophysics e-prints*, November 2006.

S. Niklas, U. Klein, and R. Wielebinski. A&A, 322:19–28, June 1997.

K. K. Nilsson and K. Meisenheimer. *New Astronomical Reviews*, 53:37–39, June 2009.

K. K. Nilsson, J. P. U. Fynbo, P. Møller, and A. Orsi. In J. Afonso, H. C. Ferguson, B. Mobasher, & R. Norris, editor, *Deepest Astronomical Surveys*, volume 380 of *Astronomical Society of the Pacific Conference Series*, pages 61–+, December 2007a.

K. K. Nilsson, A. Orsi, C. G. Lacey, C. M. Baugh, and E. Thommes. A&A, 474:385–392, November 2007b.

K. K. Nilsson, O. Möller-Nilsson, P. Møller, J. P. U. Fynbo, and A. E. Shapley. MNRAS, 400:232–237, November 2009.

T. Okamoto, L. Gao, and T. Theuns. MNRAS, 390:920–928, November 2008.

A. Orsi, C. G. Lacey, C. M. Baugh, and L. Infante. MNRAS, 391:1589–1604, December 2008.

D. E. Osterbrock. ApJ, 135:195–+, January 1962.

D. E. Osterbrock. *Astrophysics of gaseous nebulae and active galactic nuclei*. 1989.

G. Östlin, M. Hayes, D. Kunth, J. M. Mas-Hesse, C. Leitherer, A. Petrosian, and H. Atek. AJ, 138:923–940, September 2009.

M. Ouchi, K. Shimasaku, M. Akiyama, K. Sekiguchi, H. Furusawa, S. Okamura, N. Kashikawa, M. Iye, T. Kodama, T. Saito, T. Sasaki, C. Simpson, T. Takata, T. Yamada, H. Yamanoi, M. Yoshida, and M. Yoshida. ApJ, 620:L1–L4, February 2005.

M. Ouchi, K. Shimasaku, M. Akiyama, C. Simpson, T. Saito, Y. Ueda, H. Furusawa, K. Sekiguchi, T. Yamada, T. Kodama, N. Kashikawa, S. Okamura, M. Iye, T. Takata, M. Yoshida, and M. Yoshida. ApJS, 176:301–330, June 2008.

M. Ouchi, K. Shimasaku, H. Furusawa, T. Saito, M. Yoshida, M. Akiyama, Y. Ono, T. Yamada, K. Ota, N. Kashikawa, M. Iye, T. Kodama, S. Okamura, C. Simpson, and M. Yoshida. *ArXiv e-prints*, July 2010.

R. B. Partridge and P. J. E. Peebles. ApJ, 147:868–+, March 1967.

S. Pascual, J. Gallego, A. Aragón-Salamanca, and J. Zamorano. A&A, 379:798–806, December 2001.

J. A. Peacock, P. Schneider, G. Efstathiou, J. R. Ellis, B. Leibundgut, S. J. Lilly, and Y. Mellier. Technical report, October 2006.

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

P. J. E. Peebles. *Principles of physical cosmology*. 1993.

P. J. E. Peebles and E. J. Groth. ApJ, 196:1–11, February 1975.

P. J. E. Peebles and M. G. Hauser. ApJS, 28:19–+, November 1974.

P. J. E. Peebles and J. T. Yu. ApJ, 162:815–+, December 1970.

W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay. MNRAS, 381: 1053–1066, November 2007.

S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. ApJ, 517:565–586, June 1999.

W. H. Press and P. Schechter. ApJ, 187:425–438, February 1974.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in FORTRAN. The art of scientific computing*. 1992.

R. Quadri, D. Marchesini, P. van Dokkum, E. Gawiser, M. Franx, P. Lira, G. Rudnick, C. M. Urry, J. Maza, M. Kriek, L. F. Barrientos, G. A. Blanc, F. J. Castander, D. Christlein, P. S. Coppi, P. B. Hall, D. Herrera, L. Infante, E. N. Taylor, E. Treister, and J. P. Willis. AJ, 134:1103–1117, September 2007.

M. Rauch, M. Haehnelt, A. Bunker, G. Becker, F. Marleau, J. Graham, S. Cristiani, M. Jarvis, C. Lacey, S. Morris, C. Peroux, H. Röttgering, and T. Theuns. ApJ, 681:856–880, July 2008.

J. Retzlaff, P. Rosati, M. Dickinson, B. Vandame, C. Rité, M. Nonino, C. Cesarsky, and GOODS Team. A&A, 511: A50+, February 2010.

J. E. Rhoads, A. Dey, S. Malhotra, D. Stern, H. Spinrad, B. T. Jannuzi, S. Dawson, M. J. I. Brown, and E. Landes. AJ, 125:1006–1013, March 2003.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. AJ, 116:1009–1038, September 1998.

N. Ross, E. S. Sheldon, A. D. Myers, C. Yeche, G. T. Richards, R. G. McMahon, J. F. Hennawi, K. Lee, W. M. Wood-Vasey, A. Weyant, P. Petitjean, D. J. Eisenstein, R. C. Nichol, N. Padmanabhan, D. J. Schlegel, D. P. Schneider, M. A. Strauss, D. H. Weinberg, and M. White. In *Bulletin of the American Astronomical Society*, volume 41 of *Bulletin of the American Astronomical Society*, pages 517–+, January 2010.

N. P. Ross, T. Shanks, and J. Cruz da Ângela. In N. Metcalfe & T. Shanks, editor, *Cosmic Frontiers*, volume 379 of *Astronomical Society of the Pacific Conference Series*, pages 68–+, December 2007.

A. G. Sánchez, M. Crocce, A. Cabré, C. M. Baugh, and E. Gaztañaga. MNRAS, 400:1643–1664, December 2009.

M. R. Santos, R. S. Ellis, J.-P. Kneib, J. Richard, and K. Kuijken. ApJ, 606:683–701, May 2004.

W. Saunders, C. Frenk, M. Rowan-Robinson, A. Lawrence, and G. Efstathiou. Nature, 349:32–38, January 1991.

D. Schaerer. *ArXiv e-prints*, June 2007.

D. Schaerer and A. Verhamme. A&A, 480:369–377, March 2008.

R. Scoccimarro. Phys. Rev. D, 70(8):083007–+, October 2004.

M. Seldner and P. J. E. Peebles. ApJ, 225:7–20, October 1978.

A. E. Shapley, C. C. Steidel, M. Pettini, and K. L. Adelberger. ApJ, 588:65–89, May 2003.

R. K. Sheth, H. J. Mo, and G. Tormen. MNRAS, 323:1–12, May 2001.

H. Shim, J. Colbert, H. Teplitz, A. Henry, M. Malkan, P. McCarthy, and L. Yan. ApJ, 696:785–796, May 2009.

K. Shimasaku, T. Hayashino, Y. Matsuda, M. Ouchi, K. Ohta, S. Okamura, H. Tamura, T. Yamada, and R. Yamauchi. ApJ, 605:L93–L96, April 2004.

K. Shimasaku, N. Kashikawa, M. Doi, C. Ly, M. A. Malkan, Y. Matsuda, M. Ouchi, T. Hayashino, M. Iye, K. Motohara, T. Murayama, T. Nagao, K. Ohta, S. Okamura, T. Sasaki, Y. Shioya, and Y. Taniguchi. PASJ, 58:313–334, April 2006.

Y. Shioya, Y. Taniguchi, S. S. Sasaki, T. Nagao, T. Murayama, M. I. Takahashi, M. Ajiki, Y. Ideue, S. Mihara, A. Nakajima, N. Z. Scoville, B. Mobasher, H. Aussel, M. Giavalisco, L. Guzzo, G. Hasinger, C. Impey, O. Le Fèvre, S. Lilly, A. Renzini, M. Rich, D. B. Sanders, E. Schinnerer, P. Shopbell, A. Leauthaud, J.-P. Kneib, J. Rhodes, and R. Massey. ApJS, 175:128–137, March 2008.

L. Silva, G. L. Granato, A. Bressan, and L. Danese. ApJ, 509:103–117, December 1998.

P. Simon. A&A, 473:711–714, October 2007.

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

R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. MNRAS, 341:1311–1332, June 2003.

G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. ApJ, 396:L1–L5, September 1992.

D. Sobral, P. N. Best, J. E. Geach, I. Smail, J. Kurk, M. Cirasuolo, M. Casali, R. J. Ivison, K. Coppin, and G. B. Dalton. MNRAS, 398:75–90, September 2009.

V. Springel and L. Hernquist. MNRAS, 339:289–311, February 2003.

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Nature, 435:629–636, June 2005.

D. P. Stark, R. S. Ellis, J. Richard, J.-P. Kneib, G. P. Smith, and M. R. Santos. ApJ, 663:10–28, July 2007.

G. Stasińska. A&AS, 83:501–538, June 1990.

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

C. C. Steidel, K. L. Adelberger, M. Giavalisco, M. Dickinson, and M. Pettini. ApJ, 519:1–17, July 1999.

D. Strickland. In R. Fusco-Femiano & F. Matteucci, editor, *Chemical Enrichment of Intracluster and Intergalactic Medium*, volume 253 of *Astronomical Society of the Pacific Conference Series*, pages 387–+, 2002.

M. Sullivan, M. A. Treyer, R. S. Ellis, T. J. Bridges, B. Milliard, and J. Donas. MNRAS, 312:442–464, February 2000.

R. S. Sutherland and M. A. Dopita. ApJS, 88:253–327, September 1993.

Y. Taniguchi, M. Ajiki, T. Nagao, Y. Shioya, T. Murayama, N. Kashikawa, K. Kodaira, N. Kaifu, H. Ando, H. Karoji, M. Akiyama, K. Aoki, M. Doi, S. S. Fujita, H. Furusawa, T. Hayashino, F. Iwamuro, M. Iye, N. Kobayashi, T. Kodama, Y. Komiyama, Y. Matsuda, S. Miyazaki, Y. Mizumoto, T. Morokuma, K. Motohara, K. Nariai, K. Ohta, Y. Ohyama, S. Okamura, M. Ouchi, T. Sasaki, Y. Sato, K. Sekiguchi, K. Shimasaku, H. Tamura, M. Umemura, T. Yamada, N. Yasuda, and M. Yoshida. PASJ, 57:165–182, February 2005.

A. Tasitsiomi. ApJ, 645:792–813, July 2006.

M. Tegmark, M. A. Strauss, M. R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D. H. Weinberg, I. Zehavi, N. A. Bahcall, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, A. Berlind, T. Budavari, A. Connolly, D. J. Eisenstein, D. Finkbeiner, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R. C. Nichol, J. P. Ostriker, A. Pope, R. Scranton, U. Seljak, R. K. Sheth, A. Stebbins, A. S. Szalay, I. Szapudi, Y. Xu, J. Annis, J. Brinkmann, S. Burles, F. J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, B. Gillespie, G. Hennessy, D. W. Hogg, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, P. Kunszt, J. A. Munn, L. O'Connell, J. Peoples, J. R. Pier, M. Richmond, C. Rockosi, D. P. Schneider, C. Stoughton, D. L. Tucker, D. E. vanden Berk, B. Yanny, and D. G. York. Phys. Rev. D, 69(10):103501–+, May 2004.

H. I. Teplitz, J. P. Gardner, E. M. Malumuth, and S. R. Heap. ApJ, 507:L17–L20, November 1998.

The SDSS Collaboration, I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall, A. A. Berlind, J. Brinkmann, J. A. Frieman, J. E. Gunn, R. H. Lupton, R. C. Nichol, W. J. Percival, D. P. Schneider, R. A. Skibba, M. A. Strauss, M. Tegmark, and D. G. York. *ArXiv e-prints*, May 2010.

D. Thompson, F. Mannucci, and S. V. W. Beckwith. AJ, 112:1794–+, November 1996.

R. I. Thompson, L. J. Storrie-Lombardi, R. J. Weymann, M. J. Rieke, G. Schneider, E. Stobie, and D. Lytle. AJ, 117:17–39, January 1999.

T. X. Thuan and Y. I. Izotov. ApJ, 489:623–+, November 1997.

V. Tilvi, J. E. Rhoads, P. Hibon, S. Malhotra, J. Wang, S. Veilleux, R. Swaters, R. Probst, H. Krug, S. L. Finkelstein, and M. Dickinson. *ArXiv e-prints*, June 2010.

M. Trenti and P. Hut. *ArXiv e-prints*, June 2008.

L. Tresse and S. J. Maddox. ApJ, 495:691–+, March 1998.

L. Tresse, S. J. Maddox, O. Le Fèvre, and J.-G. Cuby. MNRAS, 337:369–383, November 2002.

B. P. Venemans, R. G. McMahon, I. R. Parry, D. J. King, J. Bland-Hawthorn, and A. J. Horton. In A. Moorwood, editor, *Science with the VLT in the ELT Era*, pages 187–+, 2009.

A. Verhamme, D. Schaerer, and A. Maselli. A&A, 460:397–413, December 2006.

A. Verhamme, D. Schaerer, H. Atek, and C. Tapken. A&A, 491:89–111, November 2008.

V. Villar, J. Gallego, P. G. Pérez-González, S. Pascual, K. Noeske, D. C. Koo, G. Barro, and J. Zamorano. ApJ, 677: 169–185, April 2008.

Y. Wang. *Journal of Cosmology and Astro-Particle Physics*, 5:21–+, May 2008.

R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov, and B. Allgood. ApJ, 652:71–84, November 2006.

S. D. M. White and M. J. Rees. MNRAS, 183:341–358, May 1978.

S. D. M. White, M. Davis, and C. S. Frenk. MNRAS, 209:27P–31P, July 1984.

L. Yan, P. J. McCarthy, L. J. Storrie-Lombardi, and R. J. Weymann. ApJ, 503:L19+, August 1998.

L. Yan, P. J. McCarthy, W. Freudling, H. I. Teplitz, E. M. Malumuth, R. J. Weymann, and M. A. Malkan. ApJ, 519: L47–L50, July 1999.

D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, and N. Yasuda. AJ, 120:1579–1587, September 2000.

Y. B. Zel'dovich and M. Y. Khlopov. *Soviet Physics Uspekhi*, 24:755–774, September 1981.

Z. Zheng and J. Miralda-Escudé. ApJ, 578:33–42, October 2002.