The Life Cycle of Nearby Galaxies: secular and external processes regulating atomic gas content

Toby Brown

Presented in fulfillment of the requirements of the degree of Doctor of Philosophy

–

Faculty of Science, Engineering and Technology
Swinburne University

Abstract

This thesis investigates how secular and external processes combine to regulate gas content and drive the life cycle of nearby galaxies. We use a volume and stellar mass limited, multi-wavelength parent sample of ∼30,000 galaxies ($10^9 \leq M_\star/M_\odot \leq 10^{11.5}$, $0.02 \leq z \leq 0.05$), selected from the Sloan Digital Sky Survey with H I data from the Arecibo Legacy Fast ALFA survey. This statistically-powerful sample is representative of gas content, star formation and environment in the local Universe and, as such, is perfectly suited for disentangling the internal and external processes that regulate galaxy evolution. The main questions this thesis addresses are: i) how does gas content regulate the star formation process and build up of stellar mass in galaxies?, ii) where and how precisely does external influence on H I reservoirs become important? and iii) how does the cycling of gas into and out of galaxies affect the star formation cycle?

We revisit key gas fraction ($M_{HI}/M_\star$) scaling relations, taking advantage of the H I spectral stacking technique to demonstrate that specific star formation rate (sSFR) is the most important parameter for tracing H I content. In fact, the gas fraction scaling relations with stellar mass and stellar surface density are primarily driven by a combination of the underlying galaxy bimodality in sSFR and the integrated Kennicutt-Schmidt law. Controlling for stellar mass and sSFR, we go on to examine the external processes responsible for the quenching of satellite galaxies, showing that systematic gas suppression begins in the group regime and continues into the cluster. This externally-driven depletion is both fast-acting and more closely associated to halo mass than local density. We invoke ram-pressure stripping to explain this. Our results are then compared with state-of-the-art theoretical models and discussed within this context, showing that more work is needed if theory is to reproduce the observations. Finally, we quantify the strong anti-correlation between H I mass and metallicity at fixed stellar mass. In addition, we establish that the dependence of metallicity on star formation is comparatively weak and heavily reliant upon the abundance and star formation rate estimates used. These trends support a scenario where galaxies exist in an evolving equilibrium between gas, metallicity and star formation. The fact that deviations from this equilibrium are most strongly correlated with gas mass suggests that the scatter in the mass-metallicity relation is primarily driven by fluctuations in gas accretion. In summary, these are new insights into the relationships between gas content, galaxy properties and environment in the local Universe, providing strong constraints for galaxy formation and evolution models. They also demonstrate the importance of H I stacking as a versatile tool for statistical studies of H I content.

Acknowledgements

First things first. This work could not have happened without my supervisors and mentors, Barbara Catinella and Luca Cortese, both of whom have been brilliant and patient in equal measure. Thanks are also due to Virginia Kilborn for providing balance and valuable counsel when the project was in need.

I have been fortunate to work with a group of colleagues who have been as honest in their criticism as generous in their friendship. They include Katinka Gereb, Adam Stevens, Katharina Lutz, David Lagattuta, Claudia Lagos and Steven Janowiecki. I am also hugely grateful for the encouragement and advice of those with a bit more experience in this game. Among them I would like to single out Jeff Cooke, Mathew Bailes, Glenn Kacprzak, Chris Blake and Thibault Garel.

Going back a little more. At Queen Elizabeth School, Stephen Bell and Stewart Rushton provided early encouragement and inspired me to study physics. In Liverpool, Toby Moore, David Eden, Phil James and Steve Barrett taught me what physics really was and helped me to follow in their example and secure a PhD scholarship.

Finally, no one could have asked for a better companion on this journey than my partner Sophie Dixon. Writing a thesis is a stressful, time consuming business but Sophie's love and support kept me calm and saw me through. I also wish to thank Mum, Dad and Zach, who taught me much, including the importance of being curious. Without failure, they have all been there for me time and time again and for that I will be forever grateful.

Declaration

The work presented in this thesis has been carried out in the Centre for Astrophysics & Supercomputing at Swinburne University of Technology between 2013 and 2017. This thesis contains no material that has been accepted for the award of any other degree or diploma. To the best of my knowledge, this thesis contains no material previously published or written by another author, except where due reference is made in the text of the thesis. The content of the chapters listed below has appeared in refereed journals. Minor alterations have been made to the published papers in order to maintain argument continuity and consistency of spelling and style.

- Sections of Chapter 2 have been published as Brown, T., Catinella, B., Cortese, L., Kilborn, V., Haynes, M. P., Giovanelli, R., *The effect of structure and star formation on the gas content of nearby galaxies*, MNRAS, 452, 2479 and Brown, T., Catinella, B., Cortese, L., Lagos, C.d.P., Davé, R., Kilborn, V., Haynes, M. P., Giovanelli, R., Rafieferantsoa, M., *Cold gas stripping in satellite galaxies: from pairs to clusters*, MNRAS, 466, 1275.

- Sections of Chapter 3 have been published as Brown, T., Catinella, B., Cortese, L., Kilborn, V., Haynes, M. P., Giovanelli, R., *The effect of structure and star formation on the gas content of nearby galaxies*, MNRAS, 452, 2479 and Brown, T., Catinella, B., Cortese, L., Lagos, C.d.P., Davé, R., Kilborn, V., Haynes, M. P., Giovanelli, R., Rafieferantsoa, M., *Cold gas stripping in satellite galaxies: from pairs to clusters*, MNRAS, 466, 1275.

- Chapter 4 has been published as Brown, T., Catinella, B., Cortese, L., Kilborn, V., Haynes, M. P., Giovanelli, R., *The effect of structure and star formation on the gas content of nearby galaxies*, MNRAS, 452, 2479.

- Chapters 5 and 6 have been published as Brown, T., Catinella, B., Cortese, L., Lagos, C.d.P., Davé, R., Kilborn, V., Haynes, M. P., Giovanelli, R., Rafieferantsoa, M., *Cold gas stripping in satellite galaxies: from pairs to clusters*, MNRAS, 466, 1275.

- Chapter 7 has been drafted as a paper with the title *Gas as the primary driver of the mass-metallicity relation.*

- Appendix A has been published as part of Brown, T., Catinella, B., Cortese, L., Kilborn, V., Haynes, M. P., Giovanelli, R., *The effect of structure and star formation on the gas content of nearby galaxies*, MNRAS, 452, 2479

Toby Brown
Perth, Australia
March 2017

To my parents.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Galaxy Evolution

How do galaxies evolve? is one of the grandest questions in all of science. Simply put, galaxies are defined as vast, dynamically bound ensembles of stars. This statement, however, belies the fact that they are the most visually striking and diverse class of objects in the Universe, whose nature and origins have intrigued astronomers for centuries.

Until the latter half of the twentieth century, galaxies were regarded as "island universes" - permanent systems uniformly distributed throughout space. Explicitly or implicitly, they were considered to be distinct from their neighbours, existing as gravitationally and chemically self-contained units. A major paradigm shift in extragalactic astronomy came with the recognition that galaxies are not isolated, but occupy a variety of structures (Holmberg, 1941; Abell, 1958; Burbidge & Burbidge, 1961; de Vaucouleurs, 1975). We now have a concordance model of cosmology (Lambda cold dark matter or ΛCDM; Peebles, 1982; Blumenthal et al., 1984; Davis et al., 1985; Planck Collaboration et al., 2016) that relies fundamentally upon the hierarchical formation of galaxies and large-scale structure, the main aspects of which are the following: i) galaxies, even those in the field, are embedded in dark matter halos whose filaments are connected to the cosmic web and halos of other galaxies; ii) galaxies, and their dark matter halos, grow by cannibalising their neighbours; and iii) baryons, mainly in the form of primordial or recycled gas, are accreted smoothly and continuously onto galaxies (see Larson & Tinsley, 1978; White & Rees, 1978; Walker et al., 1996; Mihos & Hernquist, 1996; Kereš et al., 2005; Oppenheimer et al., 2010; Nelson et al., 2013; Finlator, 2016).

In the 1930's, Edwin Hubble used his famous tuning fork diagram to establish that there are two basic types of galaxies; spirals and ellipticals. Spiral galaxies, such as the Milky Way, have a highly flattened disk that contains stars, gas and dust, and are pre-

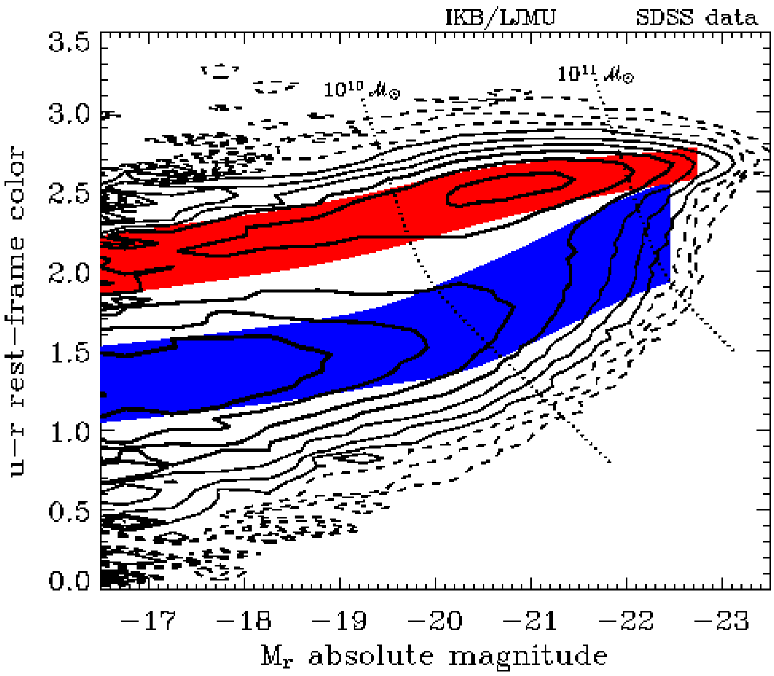

Figure 1.1: Galaxy colour versus absolute magnitude for ∼70,000 SDSS galaxies at $z < 0.1$. Contours represent number density and are logarithmically spaced. Dotted lines denote constant stellar mass assuming the mass-to-light ratios of Kauffmann et al. (2003a). There are two clear sequences in colour, red and blue, and their 1σ scatter is traced by the shaded regions. The sample has been corrected for Malmquist bias. This figure was produced by Ivan Baldry and is adapted from Figure 2 in Baldry et al. (2004).

dominantly supported by their own rotation. On the other hand, elliptical galaxies are smooth, featureless systems supported by the random motion of their stars and containing very little dust or gas. Of course, in reality these classifications sit at opposing ends of a continuous morphological sequence with most galaxies residing somewhere in the middle. Traversing this sequence from ellipticals to spirals, galaxies are said to change from early- to late-type. Although this nomenclature is widely used (including in this thesis), it has no direct physical basis. Since morphologies are determined by physical processes, the visual classification of galaxies remains important, however, it is not always clear which morphological signatures are fundamental to galaxy formation. In light of this, and driven by advances in data and methods, the field of galaxy evolution is moving away from subjective classifications of morphology, instead choosing more physically motivated, quantifiable parameters (e.g. mass density, angular momentum) to characterise the evolutionary state of galaxies.

Perhaps the most interesting aspect of the Hubble sequence is its strong correlation with intrinsic galaxy properties such as mass, star formation and gas content. To first order, the separation of galaxies into blue (star forming) and red (quiescent) sequences as function of magnitude or stellar mass shown in Figure 1.1 simply reflects the division into late- and early-types described above (e.g. Baldry et al., 2004; Blanton, 2006; Salim et al., 2007; Schiminovich et al., 2007; Wyder et al., 2007; Schawinski et al., 2014). However, the colour-morphology relationship is not a tight one. Late-type galaxies can be red and, although rarer, early-type galaxies can be blue. Thus morphology is not a fundamental driver of this picture, rather the colour bimodality tells us that, in most galaxies, the star formation cycle is either still ongoing or was quenched billions of years ago. The astrophysical mechanisms behind this systematic quenching of star formation remain among the most hotly debated topics in astronomy today.

Since Hubble's time, the statistical classification and study of galaxies has come along way. Our Galaxy is one of billions, each one different from the last, and yet this diverse population is found to obey a complex network of scaling relations between their global properties, some of which show remarkably small scatter. These relationships are governed by the physical mechanisms that drive galaxy formation and evolution, thus, if we are to develop a successful theory of galaxy evolution, we must be able to explain their origin.

Over the last few decades, this task has been made easier by the arrival of large infrared (IR), optical and ultra-violet (UV) galaxy surveys providing highly complete, multi-wavelength datasets that link star formation with stellar content for more than a million objects across cosmic time (e.g. the Two-degree-Field Galaxy Redshift Survey (2dFGRS), the Sloan Digital Sky Survey (SDSS), the Galaxy Evolution Explorer (GALEX) survey and the Galaxy And Mass Assembly (GAMA) survey; Colless et al., 2001; York et al., 2000; Martin et al., 2005; Driver et al., 2011). These programs are now an indispensable tool for those wishing to understand the physical mechanisms that drive galaxy evolution from a statistical perspective.

One of the most important examples of a global galaxy scaling relation is the tight correlation between star formation rate (SFR) and stellar mass known as the star forming "main sequence" (Brinchmann et al., 2004; Salim et al., 2007; Noeske et al., 2007). Most galaxies that are forming stars are doing so at a rate that is roughly proportional to their stellar mass. The existence of this relation suggests that, rather than growing erratically, galaxies exist in a steadily evolving state where their ability to form stars is in quasi-equilibrium with the supply of gas and the amount of material returned to the intergalactic medium through outflows (Lilly et al., 2013; Davé et al., 2011b).

In actively star forming galaxies, the interplay between gas processing and the exchange of material between galaxies and their environments is also reflected in the relationship

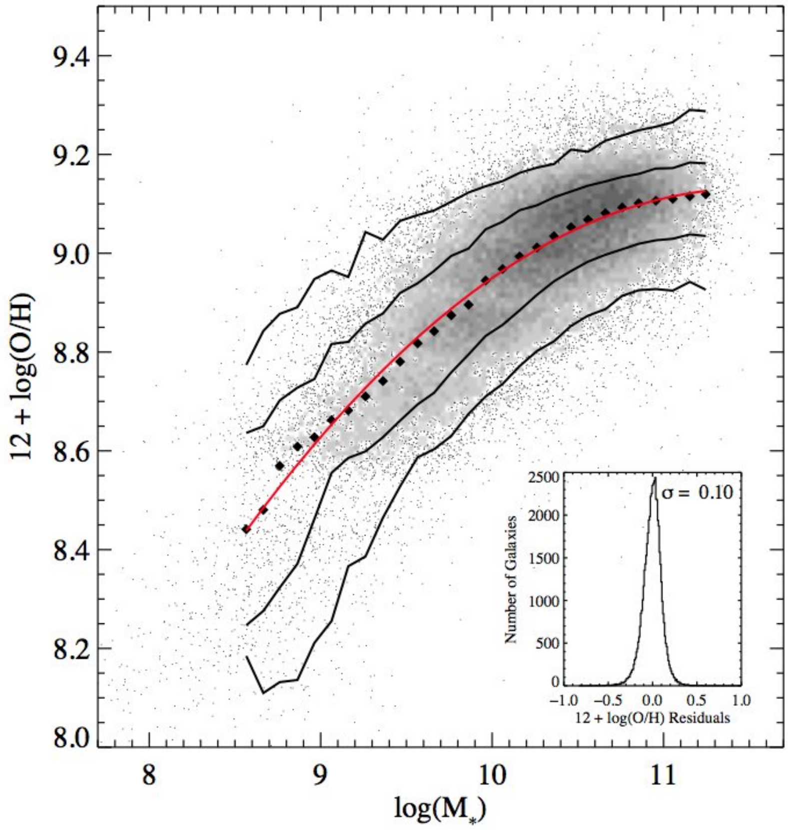

Figure 1.2: The mass-metallicity relation for 53,400 star forming galaxies in SDSS. Stellar mass is in solar units and gas-phase metallicity is expressed in terms of the abundance by number of oxygen to hydrogen, O/H. This figure is taken from Tremonti et al. (2004).

connecting stellar mass with gas-phase metallicity, the mass-metallicity relation (MZR; Tremonti et al., 2004). Stellar mass, as the product of many successive generations of interstellar medium (ISM), and gas-phase metallicity, the metals returned to the ISM via star formation and stellar evolution, are intimately related to a galaxy's star formation and accretion histories. The MZR for ~50,000 SDSS galaxies is shown in Figure 1.2. Metal content, as measured by optical emission lines, is expressed as the abundance ratio of oxygen to hydrogen and increases with stellar mass up to a point of saturation around a few 10^{10} M$_{\odot}$.

As with most scaling relations, the dispersion in the MZR contains information on the prominence of different physical mechanisms in regulating galaxy evolution. It is therefore interesting that the scatter in the MZR appears to be dependent on SFR, in the sense that highly star forming galaxies are typically found to have lower metallicities (Ellison et al., 2008; Lara-López et al., 2010; Mannucci et al., 2010). There are a small number of studies who have shown a similar correlation with the atomic gas content, with gas-rich galaxies being metal-poor (Hughes et al., 2013; Bothwell et al., 2013; Lara-López et al., 2013b). This result simultaneously downplays the importance of mergers as a method of delivering fresh gas into star forming galaxies (Papovich et al., 2011; Behroozi et al., 2013) and supports the role of galactic gas flows in modulating the processes of galaxy evolution. Understanding the chemical enrichment of galaxies is a fundamental challenge for studies of galaxy evolution and while it is clear that the stellar mass and gas-phase metallicity are two of the most fundamental properties of galaxies, the physical nature and origin of their relationship and secondary dependencies remain widely debated.

1.2 Galaxy Environment

Galaxies are not found in isolation. Instead, they exist in a hierarchical, large-scale structure that stretches from vast, low density voids to massive clusters containing many thousands of systems in close proximity. By definition, the group environment encompasses everything in between; from associations and chance encounters of just a few galaxies, to large dark matter halos containing many tens or even hundreds of gravitationally bound members. The dividing line between these regimes is, to a large extent, arbitrary and therefore variable. However, a dark matter halo mass of 10^{14} M$_{\odot}$ and above is frequently cited as the criteria for clusters while an association of two or more objects is considered to be a group. Field galaxies are systems whose neighbours reside below the sensitivity limit of observations or at such great distances that they may be considered isolated objects. These broad definitions are employed throughout this thesis.

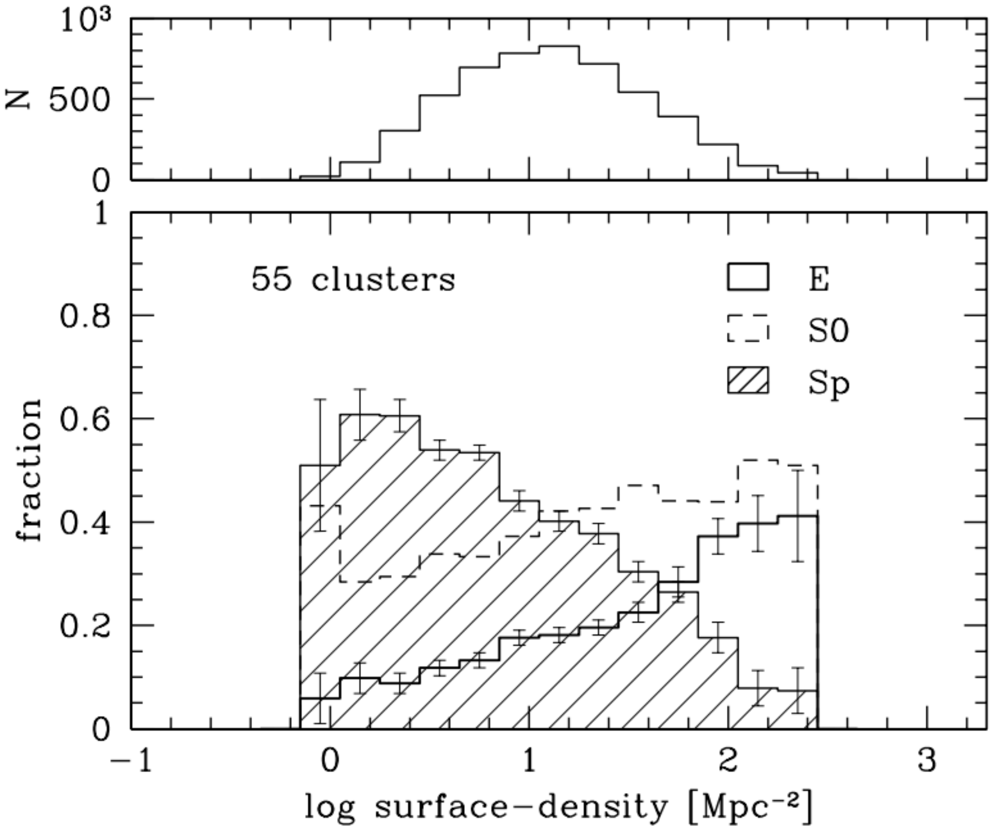

Figure 1.3: The morphology-density relation shown as the fraction of elliptical, lenticular and spiral galaxies (denoted as E, S0, Sp respectively) as a function of projected galaxy surface density for the Dressler (1980) sample of 55 local clusters. This figure is taken from Mo et al. (2010) and is based upon the original work by Dressler (1980).

The famous morphological-density relation of Dressler (1980) shown in Figure 1.3 simultaneously established and solidified the importance of environment in governing galaxy evolution. The general sense of this relation is that the fraction of early-type galaxies in the overall galaxy population increases as a function of local density, defined as the number of galaxies per Mpc^2. The inverse of this statement, quite obviously, is also true. There are more late-type galaxies at lower environmental densities.

Even in galaxies of similar mass and/or morphology, studies show that star formation properties correlate strongly with environment. As galaxies transition into higher density regions their star formation is quenched on timescales that differ depending on the mechanisms involved (Balogh et al., 1998; Gómez et al., 2003; Hogg et al., 2004; Baldry et al., 2006; Wetzel et al., 2009; Grützbauch et al., 2011; Peng et al., 2012; Wilman & Erwin, 2012; Wetzel et al., 2012, 2013). Furthermore, at fixed star formation rate the trend of morphology with environment is almost removed (Kauffmann et al., 2004; Blanton et al., 2005; Bamford et al., 2009). This trend highlights the role of environment in the shut down of star formation activity, with different regimes driving galaxies down different evolutionary paths.

There is a range of physical mechanisms believed to be responsible for the quenching of star formation across different environments. These are frequently divided into two categories; gravitational and hydrodynamical. In the first camp, the dynamical interaction of two or more galaxies can result in the tidal stripping of stars and gas (Moore et al., 1999). If the velocity and frequency of galaxy-galaxy encounters is high enough (i.e. large groups or clusters), harassment can tidally heat galaxies shutting off inflow and causing evaporation of H I content (Moore et al., 1998). As well as altering the structure and morphology of galaxies, mergers can cause gas consumption that boosts star formation in the short term but eventually quenches the system (Makino & Hut, 1997; Wetzel et al., 2009; White et al., 2010). In the second category, the hydrodynamical process of starvation - or strangulation - is the lack of gas cooling, cutting off in the external gas supply, either to the halo or to the galaxy itself, so that reservoirs are not replenished (Larson et al., 1980; Balogh et al., 2000). For galaxies entering a cluster, the interaction between the ISM and intergalactic medium (IGM) has been shown to be strong enough to rapidly remove gas from the disk in a process known as ram-pressure stripping, inevitably shutting off star formation (Gunn & Gott, 1972; Hester, 2006; Chung et al., 2009). On slower timescales, viscous stripping or thermal evaporation of the cold gas content of galaxies surrounded by a hot IGM can also suppress star formation (Cowie & McKee, 1977; Nulsen, 1982).

Tellingly, the suppression or destruction of galaxy gas reservoirs is the common denominator across all these scenarios. It is for this reason that H I is of primary importance for those wishing to understand how environment influences galaxy evolution. We cannot form a complete picture without extensive characterisation of the gas content of galaxies across the full range of environments.

1.3 Neutral Atomic Hydrogen

One of the more obvious statements in astrophysics is that star formation requires a source of gas. However, this simple fact is the end result of a complex cosmic supply chain that is determined by structure on the largest scales and spans almost 30 orders of magnitude in density. As ionized gas (H II) contained within a dark matter halo cools, it loses pressure support and flows inward to the bottom of the potential well. Once the condensing gas reaches a critical density, and if the gas cooling in this process is efficient, the H II will be converted into neutral atomic hydrogen (H I), and then into the dense clouds and cores of molecular gas (H_2), inside of which stars can form.

In extragalactic astronomy, the clearest observational evidence for the importance of gas to the star formation process is the scaling relation between the integrated SFR surface density and total gas (atomic + molecular) mass surface density (Schmidt, 1959; Kennicutt, 1998a,b). This so-called Kennicutt-Schmidt law is shown in Figure 1.4. The ubiquitous nature of this relationship reflects the fact that gas is the input driver of star formation and, as such, a principle agent of galaxy evolution.

Since the formation of stars is a local, small scale process when compared to the host galaxy, it is intriguing that the global star formation scaling relations exists with such small scatter. Following this, recent studies have used resolved observations to examine local star formation as function of atomic and molecular gas in nearby galaxies, showing that the strongest driver of local SFR surface density is the molecular hydrogen, not H I (Wong & Blitz, 2002; Leroy et al., 2008; Bigiel et al., 2008). Furthermore, there is evidence to suggest that, due to self-shielding and the formation of H_2, the H I saturates at a surface density of around 10 $M_\odot$ pc^2 while SFR continues to increase. This work suggests the global Kennicutt-Schmidt law is actually a superposition of two more fundamental processes, the conversion of H I into H_2 and the conversion of H_2 into stars.

H I is the principle component of interstellar gas in galaxies (~90 per cent by number), thus, any theory of galaxy formation must address the question of how atomic hydrogen gas is regulated by both the star formation cycle and environment.

Although large galaxy surveys at IR-to-UV wavelengths have been standard practice for decades, folding 21 cm ($\nu = 1420.4$ MHz) spectral line observations of gas content into multi-wavelength, large statistical studies is a comparatively recent advance. This change has been driven by new and improved instrumentation (e.g. multi-beam receivers) that have sped up the process of accumulating H I observations, making it possible to add gas masses to the list of physical properties measurable over large galaxy samples (e.g. Meyer et al., 2004; Springob et al., 2005; Wong et al., 2006; Haynes et al., 2011; Catinella et al., 2013).

A growing body of H I-focussed work has established the interwoven scaling relations that exist between galaxy gas content and the stellar, star formation and environmental properties of galaxies (Bothun, 1984; Broeils & Rhee, 1997; Kannappan, 2004; Boselli & Gavazzi, 2006; Catinella et al., 2010; Fabello et al., 2011a,b, 2012; Cortese et al., 2011; Oh et al., 2011; Fabello et al., 2012; Leroy et al., 2013; Huang et al., 2014). In particular, the relationship between gas fraction, defined as $M_{HI}/M_\star$, and stellar mass is well studied, becoming an important constraint for theoretical predictions of star formation in the local

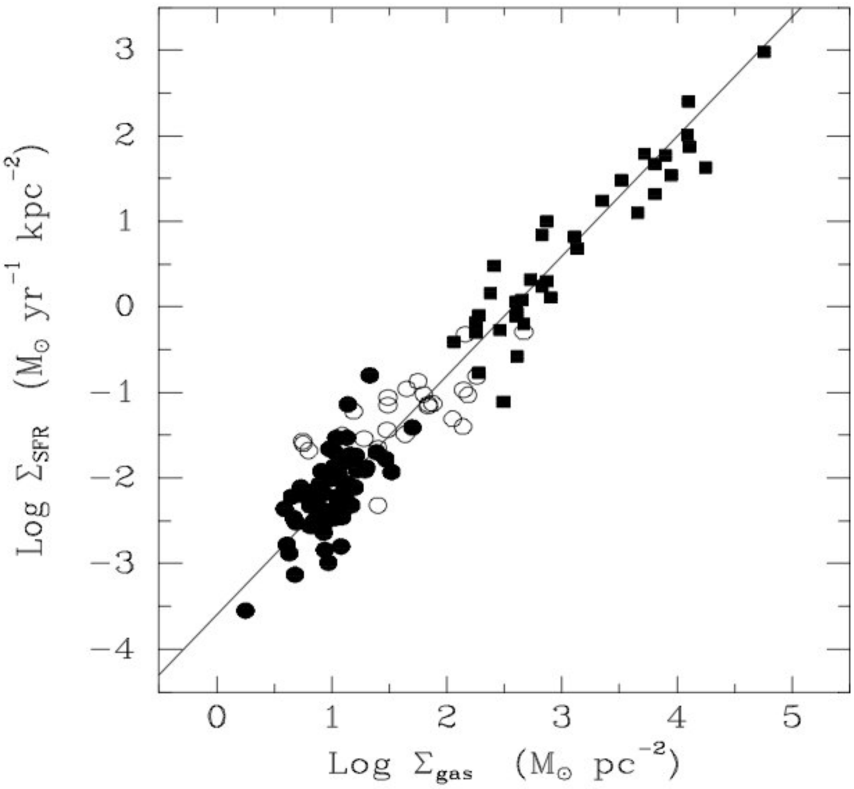

Figure 1.4: The global Kennicutt-Schmidt law as characterised by the relationship between total (H I + H_2) gas mass density, Σ_{gas}, and star formation rate surface density, Σ_{SFR}. This figure is taken from Kennicutt (1998a).

Universe (Catinella et al., 2013; Davé et al., 2013). However, while it is known that the complex relationships between these parameters are regulated by the processes that govern galaxy evolution and the timescales over which they occur, the extent of the interplay between global galaxy properties and the cold gas component remains unclear.

To date, most H I observations of galaxies and their environment have been focused either in the field, or in the high-density cluster regime. In a series of seminal papers throughout the 1980's, Haynes and Giovanelli presented the evidence that externally driven processes play a major role in the suppression of gas content within clusters (Giovanelli et al., 1981, 1982; Giovanelli & Haynes, 1983; Haynes et al., 1984; Giovanelli & Haynes, 1985a,b; Haynes & Giovanelli, 1986; Giovanelli & Haynes, 1989). A steady stream of works have followed and we are now at a point where the deficiency of gas in the very densest regions of the local Universe is well studied (Kennicutt, 1983; Abadi et al., 1999; Moore et al., 1999; Bravo-Alfaro et al., 2000; Chengalur et al., 2001; Solanes et al., 2001, 2002; Kenney et al., 2004; Hester, 2006; Gavazzi et al., 2006; Boselli & Gavazzi, 2006; Boselli et al., 2006; Cortese et al., 2011). The VLA Imaging of Virgo in Atomic gas survey (VIVA; Chung et al., 2009) demonstrates in great detail the importance of cluster environments in shutting down star formation via strong gas depletion mechanisms. Although it has been explored less extensively, investigations into the intermediate density group environment have recently hinted that significant removal of gas, through harassment, tidal stripping and strangulation of H I reservoirs, begins to occur well before galaxies enter the cluster environment (e.g. Kilborn et al., 2009; Rasmussen et al., 2012; Hess & Wilcots, 2013; Catinella et al., 2013; Yoon & Rosenberg, 2015; Stark et al., 2016). Galaxies in isolation or voids are seen to exhibit H I normalcy or excess more often than those in higher densities (e.g. Haynes et al., 1984; Kreckel et al., 2012). Despite such studies, there remains a shortage of work investigating the impact of environment on the gas content of galaxies from a statistical, multi-wavelength perspective. Further work is required if we are to provide a complete explanation for the externally-driven suppression of gas content over the full range of galactic environments, revealing what mechanisms are at work and in which regimes.

From a theoretical viewpoint, semi-analytic models and hydrodynamical simulations have thus far provided a variety of frameworks for the H I content of galaxies (e.g. Croton et al., 2006; Obreschkow et al., 2009; Duffy et al., 2012; Davé et al., 2012; Popping et al., 2014; Gonzalez-Perez et al., 2014; Schaye et al., 2015). Both simulations and models are now capable of reproducing basic scaling relation between gas and stellar masses that are in agreement with H I observations (e.g. Lagos et al., 2011b; Davé et al., 2013; Stevens et al., 2016; Crain et al., 2017). Given that the depletion of gas content due to environment is the result of hydrodynamical and gravitational mechanisms, theoretical frameworks are also a powerful tool for revealing the prevalence of these phenomena and how gas removal rates change as a function of environment and time. In pursuit of this, clear theoretical predictions are now available for the second-order environmental dependencies of gas content (McCarthy et al., 2008; Rafieferantsoa et al., 2015; Marasco et al., 2016), however, an in-depth comparison with H I observations has so far not been possible due to a lack of large representative samples of galaxies across the entire environment regime.

1.4 H I Surveys and Spectral Stacking

Large area comprehensive surveys of extragalactic H I have historically lagged behind their optical counterparts in both statistics and depth. However, over the past decade or so, 21 cm surveys have dramatically increased the quantity and quality of data available, characterising the neutral hydrogen content for thousands of galaxies. The current collection of single-dish observations that provide statistical inventories of extragalactic gas content can be placed into two camps. Firstly, targeted surveys of optically selected galaxy samples that provide flux, mass or gas fraction limited measurements of atomic gas content. One such program is the GALEX Arecibo SDSS Survey (GASS; Catinella et al., 2010). GASS details unresolved H I properties for a representative sample of ∼800 galaxies selected only by redshift and stellar mass ($z \leq 0.05$; $\mathrm{M}_\star \geq 10^{10}\mathrm{M}_\odot$) from SDSS, observing each target until a 21 cm detection is achieved or a gas fraction ($\mathrm{M}_{HI}/\mathrm{M}_\star$) upper limit of ∼2 per cent is reached. The main advantage of GASS and other targeted H I surveys is the depth delivered by the long integration times and targeted observing strategy that allows investigation of low and distant gas content. This strength is also their weakness as observations, even for relatively small galaxy samples, are exceptionally time consuming, making it a difficult and inefficient method for statistical characterisation of H I content.

Secondly, blind H I surveys are capable of detecting much larger samples of galaxies than their targeted counterparts (a few 10^4). Keeping the integration time per beam solid angle (approximately) constant as the telescope *blindly* maps the sky means that survey speed, and therefore volume observed, can be greatly increased. Using an integration time of ∼450s beam^{-1}, the H I Parkes All Sky Survey (HIPASS; Meyer et al., 2004; Wong et al., 2006) scanned ∼75 per cent of the sky (30,000 deg^2) and detected ∼5000 nearby galaxies. Its successor, the Arecibo Legacy Fast ALFA (ALFALFA; Giovanelli et al., 2005) survey detects ∼30,000 galaxies over ∼7000 deg^2 of sky, with an integration time of ∼50s beam^{-1}. Arecibo's superior dish size (305m) and angular resolution (3.5 arcmin)

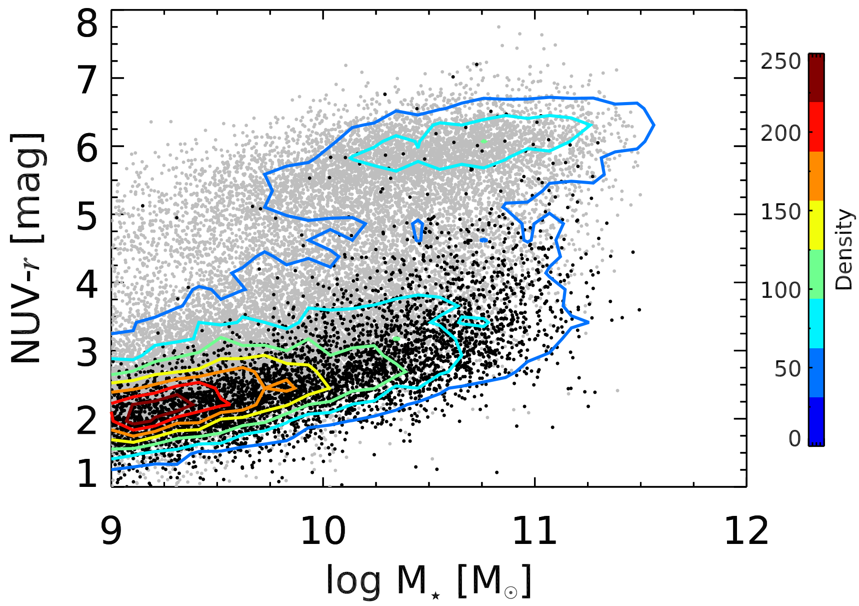

Figure 1.5: The colour-mass diagram for galaxies in our sample with NUV-r (see Chapter 2). Contours represent density levels from a minimum of 20 galaxies per bin up to 250 per bin. The points show positions in the plane for the full sample of 24,337 galaxies. Black points are the 4,610 galaxies detected by the ALFALFA 21 cm survey, grey points are undetected galaxies in the same ALFALFA observations.

means that ALFALFA is 8 times more sensitive and detects 20 times more sources per deg^2 than HIPASS. While obviously such large samples are extremely useful for statistical characterisation of H I content, the short integration times and subsequent shallow depth of blind surveys mean that they are biased towards detecting the gas-rich populations within their volume.

We highlight the bias of shallow H I surveys toward the H I rich population by plotting the galaxy bimodality for the representative sample of 24,000 nearby galaxies used in this thesis (see Chapters 2 and 4) in Figure 1.5. Galaxies that are detected by ALFALFA are denoted using black points while galaxies that are undetected are shown in grey. The superimposed number density contours trace the blue, star forming and red, quiescent sequences as function of stellar mass. This figure clearly shows that blind surveys such as ALFALFA preferentially detect blue, star forming objects (NUV-$r \leq 3.5$) and though red sequence detections do occur, they are rare. As well as preventing a representative quantification of gas content, this bias also severely hampers investigation of gas depletion due to environment (see Huang et al., 2012; Yoon & Rosenberg, 2015).

In an effort to overcome this limitation and push H I surveys significantly beyond their nominal sensitivity limit, studies have increasingly begun to exploit a signal processing technique called H I spectral stacking. By co-adding 21 cm line spectra, extracted from a given H I survey volume using optical position and redshift, stacking may be used to obtain statistical estimates of average H I properties for galaxies that are not necessarily individually detected in H I. Assuming a large enough number of accurate (spectroscopic) redshift measurements within the survey volume, one is able to select statistically powerful samples based upon stellar mass, rather than H I mass, which enables a representative view of gas content in the local Universe.

Indeed, even with next generation facilities such as the Square Kilometre Array and its pathfinders surveying vast volumes of the Universe with unprecedented sensitivity, these advantages mean that the most stringent constraints on average gas content for the largest galaxy samples will always come from the stacking of H I spectra.

The first stacking study to characterise the H I emission from galaxies was Chengalur et al. (2001), who showed that the average gas reservoir of galaxies in the central regions of a nearby cluster are lower than those on the outskirts. Since then, in an effort to probe the evolution of gas with cosmic time, stacking has mostly been employed to estimate variance in the H I mass density over the redshift range, $0 \lesssim z \lesssim 1$ ($\Omega_{\rm HI}$; Lah et al., 2007, 2009; Rhee et al., 2013; Delhaize et al., 2013; Rhee et al., 2016; Kanekar et al., 2016).

Other works have stacked intermediate redshift samples to quantify the amount of gas in the cluster environment above $z \sim 0.2$ (Verheijen et al., 2007; Lah et al., 2009; Stroe et al., 2015) and the typical H I-luminosity ratio out to $z \sim 0.1$ (Geréb et al., 2015).

The application of stacking to the gas content scaling relations in nearby galaxies is a relatively recent development. Fabello et al. (2011a), and their follow up works Fabello et al. (2011b, 2012), used the technique to quantify the scaling relations with stellar mass, concentration index, stellar surface density and NUV-r using a large multi-wavelength sample of ~5,000 massive galaxies ($M_\star \geq 10^{10}$ $M_\odot$). In addition to establishing the stacked H I scaling relations, the authors find that the total gas content of these galaxies is not driven by the size of their bulge nor the presence of active galactic nuclei once stellar mass is controlled for. They also provide tentative evidence for the effect of environment by showing that the mean gas fraction of galaxies with stellar masses between $10^{10} \leq M_\star/M_\odot < 10^{10.5}$ decreases as function of local density, moreover, this decrease is sharper than the corresponding decline in star formation. Likewise, Geréb et al. (2013) used stacking to perform a comparison of H I properties for star forming and quiescent galaxies populations in the Lockman Hole region, confirming the trend of blue galaxies being more H I-rich but highlighting that significant gas remains present in the red population.

The versatility of stacking is becoming ever more apparent with its application to different astrophysical questions. For example, Meyer et al. (2016) explore the possibility of extending the Tully-Fisher relation (rotational velocity vs. luminosity; Tully & Fisher, 1977) to greater distances and smaller galaxies, finding the stacked relation to be in good agreement with the individual detections. It must be noted that their method does not yet include non-detections, a crucial step if this approach is to reach its full potential. In absorption, Gereb et al. (2014) used stacking to quantify the gas content of radio galaxies.

1.5 Objectives of this Thesis

It should be clear from the discussion above that galaxy evolution is a subject of great complexity. As such, fundamental questions remain surrounding the role of H I in the life cycle of galaxies that this work aims to address: i) how does gas content regulate the star formation process and build up of stellar mass in galaxies?, ii) where and how precisely does external influence on H I reservoirs become important? and iii) how does the cycling of gas into and out of galaxies drive the star formation cycle?

If we are to translate our current phenomenological understanding into an successful model of galaxy evolution we must disentangle the complex relationships between galaxy gas content, star formation, metallicity and environment. In this context, *the primary goal of this thesis is to understand how internal and external processes combine to regulate gas content and drive the life cycle of nearby galaxies.*

1.6 Thesis Outline

To address our objectives we have adopted a multi-wavelength approach, covering a range of topics (from gas fraction scaling relations to the chemical analysis of galaxies), methods (from observations to theory) and datasets (from environment to emission line fluxes). This work is presented in 7 chapters, each of which is outlined below:

2. *Sample Characterisation:* As a starting point for this research, large, representative samples are required. In Chapter 2 we present the parent sample of ∼30,000 galaxies selected according to stellar mass from the overlap of SDSS and the ALFALFA volumes. We also fully describe the 3 subsets of this sample (A, B, C) used to tackle the different science questions in Chapters 4, 5, 6 and 7.

3. *H I Data and Spectral Stacking:* Chapter 3 describes all major aspects of the H I stacking technique, from the initial extraction, flagging and cleaning of 21 cm line H I spectra to the principle of stacking and its caveats.

4. *Disentangling the Key Gas Fraction Scaling Relations:* This investigation focusses on providing a framework of H I scaling relations deep enough to probe representative galaxy H I masses, and comprehensive enough in their scope that the independent influence of each variable on gas content may be established.

5. *Gas Stripping in Satellite Galaxies: from Pairs to Clusters:* In Chapter 5, we select satellite galaxies from the parent sample in order to determine the extent to which H I loss may be attributed to environment, what processes are at work and in which regimes.

6. *Comparing Observations with Theoretical Predictions:* In this chapter, we compare the results of Chapter 5 to the predictions of semi-analytic models and hydrodynamical simulations (Davé et al., 2013; Gonzalez-Perez et al., 2014).

7. *Gas as the Primary Regulator of the Mass-Metallicity Relation:* We stack H I spectra for ~10,000 nearby galaxies along the mass-metallicity relation, quantifying the relative importance of gas content and star formation as drivers of scatter and establishing the most physically motivated dependence of the mass-metallicity relation. We interpret our results in the context of theoretical predictions.

8. *Conclusions:* Finally, in Chapter 8 we summarise and review the conclusions of all the research presented in this thesis and discuss future work.

2
Sample Characterisation

Over the past two decades, the astronomical community has built up a vast collection of observations that are both representative of the local galaxy population and span the full range of galactic environments. In particular, large scale photometric and spectroscopic surveys such as the Galaxy Evolution Explorer surveys (GALEX; Martin et al., 2005; Morrissey et al., 2007) in the ultraviolet (UV), and the Sloan Digital Sky Survey (SDSS; York et al., 2000) in the optical, cover a significant fraction of the sky and provide imaging and spectra for over a million galaxies.

In order to answer the questions outlined in the introduction, this thesis requires a dataset with wavelength coverage spanning the UV to the radio. Building upon SDSS photometry and spectroscopy, we use UV photometry from GALEX and H I spectral line data from ALFALFA. We also make use of archival stellar mass, star formation, metallicity and environment data from SDSS value-added catalogues. Combining multi-wavelength survey efforts in this way allows an in-depth and statistical approach to the study of galaxy properties and environments.

All optical, UV and derived quantities used in this work are presented in this Chapter. In Section 2.1 we describe the parent sample selection criteria, Section 2.2 is a summary of the various sample subsets used to address the science questions in this thesis, an overview of the optical and UV surveys and data used in this thesis is given in Sections 2.3 and 2.4 respectively. Section 2.5 is a description of derived stellar mass, star formation rate (SFR) and metallicity estimates and, finally, Section 2.6 is a discussion of the different metrics used to define the environment of a galaxy. All H I line data is characterised in Chapter 3.

Where relevant, all astrophysical quantities in this thesis were either derived using or converted to a Chabrier (2003) initial mass function (IMF). Throughout this thesis the distance dependent quantities are computed assuming a ΛCDM cosmology with $\Omega_{\rm M} = 0.3$, $\Omega_\Lambda = 0.7$, and a Hubble constant $\rm H_0 = 70$ km s^{-1} Mpc^{-1}.

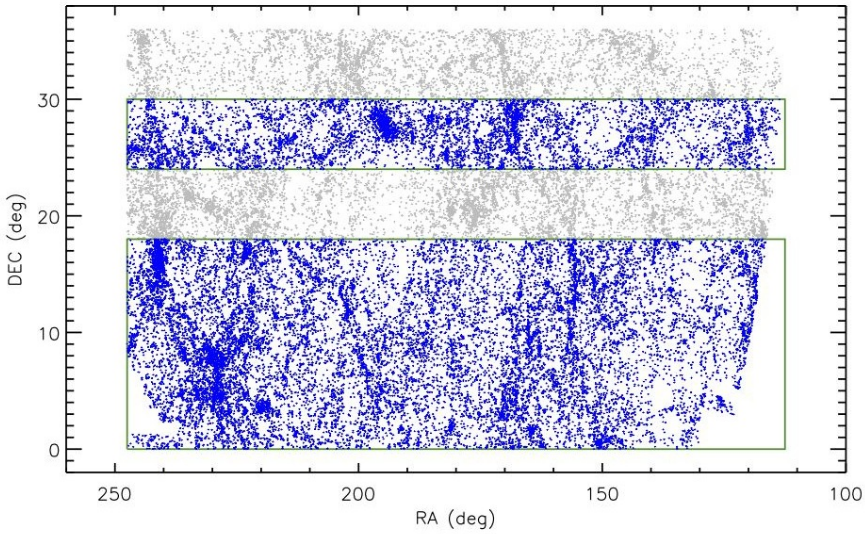

Figure 2.1: ALFALFA 'spring sky' footprint for which data cubes were processed in 2013 (green boxes). This covers two sky areas between 112.5° (7h30m) to 247.5° (16h30m) in right ascension, and from 0° to 18° and 24° to 30° in declination. This is overlaid on SDSS galaxies (grey points) within the same sky region, redshift range $0.01 \leq z \leq 0.05$ and stellar mass $10^9 M_\odot \leq M_\star \leq 10^{11.5} M_\odot$. Galaxies that overlap are shown in blue.

2.1 Parent Sample Selection Criteria

Using the publicly available SDSS Data Release 7 (DR7; Abazajian et al., 2009), we select a volume-limited parent sample of galaxies for which photometry and spectroscopy are available. The footprint of this sample is shown by the green boxes in Figure 2.1 and is restricted to the sky area for which ALFALFA data cubes had been processed by February 2014. For a full description of the ALFALFA survey and data, we refer the reader to Chapter 3.

The stellar mass, $M_\star$, and volume selection criteria for this sample are as follows:

$$10^9\ M_\odot \leq\ M_\star \leq 10^{11.5}\ M_\odot$$
$$112.5^\circ \leq \alpha \leq 247.5^\circ$$
$$0^\circ \leq \delta \leq 18^\circ\ \&\ 24^\circ \leq \delta \leq 30^\circ$$
$$0.02 \leq z \leq 0.05$$

This selection yields 30,368 galaxies.

The stellar mass distribution of the parent sample is shown by the black histogram in Figure 2.2a. Selecting galaxies with stellar mass greater than $10^9\ M_\odot$ ensures we are representative of star forming or quiescent populations above this mass and that our

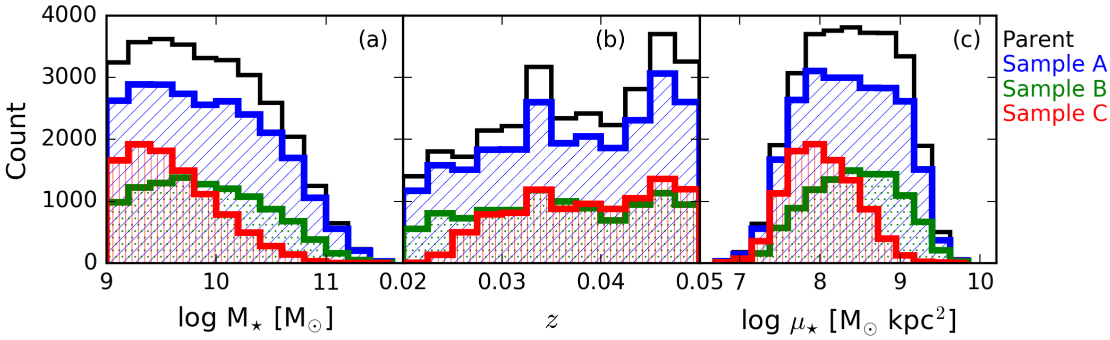

Figure 2.2: Distribution of sample galaxies across stellar mass (a), redshift (b) and stellar surface density (c, Equation 2.3). The black histograms show the parent sample. The blue, diagonally hatched histograms represent the 24,337 galaxies used in Chapter 4 that are referred to as Sample A throughout this thesis. The green dotted histograms show the 10,567 satellite galaxies used in Chapters 5 and Chapter 6 (Sample B). The 9,720 star forming galaxies that are used in Chapter 7 (Sample C) are denoted by the red, vertically hatched histograms.

sample straddles the 'transition mass' in galaxy populations of $\sim 3\times10^{10}$ $M_{\odot}$ identified by Kauffmann et al. (2003b).

The redshift distribution of the parent sample is shown in Figure 2.2b (black histogram). We avoid the significant contribution of peculiar velocities to galaxy redshift measurements by setting a lower redshift limit of $z = 0.02$ while the frequency range over which the San Juan airport radar affects Arecibo's observation of redshifted H I emission is removed by the ceiling of $z = 0.05$.

2.2 Summary of Samples

For the analysis in this thesis, we select 3 subsets of our parent sample where accompanying multi-wavelength and value-added data are best suited to addressing the question at hand. We refer to these subsets as Sample A (Chapter 4), Sample B (Chapters 5 and 6) and Sample C (Chapter 7). The stellar mass, redshift and stellar surface density distributions of each subset are shown in Figure 2.2.

In order to provide a reference point for the reader and some context with which to understand the data described in this chapter, here we present a very brief overview of the various multi-wavelength subsets used in this thesis, the criteria with which they were defined and the chapters where they are used. Table 2.1 provides the reader with a quick reference to each sample and the quantities used in their selection while a more detailed description of each subset is also presented in the relevant Chapter.

- **Parent Sample:** A volume-limited, representative sample of 30,368 galaxies selected according to stellar mass ($M_\star/M_\odot \geq 10^9$) and redshift ($0.02 \leq z \leq 0.05$) from the intersection of the SDSS and ALFALFA surveys. All galaxies have high quality optical (SDSS) and H I 21 cm (ALFALFA) data.

- **Sample A:** 24,337 galaxies selected from the parent sample to have accompanying GALEX NUV photometry. This sample forms the basis of the analysis in Chapter 4.

- **Sample B:** For the analysis in Chapters 5 and 6, we focus only on satellite galaxies (see Section 2.6) and restrict the parent sample to objects for which the full complement of environment data and Brinchmann et al. (2004) total SFRs are available, 10,567 in total.

- **Sample C:** We construct a sample of 9720 star forming galaxies with $10^9 \leq M_\star/M_\odot \leq 10^{11}$ for which Mannucci et al. (2010) and Tremonti et al. (2004) metallicity estimates are available. In addition, we require all galaxies in this sample to have Brinchmann et al. (2004), Salim et al. (2016) and Hα SFR measurements. The resulting sample is used to investigate the mass-metallicity relation in Chapter 7.

Table 2.1: Reference table for the 3 samples used in this thesis and the key criteria by which they were selected. All the data are described in detail in Chapters 2 and 3. In order, the columns are i) sample name, ii) number of galaxies, iii) stellar mass cut, iv) GALEX cut, v) environment cut, vi) gas-phase metallicity cut and vii) accompanying notes. N/A means that no selection criteria were applied on this parameter.

Sample	N	$M_\star/M_\odot$	NUV	$M_{\rm halo}$	$Z_{\rm gas}$	Notes
Parent	30,368	$\geq 10^9$	N/A	N/A	N/A	Stellar mass selected from the overlapping volume of SDSS and ALFALFA.
A	24,337	$\geq 10^9$	$1 \leq \mathrm{NUV}-r \leq 8$	N/A	N/A	Selected to have accompanying near-UV GALEX data (96% MIS, 4% AIS).
B	10,567	$\geq 10^9$	N/A	Satellite galaxies only	N/A	Halo masses and satellite designation are taken from Yang et al. (2007) SDSS group catalogue.
C	9,720	$10^9 \leq M_\star \leq 10^{11}$	N/A	N/A	Valid Tremonti et al. (2004) and Mannucci et al. (2010) log (O/H) + 12 estimates	By definition, this sample contains only star forming galaxies.

2.3 SDSS Data

All optical data are taken from SDSS, a major imaging and spectroscopic campaign that saw first light 17 years ago. Since then SDSS has observed almost 1 million galaxies across 10,000 deg^2 of the sky, making it the most comprehensive galaxy survey ever undertaken. The 'main' galaxy sample consists of photometry taken in five broad-bands, *u, g, r, i, z* (AB system; Fukugita et al., 1996), using a drift scan, wide-field camera mounted on the dedicated 2.5 meter Sloan telescope located at Apache Point Observatory, New Mexico (Gunn et al., 1998). After astrometric and photometric calibration (Hogg et al., 2001; Pier et al., 2003), imaging catalogues are used to identify objects for spectroscopic follow up with two 320 fibre-fed multi-object spectrographs. Target selection for the main spectroscopic galaxy sample is based upon the r-band Petrosian magnitude ($r \leq 17.77$ mag) and half-light surface brightness ($\mu_{50} \leq 24.55$ mag arcsec^2). Further details are given in Strauss et al. (2002).

2.3.1 SDSS Photometry

Optical magnitudes and sizes for the 30,368 galaxies in our parent sample are obtained from the SDSS DR7 database server *CasJobs*[1] via Structured Query Language queries. Magnitudes are taken from each of the *u, g, r, i, z* bands. For magnitudes and colours we use the 'model' magnitudes. These are the optimal fit of either a pure de Vaucouleurs or a pure exponential profile to the galaxy flux in each band. All optical photometric data are corrected for Galactic extinction following Schlegel, Finkbeiner, & Davis (1998) and using the respective band extinction provided in the SDSS DR7 *PhotObj* catalogue. We compute distance measurements and rest-frame shift H I spectra using fibre spectroscopic redshifts.

2.3.2 SDSS Fibre Spectroscopy

The 640-fibres of the two SDSS spectrographs cover a wavelength range of 3800-6150Å (blue spectrograph) and 5800-9200Å (red spectrograph) at an average resolution of R=$\lambda/\Delta\lambda \simeq$ 2000. Each fibre has an angular diameter of 3 arcsec. Spectra for target galaxies are extracted, calibrated and classified by the SDSS pipeline, and redshifts are estimated with an accuracy of ~ 30 km s^{-1}. We use CasJobs SQL queries to select spectroscopic redshifts for all of our objects.

[1] `http://skyserver.sdss3.org/casjobs/`

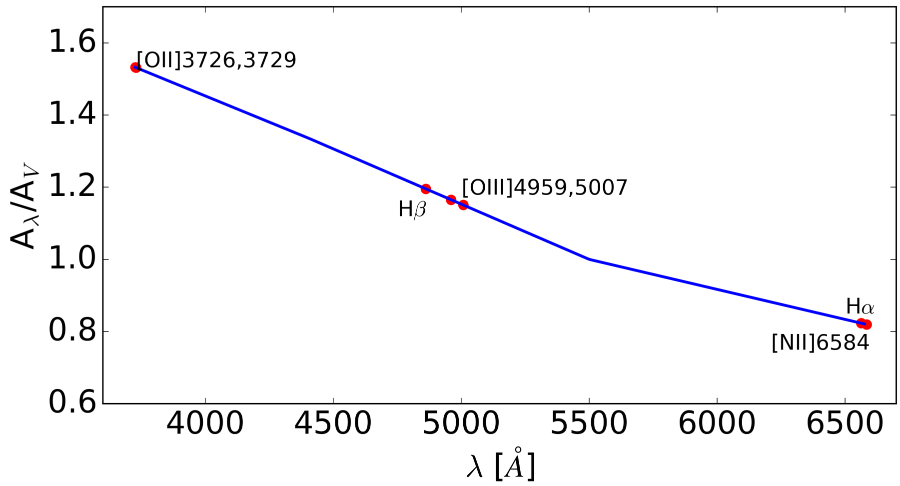

Figure 2.3: The blue line shows the Cardelli, Clayton, & Mathis (1989) reddening law with the wavelengths of the emission lines used in this thesis marked by red circles.

Spectral line measurements for 29,928 galaxies in our parent sample are obtained by cross-matching with the publicly available Max Planck Institute for Astrophysics-Johns Hopkins University (MPA-JHU) reduction of SDSS DR7 spectra[2]. The methodology used by the MPA-JHU group is described in Tremonti et al. (2004), hereafter T04. The authors employ a sophisticated pipeline customised for SDSS spectra to derive continuum-subtracted emission line fluxes. The procedure uses a least-squares fit to determine the 'best' stellar population synthesis model (Bruzual & Charlot, 2003) and dust attenuation. For each galaxy, the best fitting model of the stellar continuum is subtracted from its spectrum (along with any residuals) and each of the optical Balmer lines (Hδ, Hγ, Hβ, Hα) and eight forbidden lines ([OII]λ3726, [OII]λ3729, [OIII]λ4959, [OIII]λ5007, [NII]λ6548, [NII]λ6584, [SII]λ6717, [SII]λ6731) is simultaneously fit with a Gaussian.

The required dust extinction correction for each galaxy, or more accurately the HII regions within them, is calculated using the ratio of the lower Balmer lines, $F(H\alpha)/F(H\beta)$, known as the Balmer decrement. This method relies upon the relative simplicity of the hydrogen atom and our understanding of radiative ionisation followed by recombination to constrain the expected intensity ratios between the Balmer lines. Under the assumption of certain interstellar medium (ISM) temperature and density conditions, increasing of the observed Balmer decrement with respect to its predicted value can then be used to determine the reddening due to dust. We assume 'Case B' conditions where $F(H\alpha)/F(H\beta)$=2.86 at a temperature of T=10^4 K and column density of n_e $\sim$$10^2$-$10^4$

[2] http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/raw_data.html

cm^{-3} (Osterbrock, 1989).

The dust attenuation in magnitudes at a given wavelength is calculated by combining this information with an 'extinction law' such as the Cardelli et al. (1989) reddening curve used in this thesis (Figure 2.3):

$$A_\lambda/A_V = a(x) + b(x)/R_V \tag{2.1}$$

where $a(x)$ and $b(x)$ are the wavelength-dependent coefficients provided in Cardelli et al. (1989) and $R_V = A_V/E(B-V) = 3.1$ is the value of total to selective V band (5500Å) extinction for the Milky Way. The ratio between the observed and expected Balmer decrement is used to calculate the extinction in the V-band, A_V:

$$\begin{aligned} A_V &= \frac{2.5}{(A_{H\beta}/A_V) - (A_{H\alpha}/A_V)} \log \frac{F(H\alpha)/F(H\beta)}{2.86} \\ &= 6.71 \log \frac{F(H\alpha)/F(H\beta)}{2.86} \end{aligned} \tag{2.2}$$

The reader should note that the extinction law used here is different to the one employed to correct the photometry (Schlegel et al., 1998). The reason for this is that it is important for both the spectroscopic and photometric reductions in this thesis to follow the respective methods of Mannucci et al. (2010) and Catinella et al. (2013) as closely as possible.

For the analysis in Chapter 7, we discard 10,638 systems where emission line flux is contaminated by a contribution from active galactic nuclei (AGN) using the Kauffmann et al. (2003c) classification and [NII]6584/Hα < 0.6 cut on the Baldwin et al. (1981, BPT) diagram shown in Figure 2.4. Of the remaining 18,200 objects, 954 have Balmer decrements smaller than 2.5 and, since this is unphysical for these galaxies, we exclude them from our analysis (Kewley & Dopita, 2002; Salim et al., 2014). For galaxies with Balmer decrements between 2.5 and 2.86, we assume 2.86 and do not apply a correction. The number of star forming galaxies with emission line measurements in our parent sample is therefore 17,247.

2.4 GALEX Photometry

Since its launch in 2003, NASA's GALEX telescope has observed most of the sky in the far (FUV, 1344 - 1786Å) and near (NUV, 1771 - 2831Å) UV bands at a typical resolution of 4.5 and 6 arcsec respectively. The sample used in this thesis uses NUV data from two imaging surveys conducted as part of the GALEX mission. The first is the All-sky Imaging

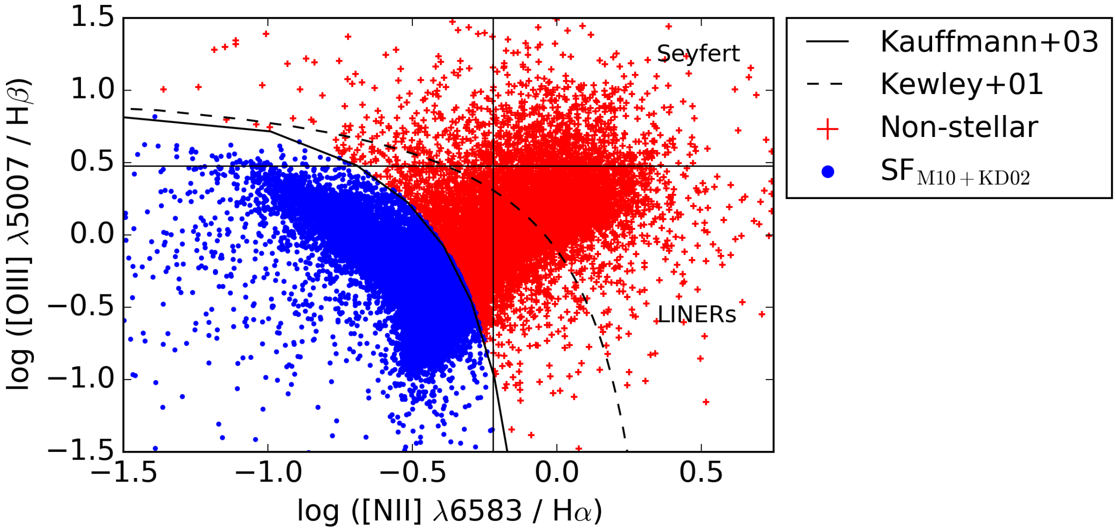

Figure 2.4: The Baldwin, Phillips, & Terlevich (1981, BPT) diagram for 28,838 galaxies in our sample for which flux measurements in all four emission lines are present. The solid line denotes the criterion for galaxy emission to be considered star forming (18,200, blue points) or AGN contaminated (10,638 red crosses) as defined by Kauffmann et al. (2003c). The blue points show the star forming sample for which we calculate the Mannucci et al. (2010) metallicity diagnostic (see Section 2.5.3). The dotted line shows the slightly less conservative cut given by Kewley et al. (2001) that is used in the Kewley & Dopita (2002) calibration.

Survey (AIS) which covers over 26,000 $\deg^2$ of sky down to a NUV-band depth of m_{AB} = 20.8 mag in NUV (FUV-band m_{AB} = 19.9 mag). Secondly, the Medium Imaging Survey (MIS) has observed ∼1000 $\deg^2$ in the footprint of the SDSS spectroscopic sample down to a NUV-band sensitivity $m_{AB} \simeq 23$ mag. For more details on both surveys, see Martin et al. (2005). GALEX NUV observations are available for 24,337 galaxies in our parent sample. Hereafter, this subset of galaxies is referred to as 'Sample A'. For more details see Section 2.2.

The UV data are obtained by cross-matching the GALEX Unique Source Catalogues[3] (GCAT; Seibert et al., 2012) and the Bianchi, Conti, Shiao Catalogue of Unique GALEX Sources[4] (BCS; Bianchi et al., 2014) to the SDSS database using an impact parameter of ten arcsec. The two UV catalogues use very similar techniques and as such any differences between their photometries are small ($\leq$ 0.1 mag). Where galaxies are present in both catalogues we selected magnitudes from the GCAT as this has a larger overlap with our sample. GCAT does not include the GALEX GR7 data release and so, when necessary, we also draw values from the BCS catalogue. For the majority of galaxies (96 per cent) with NUV data available, we use magnitudes obtained from the MIS, and for the remaining 4 per cent we use the AIS.

Reddening correction for NUV photometry is according to Wyder et al. (2007), who adopt $A(\lambda)/E(B\text{-}V) = 2.751$ for SDSS r-band and $A(\lambda)/E(B\text{-}V) = 8.2$ for GALEX NUV. Thus we convert to NUV extinction, A_{NUV}, using $A_{NUV} - A_r = 1.9807A_r$, where A_r is the r-band extinction.

The distributions of the stellar mass, redshift and stellar surface density for the parent sample (solid black line) and of the galaxies with NUV data (hatched blue histograms) are given in Figure 2.2. The ratio between sample galaxies with GALEX data and their parent sample is similarly uniform in distribution across both parameters. This illustrates that, when removing galaxies for which there are no valid NUV magnitudes, we do not inadvertently introduce bias by over or under sampling a region of the parameter space. Figure 2.5 is the NUV-r colour histogram for the sample and exhibits the well known blue cloud and red sequence bimodality (Baldry et al., 2004; Wyder et al., 2007; Schiminovich et al., 2007). The NUV-r colour is chosen as it probes young and old stellar populations on either side of the 4000Å break and thus is a sensitive tracer of a galaxy's specific SFR (sSFR=SFR/$M_\star$).

[3] `http://archive.stsci.edu/prepds/gcat/`
[4] `http://archive.stsci.edu/prepds/bcscat/`

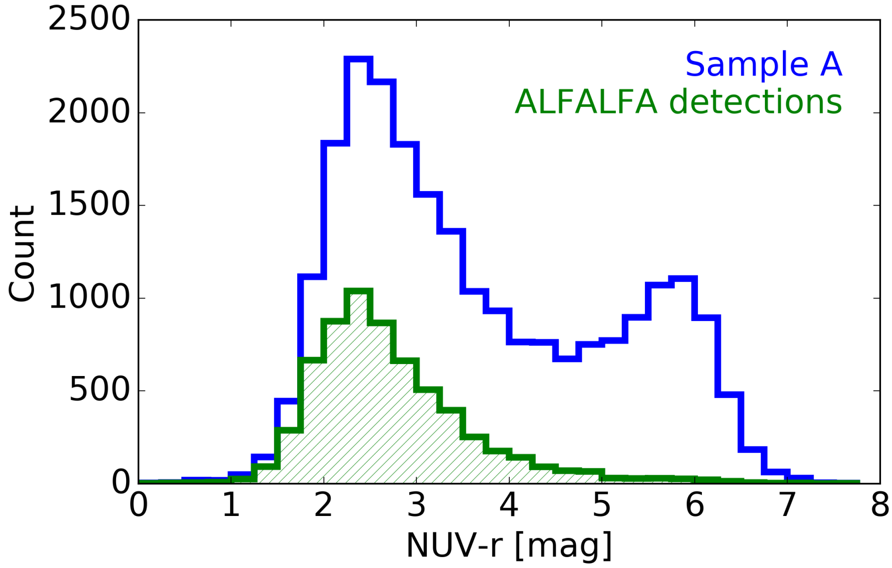

Figure 2.5: The NUV-r distribution of the 24,337 galaxies in our parent sample for which NUV data are available (Sample A) is shown by the blue histogram. The green, hatched histogram shows only those galaxies in our sample that are detected by ALFALFA.

2.5 Derived Quantities

2.5.1 Stellar Masses

All stellar masses are taken from the value-added MPA-JHU[5] catalogue. They are derived using a Bayesian approach to spectral energy distribution (SED) fitting developed by Kauffmann et al. (2003a). SED fitting is based upon the principle that encoded into a galaxy's observed SED is a large amount of information on the unresolved stellar populations. The technique's applications and limitation are discussed at length in the literature, comprehensive reviews of which are provided by Walcher et al. (2011) and Conroy (2013). The basic method of Kauffmann et al. (2003a) is to pre-compute a library of model star formation histories spanning the full range of physically plausible scenarios, including stochastic bursts. Bayesian analysis is applied to associate a model SED and parameters to the observational data. Comparisons are conducted with the SDSS broadband photometry to estimate reddening due to dust. Finally, stellar mass estimates are computed by multiplying the dust-corrected luminosity by the predicted mass-to-light ratio from the best fit SED model, weighted by its probability. The advantage of the Bayesian approach

[5] `http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/` using updated stellar masses from `http://home.strw.leidenuniv.nl/~jarle/SDSS/`. This update improves handling of poor u and z photometry in the MPA-JHU pipeline, see website for details.

over other best fit methods (likelihood, minimum χ^2) is that it provides the probability density function of galaxy properties, enabling a more rigorous characterisation and estimation of uncertainties. For this work, we select the median values of the stellar mass probability density function. The authors derive these stellar masses assuming a universal Chabrier (2003) initial mass function (IMF).

From the stellar mass, $M_\star$, we derive the stellar surface density, $\mu_\star$, as:

$$\mu_\star = \frac{M_\star}{2\pi R_{50,z}^2} \tag{2.3}$$

where $R_{50,z}$ is the Petrosian radius containing 50 per cent of the flux in z band, expressed in kpc.

2.5.2 Star Formation Rates

The primary estimate of galaxy SFRs used in this thesis also come from the MPA-JHU group's SDSS DR7 analysis. Following the methodology of Brinchmann et al. (2004, hereafter B04), SFRs are calculated within the galaxy region covered by the SDSS fibre (3 arcsec) using Hα emission line modelling where available for star forming galaxies. Where no or low signal-to-noise ratio (S/N(Hα) < 3) emission lines are present or there is a strong AGN contribution, B04 compute SFRs using the empirical relationship between sSFR and the strength of the break at 4000 Å (D4000), based upon the star forming galaxies. Aperture corrections are applied by performing SED fits to the broad-band photometry to derive SFRs for the area of the galaxy not covered by the fibre, following Salim et al. (2007). These 'out-of-fiber' SFRs are summed with the fibre SFRs so that a global value is recovered. There are 29,927 galaxies in our parent sample with MPA-JHU SFRs.

In Chapter 7, we follow Kennicutt (1998a) in deriving fibre SFRs directly from the Hα line flux for the 17,197 galaxies where S/N(Hα) > 3 and emission is not contaminated by AGN. Hα is a reliable tracer of the star formation because the presence of HII regions scales with the UV flux from OB stars. As described in Section 2.3.2, Hα fluxes are taken from the MPA-JHU catalogue and corrected for dust attenuation using the Balmer decrement method assuming the extinction curve of Cardelli et al. (1989). Once corrected, the fluxes are used to estimate the SFR within the fibre following Equation 2 in Kennicutt (1998a):

$$\mathrm{SFR}_{\mathrm{H}\alpha} = 7.9 \times 10^{-42} L(\mathrm{H}\alpha) \tag{2.4}$$

where the SFR is expressed in units of $M_\odot\,yr^{-1}$ and $L(H\alpha)$ is the Hα luminosity in ergs s^{-1}. Following Madau & Dickinson (2014), we adjust the Hα SFRs by a factor of 0.63 to convert the IMF from Salpeter (1955) to Chabrier (2003).

As a final check, we adopt two additional SFR estimates as a point of comparison with the MPA-JHU values when investigating trends with star formation rate in Chapter 7. The first of these are extracted from the GALEX-SDSS-WISE Legacy Catalog (GSWLC; Salim et al., 2016, hereafter S16). GSWLC is a database of stellar masses, SFRs and dust attenuation properties for ~700,000 SDSS DR10 galaxies out to $z \sim 0.3$ (Ahn et al., 2014). S16 derive these quantities using a similar methodology to the Bayesian approach of B04, this time SED fitting to optical *and* UV fluxes using a modified dust attenuation law (Conroy et al., 2010) and emission line modelling. For star forming systems, the SFRs of the GSWLC and MPA-JHU catalogue agree well with one another (0.18 dex). The agreement between the two methodologies is significantly worse for galaxies classified as AGN or exhibiting low S/N Hα emission (~1 dex). S16 that find their approach, in particular the use of a variable dust attenuation law, produces more realistic SFRs in this regime. A comparison of the GSWLC measurements with other SFR estimates is available in Section 8 of S16. For our sample, we select UV-optical SED fit SFRs for 27,289 galaxies from the GSWLC.

Figure 2.6 compares the three SFR estimates described above. We only plot the ~15,000 galaxies that have all three indicators (black points) which, due to the requirement for high S/N star forming Hα in the K98 method, are dominated by blue sequence objects. Panel (a) shows the excellent agreement between the B04 and S16 total SFR estimates. Given the similarity of their approach, their tight correlation is not surprising and was noted in the original S16 paper. Panels (b) and (c) demonstrate that, due to the smaller covering fraction of the fibre, the K98 SFRs are biased low, typically by a factor of 4-5, with respect to estimates of global star formation. Despite this, the ranking of each SFR is consistent, with a strong correlation between methods, particularly for the more star forming systems.

2.5.3 Metallicity Estimates

The metal abundance of interstellar gas in galaxies is commonly estimated using emission line spectroscopy. In star forming galaxies, the UV radiation field emitted by hot, young stars embedded in gaseous nebulae causes photoionisation and recombination of atoms in the gas. The photons emitted from the cascade of elections as they recombine are imparted with information as to the chemical elements from which they originated. In the optical,

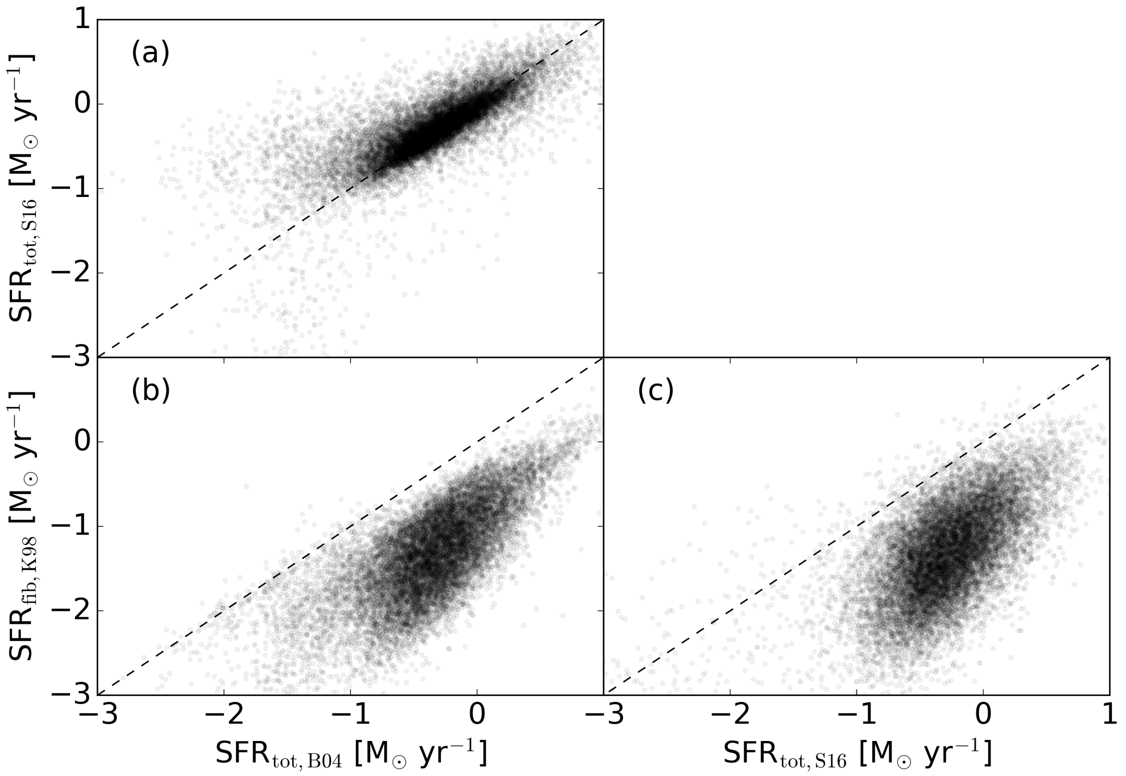

Figure 2.6: Comparison of the B04, K98 and S16 SFR estimates. Black points show the 15,363 parent sample galaxies for which all three indicators are available.

the emission lines typically used for spectroscopic analysis are two of the Balmer lines of hydrogen, Hβ and Hα (λ4861 and λ6563 respectively), as well as the forbidden strong lines of [OII]λ3726, [OII]λ3728, [OIII]λ5007, [NII]λ6584, [SII]λ6717 and [SII]λ6731.

Oxygen abundance (the ratio of oxygen to hydrogen, log (O/H) + 12) is frequently used as a tracer of the overall metal content of the gas for the following reasons: i) by mass, oxygen is the most abundant element after hydrogen and helium, ii) its strong line emission is at optical wavelengths, iii) it is observed in three ionisation states and iv) it is a good probe of local ISM conditions. The constant 12 is added as a matter of convention.

Calibrations between emission line ratios and the gas-phase oxygen abundances are typically derived using either theoretical results from photoionization models (e.g. Kewley & Dopita, 2002), empirical methods based on measurements of electron temperature (e.g. Pettini & Pagel, 2004), or a combination of both these approaches (e.g. Maiolino et al., 2008). There has been extensive discussion on the merits and drawbacks of each method and its associated metallicity calibrations in the literature. As it is beyond the scope of this thesis to conduct a detailed comparison of the various metallicity estimates, the reader is referred to the discussions in Kewley & Ellison (2008), Maiolino et al. (2008) and Salim et al. (2014).

For this thesis, we use three estimators of the gas-phase oxygen abundance to trace the metallicity of our galaxies: i) the combined theoretical and empirical calibration used in Mannucci et al. (2010, hereafter M10) that is based upon the Maiolino et al. (2008) average of the R23 and N2 methods (see Equations 2.5.3 and 2.5.3 respectively); ii) the theoretical Kewley & Dopita (2002, hereafter KD02) calibration that uses the relationship between [NII]/[OII] and (O/H); and iii) a Bayesian estimate from chemical evolution models derived by T04.

M10 Calibration

In order to ensure a proper comparison, we follow selection criteria of M10. Starting from the dust-corrected sample of 18,200 star forming galaxies described in Section 2.3.2, we select 15,667 objects that have: i) an Hα S/N of at least 25 and ii) detections in the Hβ, [OII]λ3726, [OII]λ3729, [OIII]λ4959, [OIII]5007, [NII]λ6584 lines. The high S/N cut on the Hα line is applied following M10 to ensure that fluxes are typically detected across the other, weaker optical lines, however, we find that the additional requirement for detections in each line was also necessary to ensure only reliable measurements were selected. Note that for this calibration we adjust the Balmer decrement cut given in Section 2.3.2 to follow M10. Hence we remove a further 244 galaxies where $F(H\alpha)/F(H\beta)$ <2.5 and extremely high reddening is present in the V-band (A_V >2.5), leaving 15,423 galaxies.

To ensure that the SDSS 3 arcsec fibre covers a significant fraction of each galaxy in their sample, M10 implement a redshift cut ($0.07 < z < 0.3$) that does not overlap in redshift with our sample. Although we cannot apply this cut for the main analysis in Chapter 7, in consideration of the potential biases, we are able to reproduce the M10 MZR using the full SDSS (without the accompanying H I data) with their redshift cut and provide a discussion of aperture effects in Section 7.4 of Chapter 7.

The procedure followed by M10 is based upon the calibrations of Maiolino et al. (2008), which were derived using a combination of theoretical metallicity estimates for SDSS galaxies (Kewley & Dopita, 2002) and empirical relations between strong line ratios and the oxygen abundance. Following M10, we use an average of the R_{23} and N2 estimates. The former was first formulated by Pagel et al. (1979) and is the sum of [OII] and [OIII] intensities relative to Hβ, defined as:

$$\mathrm{R}_{23} \equiv \log \frac{[OII]\lambda3726 + [OII]\lambda3729 + [OIII]\lambda4959 + [OIII]\lambda5007}{H\beta} \tag{2.5}$$

Following Storchi-Bergmann et al. (1994), the latter is given as:

$$\mathrm{N2} = \frac{[NII]\lambda6584}{H\alpha} \tag{2.6}$$

We then discard galaxies that have values of R_{23} and N2 based estimates of log (O/H) + 12 that differ by more 0.25 dex and/or where R_{23} and N2 fall outside of the range of acceptable values (R_{23} <0.9, N2<-0.35; Maiolino et al., 2008). The valid R_{23} and N2 values are then used to estimate the gas-phase oxygen abundance (log (O/H) + 12) by finding the roots of the polynomial relationship given by Maiolino et al. (2008):

$$\log R = a + bx + cx^2 + dx^3 + ex^4 \tag{2.7}$$

where R is the strong line calibration and x is the abundance relative to solar (x = log (O/H) + 12 − 8.69). The R_{23} method's constants are: $a = 0.7462; b = -0.7149; c = -0.9401; d = -0.6154; e = -0.2524$. For the N2 method: $a = -0.7732; b = 1.2357; c = -0.2811; d = -0.7201; e = -0.3330$. The typical scatter between the two estimates is ∼0.1 dex and the average is taken to obtain a final M10 metallicity estimate for each galaxy. There are 12,208 galaxies with reliable M10 metallicities in the parent sample.

KD02 Calibration

The calibrations of KD02 are derived using stellar population synthesis and photoionization models to determine the theoretical emission line ratios that arise from a given set of assumptions (i.e. conditions of star formation, electron density, IMF). They find the ratio of [NII]/[OIII] to be their best diagnostic because it correlates well with the gas-phase abundance while remaining relatively insensitive to the ionisation state of the gas (characterised by the ionisation parameter). The quoted accuracy of this measurement is ∼0.1 dex.

For log [NII]/[OIII] > -1.2, the polynomial KD02 calibration is given as:

$$\log \frac{[NII]}{[OIII]} = a + bz + cz^2 + dz^3 + ez^4 \tag{2.8}$$

where z = log (O/H) + 12 and the constants are: $a = 1106.8660; b = -532.15451; c = 96.37326; d = -7.8106123; e = 0.23928247$. When log (O/H) + 12 < 8.4 (log [NII]/[OIII] < -1.2), KD02 recommend using an average with the R_{23} parameter (lower branch), however, this metallicity range does not apply for our galaxies.

Our sample selection when calculating the KD02 metallicities follows the revised criteria of Kewley & Ellison (2008) in requiring a S/N of at least 8 in the following lines: [OII]λ3726, [OII]λ3728, Hβ, [OIII]λ5007, Hα, [NII]λ6584, [SII]λ6717 and [SII]λ6731. These authors also require that the 3 arcsec fibre of SDSS has a covering fraction of > 20 per cent on all their galaxies. This is based upon Kewley et al. (2005), who show that this cut leads to fibre metallicities that are representative of the global value. Although we do not follow this cut in our main analysis, we do investigate its effect on our results in Chapter 7. The number of galaxies with valid KD02 metallicities in our parent sample is 7670.

T04 Metallicities

T04 provide log (O/H) + 12 estimates for galaxies that are classified as star forming in the public MPA-JHU catalogue (Brinchmann et al., 2004). Their method differs significantly from other metallicity estimates in that it uses Bayesian analysis of theoretical model fits to the continuum-subtracted spectra described in Section 2.3.2 rather than strong line ratios calibrated to metallicity. The full methodology is given in T04. We select the median value of oxygen abundance given in the MPA-JHU catalogue for 13,748 galaxies in our parent sample.

The 9720 galaxies that have reliable metallicity (M10 and T04) and SFR (B04, S16 and Hα) estimates are hereafter referred to as 'Sample C' (See Section 2.2).

2.6 Measures of Galaxy Environment

There are many different metrics by which one may define the environment of a galaxy, most of which have been examined extensively in the literature (see Muldrew et al., 2012, for a thorough comparison). The majority of methods can be placed into the categories of friends-of-friends (FoF), *Nth* nearest neighbour and fixed aperture techniques. In general, approaches that are based upon FoF estimate properties such as mass and velocity dispersion, attributes closely related to gravitational potential. On the other hand, nearest neighbour or fixed aperture estimators provide the number density of objects, a property that is indicative of the probability of interaction. We employ one metric from each of these categories in order to best determine the extent and influence of processes at work as well as the subsequent scale dependency of environment driven gas suppression.

The number of galaxies in our parent sample with environment diagnostics and MPA-JHU SFRs is 27,667. We divide these cleanly into central (most massive galaxy in each group of two or more members), isolated (only one galaxy in a group) and satellite (less massive than central in groups of two or more members) galaxies based upon the Yang et al. (2007)[6] group catalogue, hereafter Y07. For the analysis in Chapters 5 and 6, we focus *only* on the satellites and restrict the parent sample to objects for which the full complement of environment data outlined below is available, 10,567 in total. As summarised in Section 2.2, throughout the thesis we refer to this sample as 'Sample B'.

Below we describe the three environment metrics used in this thesis.

2.6.1 Friends-of-Friends and Halo-Based Group Catalogue

The principle of FoF algorithms is that galaxies are associated with one another based upon their spacial proximity, defining a group using all objects at a proximity less than a given linking length. The advantage of FoF is that, once the group is defined, secondary derived properties such as mass and velocity dispersion may be assigned.

In this thesis, the dark matter (DM) halo masses for Sample B galaxies are provided by Y07. More specifically, we use the 'modelB' version which takes model magnitudes along with redshift measurements taken from SDSS, however, when necessary, it also uses redshifts from additional surveys (e.g. 2df; Colless et al., 2001). The authors apply the halo-based, FoF group finder algorithm developed by Yang et al. (2005) to SDSS DR7.

[6] *http://gax.shao.ac.cn/data/Group.html*

The basic procedure assigns centres to potential groups and assumes an initial mass-to-light ratio, allocating a provisional mass to each group using their characteristic luminosity. They then use this provisional mass to estimate the size and velocity dispersion of the host DM halo, using these properties to determine a density contrast for each halo and assign galaxies to their most likely group. The process is repeated until the group membership stabilises and the resulting catalogue is largely independent of the initial mass-to-light assumption.

Once the galaxy group association is confirmed, final halo masses are estimated by abundance matching the characteristic luminosity or stellar mass rank order of individual groups with the halo mass function of Warren et al. (2006). No halo masses are assigned to the smallest groups (log $M_h/M_\odot < 11.6$) in Y07. For our work we use halo masses based upon the stellar mass ranking. In Equation 7 of their work, Yang et al. (2008) provide an empirical relation for estimating halo masses below log $M_h/M_\odot = 11.6$ using an extrapolation of the mean relation between the stellar mass of the central galaxy and the parent halo mass. However, since our low halo mass bin is log $M_h/M_\odot < 12$, we don't apply this prescription to our sample and direct comparison between manually assigned DM masses is avoided.

In some cases, the process of abundance matching may yield halo mass values that deviate significantly from the "true" halo mass (see Duarte & Mamon, 2015). We estimate this bias in Figure 2.7 by applying the abundance matching method of Yang et al. (2007) to the GALFORM semi-analytic model (Gonzalez-Perez et al., 2014) described in Chapter 6. We show that estimated halo masses are, on average, 0.2 dex lower than than true masses with a standard deviation of 1 dex, 0.7 dex and 0.3 dex at true halo masses of log $M_h/M_\odot = 12$, 13.5 and 14.5 respectively. This spread is due to scatter in the predicted stellar mass-halo mass relation of GALFORM (Guo et al., 2015; Mitchell et al., 2016), meaning that stellar mass is not necessarily a clean predictor of halo mass. That said, the correlation is also significantly dependent on the implementation of feedback in the models and, as a consequence, other models produce a tighter stellar mass-halo mass relation (see Guo et al. 2015 for a comparison of different models). The scatter introduced by abundance matching is smaller than any halo mass bin used in Chapter 5 or 6.

At low redshift, massive groups and clusters exhibit an intergalactic medium hot enough to emit X-ray light via thermal bremsstrahlung radiation. This means that, where X-ray observations exist, the virialized nature of these systems can be independently confirmed and objective comparison can be made between their optical and X-ray derived properties. To this effect, Wang et al. (2011) identify 201 clusters in the Y07 catalogue

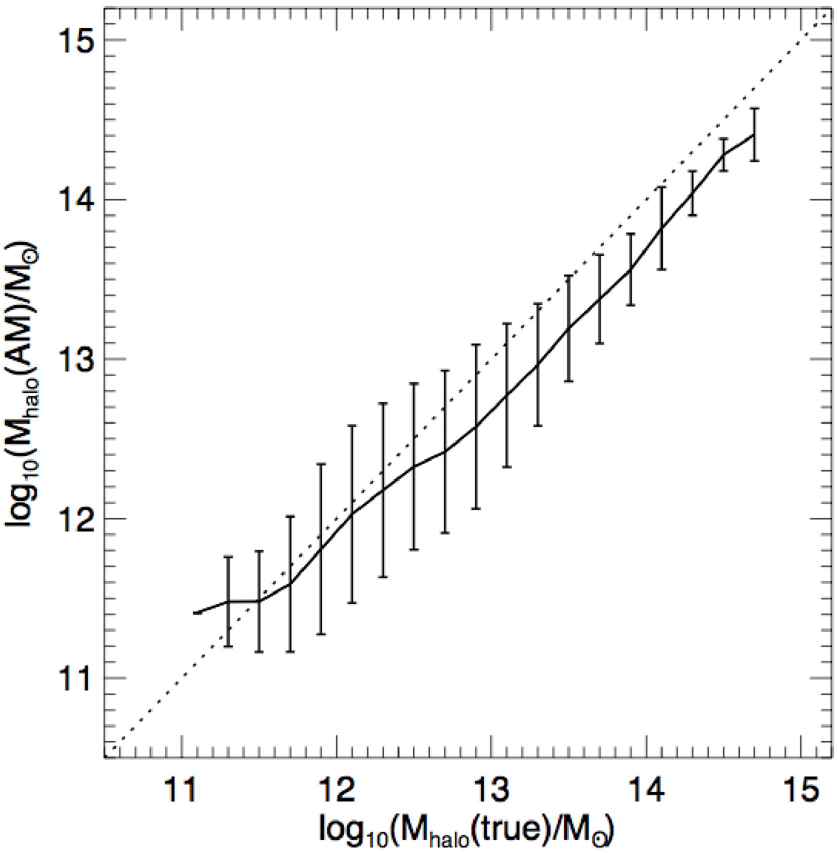

Figure 2.7: A comparison of abundance matched and "true" halo masses using the GALFORM semi-analytic model (see Chapter 6; Gonzalez-Perez et al., 2014).

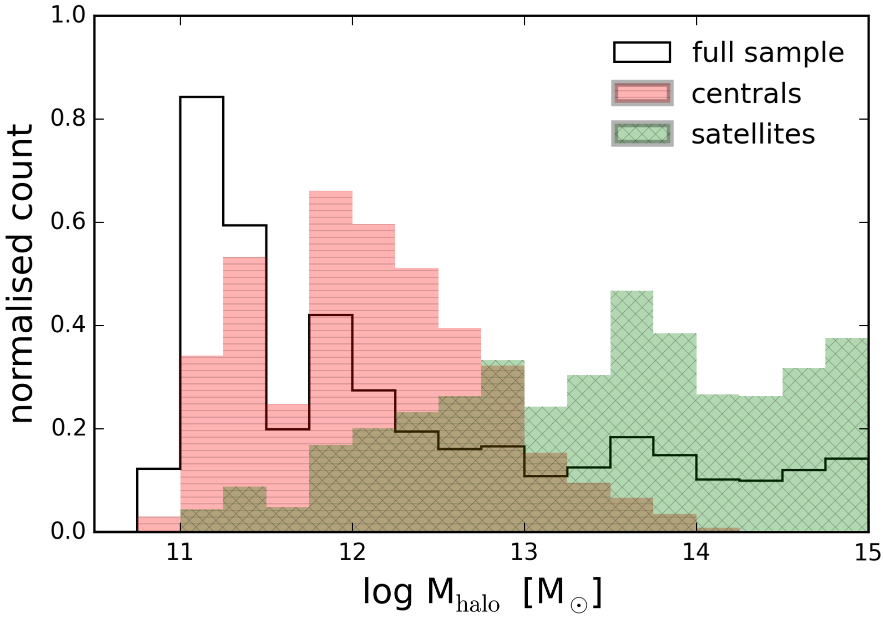

Figure 2.8: Group halo mass distribution from Y07 for galaxies within Sample B. The black solid line denotes the full sample of ∼27,700 galaxies while the red horizontal shaded histogram and green diagonally hatched histogram show the halo mass distribution of central (∼2,800) and satellite (∼10,600) galaxies respectively. Isolated galaxies (∼14,300) are not shown.

between $0.01 \leq z \leq 0.2$ with counterparts from the combined ROSAT all sky survey X-ray cluster catalogues (Ebeling et al., 1998, 2000; Böhringer et al., 2000, 2004). Encouragingly, they find reasonable agreement between halo masses derived via abundance matching in Y07 and those calculated using X-ray cluster luminosity scaling relations. Wang et al. (2011), and their subsequent paper Wang et al. (2014), show that the X-ray luminosity is correlated with the total stellar mass of the cluster and the stellar mass of the central galaxy.

The agreement between our chosen group catalogue and X-ray cluster observations, in addition to our testing of the abundance matching method, means that we can assume the halo masses of Y07 are a reliable estimate of the virial mass of groups and clusters in our sample.

Figure 2.8 shows the distribution of halo masses for the galaxies that are present in both Sample B and Y07. The sample of 27,667 galaxies is denoted by the solid black line while the 2,792 central and 10,567 satellite galaxies are displayed by the red and green shaded histograms respectively.

2.6.2 *Nth* Nearest Neighbour

Simply put, the motivation for using the Nth nearest neighbour method is that galaxies in close proximity to their neighbours, by definition, reside in dense regions of the Universe. The closer the neighbours the denser the environment.

We adopt a two dimensional nearest neighbour routine that applies a recessional velocity cut of ± 1000 km s^{-1} around the target galaxy, calculates the distance of its *Nth* neighbour and defines the density of the subsequent volume as

$$\Sigma_{\rm N} = \frac{\rm N}{\pi r_{\rm N}^2} \tag{2.9}$$

where N is the number of neighbours and $r_{\rm N}$ is the projected distance from the target galaxy to the Nth neighbour in kpc.

We apply this to DR7, using a larger volume that encompasses Sample B and all galaxies above log $M_\star/M_\odot$ = 9. This ensures completeness and removes edge effects on Sample B galaxies near the volume boundaries. Within the literature, there is no clear consensus as to the optimal number of neighbours and, as discussed in Muldrew et al. (2012), the decision depends on the scales one wishes to probe. We have investigated the differences in using 3rd, 5th, 7th and 10th nearest neighbour methods and find that, while increasing N probes larger scales, there is a strong correlation between all the neighbour-based methods. We find that the 7th nearest neighbour density (e.g. van der Wel, 2008; Muldrew et al., 2012) is a suitable match for the length scales present in Sample B, however the choice of N=7 is rather arbitrary and does not significantly affect our results.

2.6.3 Fixed Aperture

The fixed aperture technique is similar in concept to nearest neighbour, however, instead of defining a volume based upon distance to the Nth neighbour, one determines the number density of galaxies within a cylindrical volume of a given projected radius and velocity cut. Velocity cuts are intended to match the largest possible contribution of peculiar velocities to the sample, reducing interlopers that may be incorrectly placed into or out of an aperture. We compute the fixed aperture density on the larger volume used in the nearest neighbour method above. In order to match the large scales and peculiar velocities within Sample B (i.e. the Coma cluster), we set the radius of our fixed aperture at 1 Mpc (± 1000 km s^{-1}) centred on the galaxy (e.g. Grützbauch et al., 2011; Muldrew et al., 2012). We have also compared the effects of using apertures of 1 Mpc (± 500 km s^{-1}), 2 Mpc (± 1000 km s^{-1}) and 2 Mpc (± 500 km s^{-1}). Our results do not depend significantly on the aperture choice. A comparison of the 3 different environment metrics, and their effect on gas content, is given in Chapter 5.

3

H I Data and Spectral Stacking

3.1 H I Line Data

As outlined in the introduction, ALFALFA is a large blind H I survey mapping 7000 deg^2 of sky out to a redshift of $z = 0.06$ ($-2500 \lesssim v/kms^{-1} \lesssim 18,000$). Observations are carried out in drift scan mode using the Arecibo L-band Feed Array (ALFA) 7 beam receiver. Spectra from each ALFA beam are recorded separately for both orthogonal polarizations, creating two separate datasets that are then 'gridded' into three-dimensional data cubes of 2.4°× 2.4°× 5,500 km s^{-1} ($\alpha \times \delta \times$ velocity). Before any smoothing takes place, the raw spectral and angular resolution of ALFALFA data are ∼5.5 km s^{-1} (1024 channels) and 3.1×3.5 arcmin (individual ALFA beam FWHM) respectively. Cubes are constructed to have 144 pixels on each side which corresponds to an angular pixel size of 1×1 arcmin. A cartoon example of the reduced ALFALFA data cube is shown in Figure 3.1.

As part of the ALFALFA reduction, all spectra are examined by eye and flagged for radio frequency interference (RFI) and regions of low quality (due to standing waves, gain instabilities etc.). Quality weights, w, are assigned to each pixel from 0 (bad) to 20 (good) and this information is then carried forward in the construction of the three-dimensional data cubes. To correct for the variations in gain and calibration the data are bandpass subtracted and rebaselined. The final cubes are then fully processed and ready for signal extraction.

For this work, we extract H I spectra from the full volume of ALFALFA data cubes using positional coordinates, α, δ and redshift, z, drawn from the SDSS DR7 database. We extract the spectra with the raw spectral resolution (∼5.5 km s^{-1}) and velocity range (∼5,500 km s^{-1}, 25 MHz in frequency) of ALFALFA by integrating the signal over a 4×4 arcminute aperture centred on the position of the target galaxy. Galaxies in our sample have Petrosian radii containing 90 per cent of the r-band flux ($R_{90,r}$) of less than one

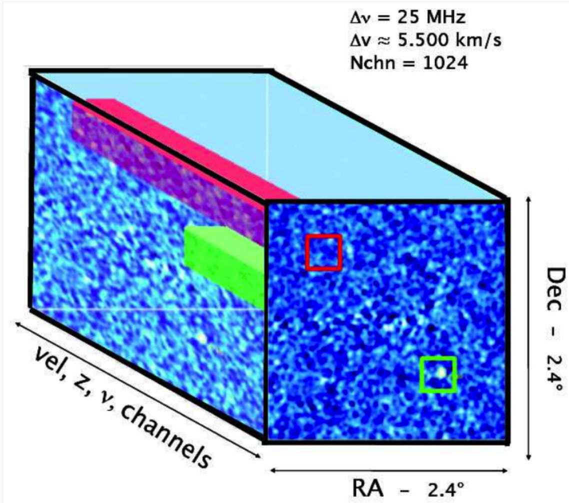

Figure 3.1: Cartoon showing a processed data cube produced by the ALFALFA pipeline. Green and red cubes show examples of the 4 × 4 arcmin cut outs centred on target galaxies from which H I spectra are extracted corresponding to detection and non-detection respectively. Figure is adapted from Fabello et al. (2011a).

arcminute, hence their H I emission is always unresolved by the Arecibo beam and we choose our aperture to match this. In order to select good quality data, we discard ∼80 spectra that do not have at least 60 per cent of their weights, w, greater than 10.

During the extraction of each spectrum, we evaluate the *root-mean-square* (rms) noise in the regions that don't contain either spurious signals (from companions or RFI) or H I line flux from the target galaxy. Since the width of an H I line profile depends on the rotational velocity of a galaxy, the region in each spectrum free from H I emission is estimated following Giovanelli et al. (1997), deriving an expected velocity width for each of our targets from the Tully-Fisher relation (Tully & Fisher, 1977) using the SDSS i-band magnitude. Gradients in the baseline are removed by fitting a low-order polynomial to the spectral channels where $w \geq 10$. The spectrum's rms is then obtained by taking the square root of the noise variance in the same high quality region.

The final quality of each spectrum is conservatively assessed by an automated routine which flags spikes in the flux density that are more than ten times the rms value and within a central 1000 km s^{-1} interval centred on the recession velocity of the galaxy. All spectra that fail this test are visually inspected and those with central RFI are discarded (∼1 per cent).

Detections within the sample are flagged using an automated routine that iterates over range of velocity widths to find the highest signal-to-noise for each spectrum, setting the detection threshold to S/N = 6.5, following Haynes et al. (2011). The ALFALFA H I

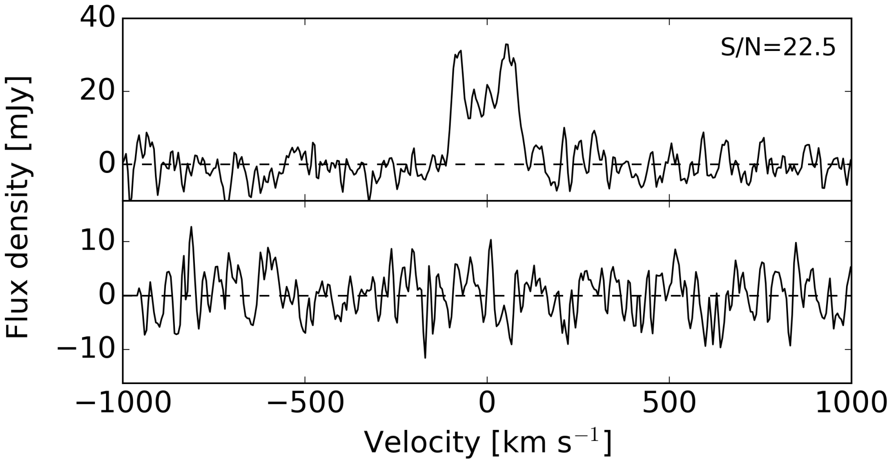

Figure 3.2: The H I spectral line profiles for two galaxies shifted to their rest-frame velocity. The top panel is a detection with the classic two-horned profile of an edge-on spiral galaxy and the bottom panel is a non-detection.

detection rate for the parent sample is 25 per cent (7552 galaxies). Figure 3.2 shows an example of detected and non-detected H I spectra from this sample.

H I masses are computed via the standard formula (Roberts, 1962):

$$\frac{\mathrm{M_{HI}}}{\mathrm{M_\odot}} = \frac{2.356 \times 10^5}{1+z}\,\mathrm{D}^2_{L(z)} \int S(v)\,dv \tag{3.1}$$

where $\mathrm{D}_{L(z)}$ is the luminosity distance at the source redshift, z, in Mpc and $\int S(v)\,dv$ is the integral of the flux density over the H I line, usually expressed in $\mathrm{Jy\,km\,s^{-1}}$. Gas fraction is then simply taken as $\mathrm{M_{HI}/M_\star}$.

3.2 Stacking of H I Spectra

Stacking makes use of the fact that contained within the three-dimensional volume of the ALFALFA data cubes is the H I emission from galaxies, regardless of whether they are formal detections or not. Each spectrum extracted from these cubes at the position of a galaxy is therefore a combination of the H I signal and a contribution of random noise that can be formalised as:

$$x_i(k) = s_i(k) + n_i(k) \tag{3.2}$$

where k is the discrete frequency (or velocity) index, $s_i(k)$ is the 21 cm emission, $n_i(k)$ is the noise contribution and i is the observation number.

Intuitively, the core principle of stacking is that by taking the linear average of N independent observations the noise term can be attenuated by the factor $\sqrt{\mathrm{N}}$. In order for stacking to behave ideally in this way, there are four conditions that have to be satisfied.

i) The signal, $s_i(k)$, and noise, $n_i(k)$, are uncorrelated.

ii) The frequency (or velocity) of $s_i(k)$ is known.

iii) Signal, $s_i(k)$, is correlated between each observation.

iv) The noise of each observation, $n_i(k)$, is uncorrelated and randomly distributed around a mean of zero with variance σ^2 that is constant across spectra (i.e. $\sigma_i^2 = \sigma^2$).

To reduce sensitivity to outliers and spikes of noise in the averaged spectrum, we adopt a weighted mean rather than the arithmetic mean when stacking:

$$\bar{x} = \frac{\sum\limits_{i=1}^{\mathrm{N}} x_i w_i}{\sum\limits_{i=1}^{\mathrm{N}} w_i} \tag{3.3}$$

where,

$$w_i = \frac{1}{\sigma_i^2} \tag{3.4}$$

The noise of the stacked spectrum $\bar{x}$ can then be calculated using the standard error propagation equation:

$$\sigma_f^2 = \sigma_x^2 \left(\frac{\partial f}{\partial x}\right)^2 + \sigma_y^2 \left(\frac{\partial f}{\partial y}\right)^2 + ... \tag{3.5}$$

Applying this to Equation 3.3 to obtain the general formula for the variance of $\bar{x}$ gives:

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^{\mathrm{N}} \left(\frac{\partial \bar{x}}{\partial x_i}\right)^2 \sigma_i^2 \tag{3.6}$$

where the derivative of $\bar{x}$ w.r.t. x_i is:

$$\begin{aligned}
\frac{\partial \bar{x}}{\partial x_i} &= \frac{\partial}{\partial x_i} \frac{\sum\limits_{j=1}^{N} x_j w_j}{\sum\limits_{h=1}^{N} w_h} \\
&= \frac{1}{\sum\limits_{h=1}^{N} 1/\sigma_h^2} \frac{\partial}{\partial x_i} \sum_{j=1}^{N} x_j/\sigma_j^2 \\
&= \frac{1}{\sum\limits_{h=1}^{N} 1/\sigma_h^2} \sum_{j=1}^{N} \frac{\partial}{\partial x_i} (x_j/\sigma_j^2) \\
&= \frac{1}{\sum\limits_{h=1}^{N} 1/\sigma_h^2} (1/\sigma_i^2)
\end{aligned} \tag{3.7}$$

Since $\frac{\partial}{\partial x_i}(x_j/\sigma_j^2) = 0$ if $i \neq j$.
Substituting the result of Equation 3.7 into Equation 3.6 yields:

$$\begin{aligned}
\sigma_{\bar{x}}^2 &= \sum_{i=1}^{N} \left(\frac{1/\sigma_i^2}{\sum\limits_{h=1}^{N} 1/\sigma_h^2} \right)^2 \sigma_i^2 \\
&= \frac{1}{(\sum\limits_{h=1}^{N} 1/\sigma_h^2)^2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \\
&= \frac{1}{\sum\limits_{h=1}^{N} (1/\sigma_h^2)}
\end{aligned} \tag{3.8}$$

If the variances of each spectrum are equal so that $\sigma_h^2 = \sigma^2$, we may simplify Equation 3.8 as:

$$\begin{aligned}
\sigma_{\bar{x}}^2 &= \frac{1}{\sum\limits_{i=1}^{N} (1/\sigma^2)} \\
&= \frac{1}{\mathrm{N}(1/\sigma^2)} \\
&= \frac{\sigma^2}{\mathrm{N}}
\end{aligned} \tag{3.9}$$

The rms noise of the stacked spectrum is defined as the square root of the variance (in the

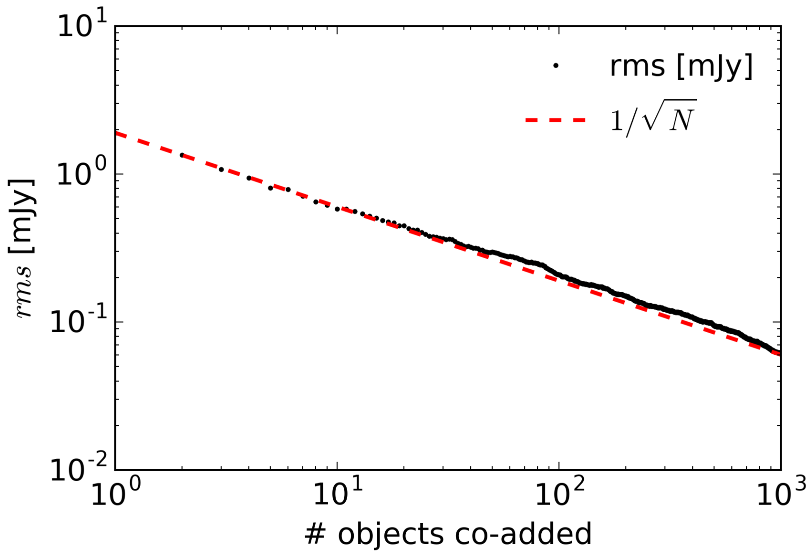

Figure 3.3: The rms of the stacked spectrum as a function of the galaxies co-added is shown by the black points. The predicted dependence of the rms upon $\sqrt{\mathrm{N}}$ is shown by the red dotted line.

regions where there is no signal) and is evaluated as:

$$rms_{\bar{x}} = \frac{\sigma_{\bar{x}}}{\sqrt{\mathrm{N}}} \tag{3.10}$$

Thus, the rms noise is inversely proportional to $\sqrt{\mathrm{N}}$.

Figure 3.3 shows this reduction in rms for the idealised case above (red dashed line) and the spectra used in this work (black points). Of course, in reality, the four conditions outlined above are somewhat violated. Most importantly, the noise level within an H I data cube is a combination of many terms, typically dominated by the Galactic and cosmic microwave backgrounds, thermal noise of the receiver and ground radiation but also including other sources such as atmospheric emission, losses in efficiency and statistical fluctuations in flux. Varying contributions between these terms means that noise is subject to change from spectrum to spectrum. Therefore, the rms is never truly random and its reduction cannot follow the $1/\sqrt{\mathrm{N}}$ prediction indefinitely, instead approaching a threshold where non-Gaussian noise begins to dominate and the reduction continues but at a decreased rate. Having said this, for samples with large numbers of spectra with stable noise characteristics, stacking remains fairly robust against such violations. This is the case for our sample where the rms reduction follows the prediction remarkably well with slight deviation as N approaches 10^3.

Before stacking, each spectrum must be shifted in frequency by an amount that aligns the recession velocity of each system with the other spectra to be stacked. We perform this shift in Fourier space to account for changes in the channel width between velocity and frequency. This is done by taking the fast Fourier transform (FFT) of each spectrum, centring the transform on the H I line rest frame frequency (1420.405752 MHz, corresponding to a recession velocity of zero) and taking the inverse FFT. The aligned rest-frame spectra are then weighted by the inverse of their variance and stacked to yield a final stacked spectrum, S_{stack}, for a sample of N galaxies, following Equation 3.3:

$$S_{stack} = \frac{\sum_{i=1}^{\mathrm{N}} S_i w_i}{\sum_{i=1}^{\mathrm{N}} w_i} \tag{3.11}$$

$$w_i = \frac{1}{\sigma_i^2} \tag{3.12}$$

where S_i and σ_i are individual flux density and rms respectively.

An example of a stacked flux spectrum containing 1000 galaxies - both detections and non-detections - is shown in Figure 3.4. The rms noise has decreased significantly compared with the typical rms of individual detections ($\sim$2 mJy, dot-dashed line) and the shape of the profile is more 'Gaussian'. This is a result of combining galaxies with different inclination in the stack i.e. face-on galaxies have narrow line profiles where H I is at low velocities with respect to the observer while edge-on systems have broader profiles as the full extent of rotation is observed.

Calculating the average H I masses from stacked flux profiles is straight forward, one simply substitutes the integrated flux density and mean distance into Equation 3.1 above. However, this approach may not always be appropriate because the average distance, and stellar mass for gas fractions, might not be representative of the distribution of galaxies in the stack. Fabello et al. (2011a) quantified the bias that this approach introduces to the stacked result for a subset of our sample, finding it to be typically around the 10 per cent level, particularly if the range of redshifts and/or masses in the sample of stacked sources is large.

In order to avoid this bias altogether, we follow the recommendation of Fabello et al. (2011a) and weight each spectrum, S_i, by its redshift, z_i, and stellar mass, $\mathrm{M}_{\star,\mathrm{i}}$, *before* stacking in order to transform the flux signal into a distance-weighted 'gas fraction spec-

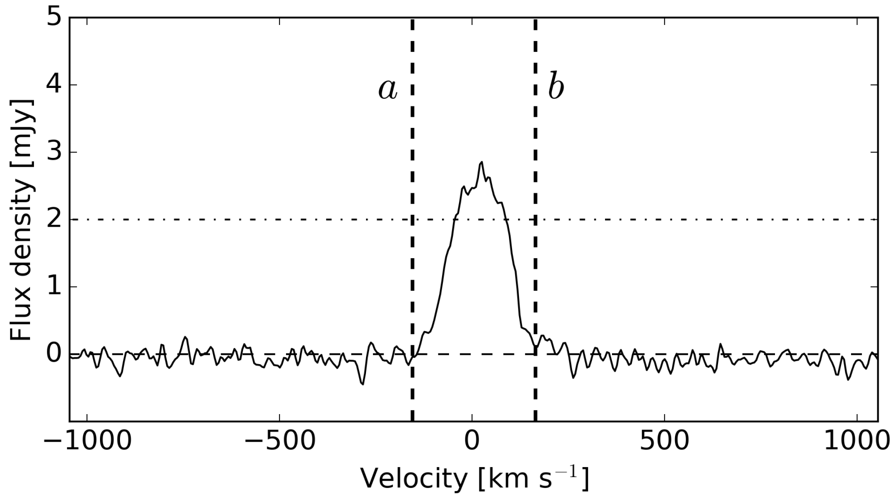

Figure 3.4: The stacked flux profile for 1000 galaxies that have been shifted to their rest-frame velocity. The limits of integration are marked a and b. The horizontal dot-dashed line shows the typical rms (∼2 mJy) of an individual ALFALFA spectrum.

trum', S'_i.

$$S'_i\,[\mathrm{Jy\,Mpc^2\,M_\odot^{-1}}] = \frac{S_i \mathrm{D}^2_{L(z),i}}{\mathrm{M_{\star,i}}} \cdot \frac{1}{1+z_i} \tag{3.13}$$

Note that for the average H I masses in Chapter 7, we don't weight spectra by stellar mass. We can then evaluate the weighted average to obtain a stacked gas fraction spectrum:

$$S'_{stack} = \frac{\sum\limits_{i=1}^{\mathrm{N}} S'_i w_i}{\sum\limits_{i=1}^{\mathrm{N}} w_i} \tag{3.14}$$

Substituting this result into Equation 3.1 yields the stacked gas fraction:

$$\left\langle \frac{\mathrm{M_{HI}}}{\mathrm{M_\star}} \right\rangle = 2.356 \times 10^5 \int\limits_a^b S'_{stack}(v)\,dv \tag{3.15}$$

where a and b are the limits of integration over the stacked emission line as illustrated in Figure 3.4.

Additionally, in some instances the stacked signal yields a non-detection (e.g. where statistics in the bins are low or galaxies are quiescent). When this occurs upper limits are computed assuming a 5σ signal with a velocity width of W = 200 km s^{-1} for bins with $\langle \mathrm{M_\star} \rangle \leq 10^{10}$ M$_\odot$, and W = 300 km s^{-1} where $\langle \mathrm{M_\star} \rangle > 10^{10}$ M$_\odot$, smoothed to a spectral

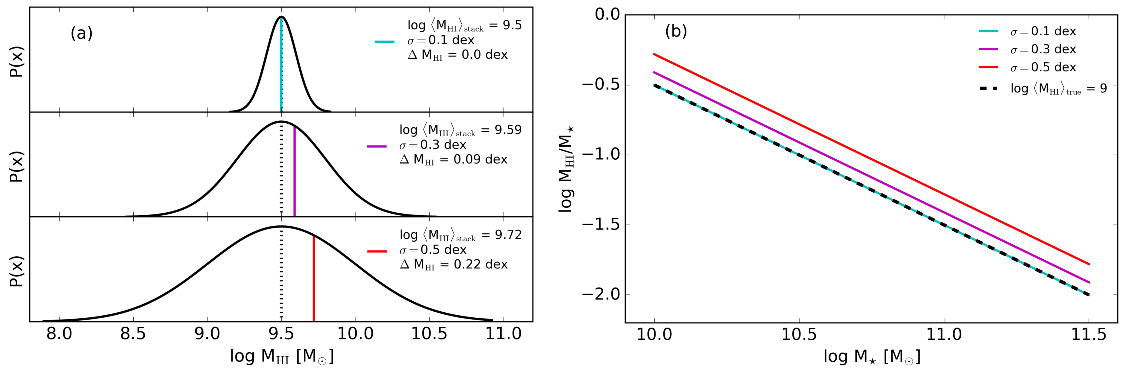

Figure 3.5: Three analytic log-normal distributions of H I masses (a), each with a "true" $\langle \mathrm{M_{HI}} \rangle$ of $10^{9.5}\,\mathrm{M_\odot}$ shown by the black dashed line. The top, middle and bottom panels have standard deviations of $\sigma = 0.1, 0.3, 0.5$ dex respectively. The average H I mass recovered from stacking each distribution is denoted by the sold coloured line. The gas fraction-stellar mass plane (b), with scaling relations for the true H I average and the stacking result from each distribution in panel (a).

resolution of (W/2) km $\mathrm{s^{-1}}$. This approach was established by Giovanelli et al. (2005) and is used in GASS (Catinella et al., 2010).

3.3 The Effect of H I Mass Distributions on the Stacking Result

One caveat of the stacking technique is that, because of the use of spectral non-detections, the distribution of individual H I masses from each stacked ensemble is not recoverable. We know from observational studies of H I detections that the distribution of H I mass in the local Universe is likely log-normal (e.g. Cortese et al., 2011). Using Figure 3.5, we show analytically how the variance of the H I distribution can affect the average mass recovered by stacking. In Panel (a), we consider three log-normal distributions for which the average H I mass is the same ($10^{9.5}\,\mathrm{M_\odot}$), we call this the "true" H I mass (dashed black line). The only difference between each distribution is the variance. The stacked H I average is denoted by the coloured lines and is clearly shown to increasingly overestimate the "true" H I mass with broadening distributions. In panel (b), we show how this trend manifests itself on the gas fraction-stellar mass scaling relation and, under the naive assumption of constant H I mass, the gas fraction is also overestimated. This issue arises because, as underlying H I mass distribution broadens in log-space, the contribution of the massive galaxies to the linear stacked average increases.

It is important to consider the impact this has on the results presented in this thesis, in particular, when we are identifying the second and third order correlations of gas content with star formation (Chapter 4) and environment (Chapter 5). For the H I masses computed via stacking to be affected, one must invoke a scenario whereby physical processes that are strongly correlated with galaxy and/or environmental properties conspire to alter the distributions of H I in a given sample significantly. Furthermore, for the bias to 'agree' with the results in this thesis, these processes must drive a strong broadening of the distribution toward higher H I masses as function of either star formation or environment. Only then would the difference between the "true" mean H I mass and the H I mass estimated from stacking mimic the observed trends between gas content, galaxy properties and environment that we see. Thus, while it is impossible to completely rule out that changes in the underlying M_{HI} distribution affect the absolute value of our stacked average, it is very unlikely to be a significant driver of the trends we see with environment and star formation. We do, however, caution that it is important to compare the relative differences between scaling relations in this thesis rather than absolute values of H I mass given by stacking. Finally, the fact that in Chapter 5, the depletion of gas content we find in the cluster regime is in agreement with previous literature using direct detections also suggests that the stacked averages we recover are robust (e.g. Cortese et al., 2011).

Stacking is an inherently linear process so care must be taken when comparing the gas fractions in this thesis with results from deep, detection dominated surveys such as GASS and the Herschel Reference Survey (HRS; Boselli et al., 2010) that the average is taken of the linear data, not of the log-scaled value, as the log of the average is not the average of the log.

3.4 Errors

Since one does not know the distribution of H I masses in the sample *a priori*, errors on the average gas fractions obtained with stacking are computed using the Delete-a-Group Jackknife method (DAGJK; Kott, 2001), a statistical estimate of the standard deviation that incorporates both the variance of the estimator and its bias. When employed using random rejection, there are no theoretical advantages to DAGJK over the traditional 'leave one out' jackknife method. The main motivation for its use is that the grouping of data allows for significant gains in computational efficiency when applied to large samples while maintaining precision in the (nearly) unbiased confidence interval of the population parameter.

Based upon the standard *Jackknife* procedure formulated by Tukey (1958), DAGJK iteratively estimates a given population parameter (i.e. mean H I gas fraction) while discarding, in turn, a separate random subset of galaxies from the sample. The weighted difference between the mean statistic measurement and each jackknifed statistic is an estimate of that group's influence on the mean value and an indicator of the variance within the dataset.

The DAGJK standard deviation estimator is:

$$\sigma(t) = \sqrt{\mathrm{Var}(t)} = \sqrt{\frac{(\mathrm{R}-1)}{\mathrm{R}} \sum_{r=0}^{\mathrm{R}} (t_{(r)} - \bar{t})^2} \tag{3.16}$$

where R is the number of replicated estimates, $t_{(r)}$ is the estimation of the population parameter *without* the rth subset and $\bar{t}$ is the mean of the replicated estimates.

When computing the DAGJK error we discard 20 per cent of the sample without replacement on each iteration, meaning that the number of estimates, R, is five.

It should be noted that the errors computed this way are not indicative of the standard deviation of the underlying distribution of individual gas fractions, this is unknown. Instead the errors show the effect of strong outliers, if present, on the final stacking result.

3.5 Source Confusion

In spectroscopy, confusion is the overlap of one or more sources in both physical space (e.g. telescope beam) and frequency space that prevents individual emission being isolated. Due to the large angular size of radio telescope beams and the extended nature of H I emission, the contribution of flux from blended sources is of particular concern when using single-dish radio observations, including when stacking. If a significant fraction (e.g. > 10 per cent; Jones et al., 2015) of the total observed flux comes from confused sources, this bias results in incorrect estimations of H I properties such as flux, mass and velocity width.

In Figure 3.6, we show an example of confusion within the thesis sample. Panel (a) is the SDSS DR7 inverted optical image[1] of three galaxies whose separation is within the Arecibo beam ($\sim$3.5 arcmin, black dashed circle) and 250 km s^{-1}. Panel (b) shows the 21 cm spectra of each galaxy, shifted to their rest frame velocity. All three galaxies exhibit signs of confusion. Galaxy A and Galaxy B have profiles that peak off-centre from their recession velocity and contain extended emission over their velocity range, while Galaxy C has an asymmetric line profile with a strong wing on the side that is approaching.

We conduct a crude estimation of the rate of confusion in the thesis sample by identifying sources within 2 arcmin (beyond which the response of the Arecibo beam is low) and $\pm$200 km s^{-1} of one another. A radius of 2 arcmin is chosen as it corresponds to a -3dB (50

[1] `http://skyserver.sdss.org/dr7/en/tools/chart/chart.asp?ra=167.4554179&dec=12.77148062`

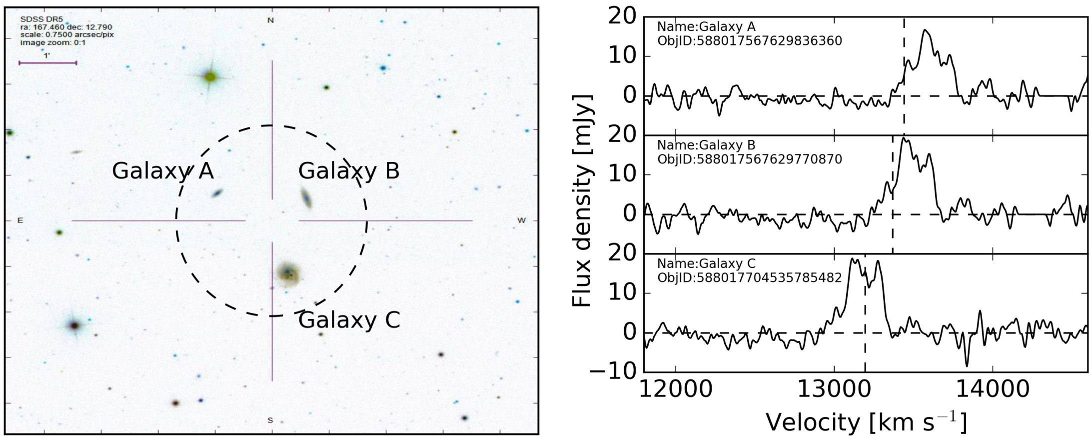

Figure 3.6: Left: SDSS DR7 optical image centred on three confused galaxies within the thesis sample. The black, dashed circle denotes the size of the Arecibo beam on the sky (~3.5 arcmin). Right: 21 cm line profiles for each galaxy with their SDSS spectroscopic recession velocities marked by the vertical dashed lines. The off centre profiles, strong wings and extended emission are the hallmarks of confused spectra. Each galaxy is labelled according an arbitrary name and SDSS DR7 ObjID.

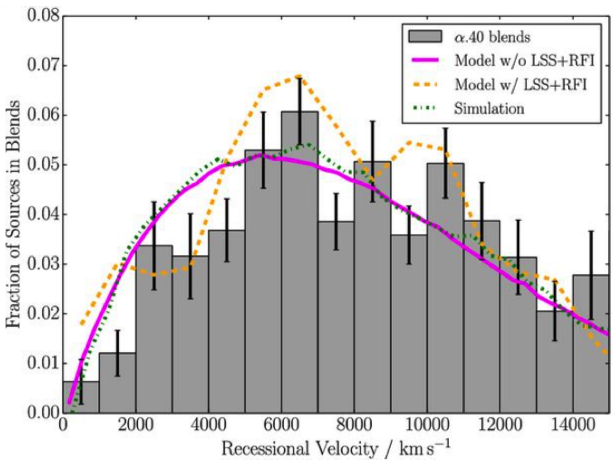

Figure 3.7: Observed rate of confused ALFALFA sources, where the H I mass of the companion source is greater than 10 per cent of the target source, in bins of recession velocity (grey histogram). Coloured lines are modelled rates of confusion within the ALFALFA data cubes (pink and orange) and a simulated ALFALFA-like catalogue (green). This figure is taken from Jones et al. (2015).

per cent) drop in the response of the Arecibo telecope beam, beyond this the contribution of flux from nearby sources to the total measurement is minimised. The velocity cut of ± 200 km s^{-1} is based upon the typical rotation velocity of a large H I disk in our sample. Excluding galaxies within this range using optical information ensures that the emission from sources contained within the beam does not overlap in frequency. Applying these cuts results in 9 per cent ($\sim$2,500 galaxies) of the sample being flagged as confused. If these criteria are increased to a far more conservative 3 arcmin (-6dB to -12dB drop in beam response) and ± 300 km s^{-1} (rotation velocity of the largest H I disks), 19 per cent of sample is flagged as confused. In this case, only the most H I-rich *and* highly inclined companions would contribute any flux to the measured spectrum and, since we do not include a colour cut on this selection to account for the reduced contribution of H I flux from quiescent objects, this is an extremely conservative upper limit on the rate of confusion within our sample. Where possible, major results in this thesis are confirmed with the former sample (9 per cent) of confused galaxies excluded.

Recently, Jones et al. (2015) modelled the rate of confusion within the ALFALFA dataset using far more sophisticated methods. Their approach is to assess the probability that two H I sources have a given physical and velocity separation using the 2D correlation function, and then assume H I properties sampled from the H I mass-width function (see Papastergis et al., 2015) to derive a prescription for ALFALFA's rate of confusion as function of distance. Over the redshift range of our parent sample, they find the rate of confusion to be no more than 10 per cent for galaxies where the H I mass of the companion exceeds 10 per cent of the target's H I mass (see Figure 3.7). The conclusion of Jones et al. (2015) is that this level of confusion is unlikely to result in serious bias in ALFALFA. Given the agreement of our rudimentary estimate with their result, we find this an acceptably low rate.

Of course, occurrences of confusion are strongly correlated with properties other than distance. For galaxies, as quantities such as mass and size increase so does the rotational velocity of the H I disk, meaning that emission from larger galaxies is more likely to overlap in frequency. Across environments, the number of blended sources within a telescope beam obviously increases in high density regions. Having said this, it is important to note that the role of confusion is to cause an overestimation with respect to the "true" H I mass, therefore, confusion will never contribute to any trend of decreasing H I mass with increasing halo mass or environmental density, both of which are recovered in this thesis.

4

Disentangling the Key Gas Fraction Scaling Relations

4.1 Introduction

As discussed briefly in the introduction to this thesis, there have been many studies into the relationships of gas content with star formation and galaxy structure for thousands of systems. Fabello et al. (2011a) applied H I stacking to a large multi-wavelength dataset, quantifying the main scaling relations of gas fraction with stellar mass, stellar surface density and colour for ∼5,000 massive galaxies. Works such as Kannappan (2004), Cortese et al. (2011), Oh et al. (2011) and Catinella et al. (2013) have looked at the same relations using targeted, deep observations. All these investigations showed a strong negative trend of gas content with stellar mass, stellar surface density and colour, identifying stellar surface density and colour, respective morphological and star formation indicators, as the two properties most tightly correlated with gas fraction.

Such studies highlight the role of internal structure and star formation in the regulation of atomic gas content. Understanding the extent and causality of this relationship is of vital importance and has been the focus of much recent work (e.g. Fabello et al., 2011a; Cortese et al., 2011; Catinella et al., 2013). However, the processes involved are not as yet fully understood and studies have frequently arrived at differing conclusions, either downplaying the importance of a bulge or bar in affecting gas content or, contrastingly, suggesting such structures may have an influence upon gas consumption (e.g. Saintonge et al., 2012; Leroy et al., 2013; Huang et al., 2014).

In this chapter we use stacking analysis with unprecedented statistics to build comprehensive scaling relations with gas content, investigating the *separate* influences of mass, structure and star formation on cold gas for the entire gas-poor to -rich regime.

4.2 Sample A and Spectral Stacking

For the analysis presented in this chapter, we use Sample A which was described in in 2. Briefly, Sample A is a volume-limited, multi-wavelength sample selected according to stellar mass ($10^9 \leq M_\star/M_\odot \leq 10^{11.5}$) and redshift ($0.02 \leq z \leq 0.05$) from the Sloan Digital Sky Survey, and with H I data from the Arecibo Legacy Fast ALFA survey. The total number of SDSS DR7 galaxies with high quality (i.e. uncontaminated by RFI) ALFALFA spectra and GALEX NUV photometry is 24,337.

With over eighty per cent H I non-detections, Sample A is ideally designed to exploit the capabilities of the stacking technique, allowing us to determine how H I content varies with galaxy properties. We use an adapted version of the software developed by Fabello et al. (2011a). A more comprehensive description of the stacking technique can be found in Chapter 3.

4.3 Gas Fraction Scaling Relations

In this section we present the main scaling relations of gas fraction versus stellar mass, stellar surface density and NUV-r colour for Sample A, based upon spectral stacking of ALFALFA data. Compared to previous work by Fabello et al. (2011a), which was based upon a subset of our sample (~5000 galaxies), we are able to extend our analysis down one order of magnitude in stellar mass ($M_\star/M_\odot \geq 10^9$) as well as investigate second order trends in gas fraction, gaining further insights into the physical drivers of these relations.

Figure 4.1 shows stacked average H I fraction as a function of galaxy stellar mass, stellar surface density and NUV-r colour, shown by the solid red lines. Errors on the average H I fractions are computed using DAGJK, as described in Section 3.4. Grey points indicate individual H I detections from ALFALFA. In Figures 4.1a and 4.1c we also plot the average *linear* gas fractions[1] (Cortese et al., 2011), provided by L. Cortese, from the Herschel Reference Survey (HRS, Boselli et al., 2010), shown by the magenta dashed line. We exclude H I-deficient galaxies - typically found within clusters - because of their significant offset to lower gas content. We do not show the HRS results in Figure 4.1b because of a difference in the definitions of stellar surface density between their work and ours. We confirm the trends of decreasing H I fraction as a function of galaxy stellar mass, stellar surface density and NUV-r colour, even with the addition of lower stellar mass galaxies. Our results are entirely consistent with the results of Fabello et al. (2011a) as well as the findings of Catinella et al. (2013), using the log of the linear gas fraction averages from the final GASS data release (dashed black line). The data points for the

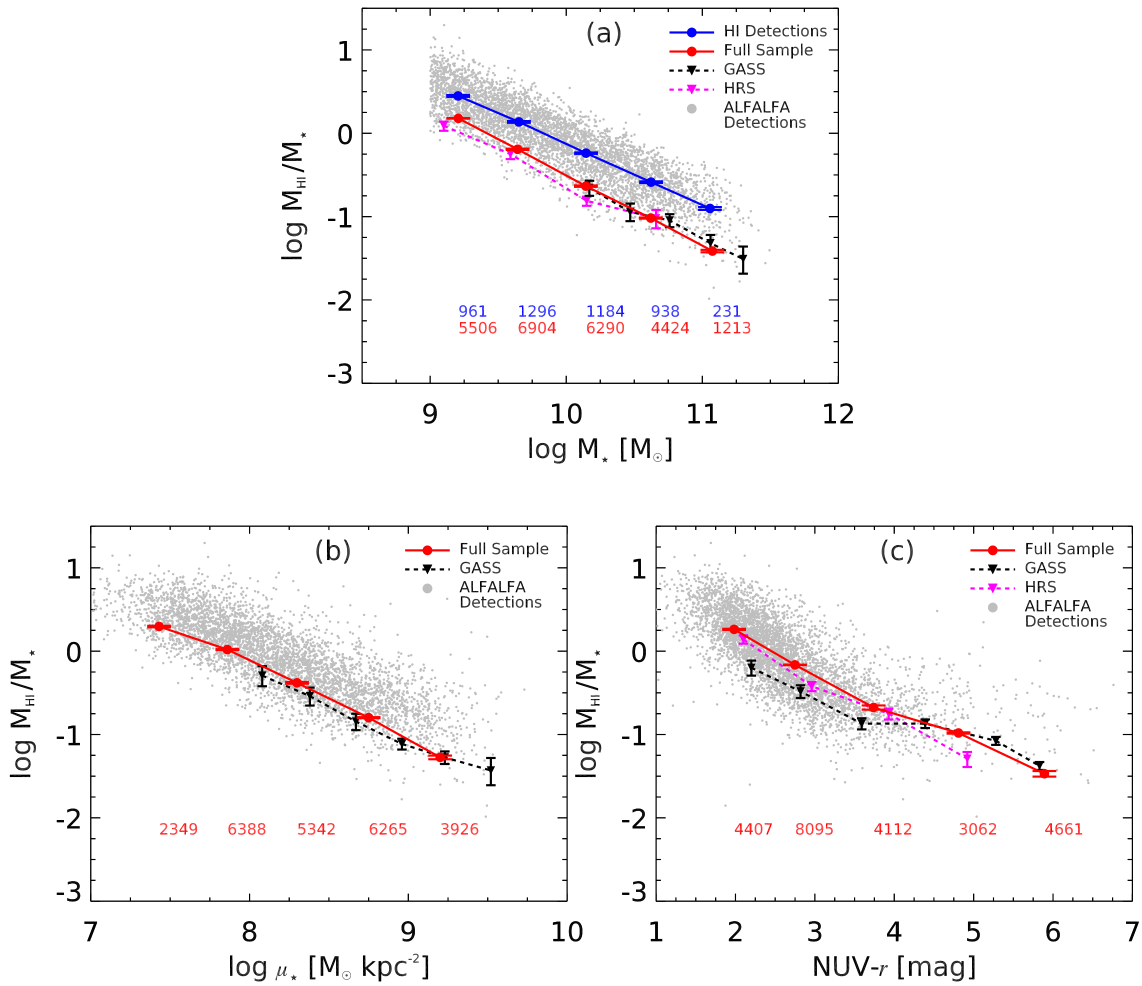

Figure 4.1: Average stacked H I gas fractions are plotted as a function of galaxy stellar mass ($M_\star$), stellar surface density ($\mu_\star$) and NUV-r colour for the whole sample. Grey points are individual detections included in the ALFALFA catalogue. The magenta dashed line is average gas fraction of nearby H I detections from the Herschel Reference Survey (HRS; Boselli et al., 2010). The dashed black line shows the scaling relation from GASS. Numbers represent the total number of co-added galaxies within each bin. The blue line in Figure 4.1a is the stacked average of the galaxies that are detected in ALFALFA. The average gas fraction data points for each relation are given in Table 4.1.

Table 4.1: Average gas fractions for the full sample scaling relations shown in Figure 4.1. The column labelled x is the galaxy property along the x-axis, $\langle x \rangle$ denotes the mean values of x within each bin, $\langle \mathrm{M_{HI}/M_\star} \rangle$ is the linear gas fraction and N gives the number of galaxies within each bin.

x	$\langle x \rangle$	$\langle \mathrm{M_{HI}/M_\star} \rangle$	N
log $\mathrm{M_\star}$	9.21	1.511 ± 0.011	5506
	9.64	0.643 ± 0.011	6904
	10.14	0.232 ± 0.005	6290
	10.62	0.096 ± 0.002	4424
	11.07	0.039 ± 0.001	1213
log $\mu_\star$	7.43	1.976 ± 0.041	2349
	7.86	1.045 ± 0.014	6388
	8.30	0.417 ± 0.010	5342
	8.75	0.159 ± 0.002	6265
	9.20	0.053 ± 0.003	3926
NUV-r	1.98	1.827 ± 0.029	4407
	2.75	0.685 ± 0.005	8095
	3.74	0.211 ± 0.012	4112
	4.81	0.104 ± 0.002	3062
	5.89	0.034 ± 0.003	4661

full sample scaling relations shown in Figure 4.1 are given in Table 4.1.

The blue line in Figure 4.1a shows the result of stacking only galaxies that are detected by ALFALFA, corroborating previous findings that H I selected samples, unless corrected, will overestimate the average gas content of galaxies within the volume (see Figure 2 in Huang et al., 2012).

To take this analysis a step further, we must establish the importance of each of the main parameters as a tracer of gas content. Stellar surface density, as a good morphological indicator, can be taken as a rough proxy for bulge-to-total ratio, another common morphological parameter, with the fraction of disk-dominated systems decreasing as $\mu_\star$ increases (Kauffmann et al., 2006). Within Sample A, more massive galaxies tend to have higher stellar surface densities and thus earlier morphologies. As the ratio of young to old stars, NUV-r colour is used as a proxy for specific star formation rate (sSFR, see

[1] As noted in Chapter 3, the distribution of gas fractions in the local Universe is more likely log-normal than Gaussian, so ideally one would compute $\langle \log \mathrm{M_{HI}/M_\star} \rangle$ (Cortese et al., 2011). However, the stacking method does not operate in log space, instead it returns the linear average of the H I content, so we must adopt log $\langle \mathrm{M_{HI}/M_\star} \rangle$ in its place. Care must be taken when comparing our results with gas fractions from deep, detection dominated surveys such as GASS and HRS that the average is taken of the linear data, not of the log-scaled value, as the log of the average is not the average of the log.

Salim et al., 2005, 2007; Schiminovich et al., 2007). We see in Cortese et al. (2011) and Catinella et al. (2013) that the H I gas fraction is most tightly correlated with $\mu_\star$ and NUV-r colour. However, no previous study has had the requisite amount of galaxies to divide the sample by two parameters simultaneously. The large statistics afforded by our sample allows us to fix a primary parameter while binning the sample in terms of a second, disentangling the individual dominance of stellar mass, stellar surface density and NUV-r colour in governing, or at least tracing, the average gas fraction of these galaxies.

Figure 4.2a shows gas fraction versus stellar mass, fixing stellar mass along the x-axis and splitting the sample into bins of stellar surface density. We then invert this in Figure 4.2b so that stellar surface density is held constant and we are separating galaxies according to their stellar mass. Blue, green and red lines denote the average gas fractions obtained by stacking. For comparison, the scaling relations for the whole sample from Figure 4.1 are shown by the dashed black line. Figure 4.2a shows a large difference ($\sim$0.8 dex) between average H I fraction for disk-dominated (low $\mu_\star$) and bulge-dominated (high $\mu_\star$) galaxies at a given stellar mass. In contrast, Figure 4.2b has a smaller variation ($\sim$0.4 dex) in gas content across the sample's mass range while holding stellar surface density constant, suggesting that the average H I fraction of a particular morphological class is, to a small degree, sensitive to the mass of the system. Using targeted H I observations, Catinella et al. (2013) found that the distributions of gas fractions averaged in the stellar mass and stellar surface density relations have a standard deviation $\sigma_{M_\star} = 0.5$ dex and $\sigma_{\mu_\star} = 0.4$ dex respectively. As the spread in gas fraction between our second parameter bins is generally comparable to or larger than these values, we can safely conclude that the differences in gas fraction between the solid lines in Figure 4.2 are significant.

Lastly, we note that the average gas fraction in the lowest stellar mass ($M_\star/M_\odot < 10^{9.75}$) and highest surface density bin in Figure 4.2b is markedly above the relation. The higher than expected gas fraction is likely due to the uncertainty involved in calculating stellar surface densities for the large fraction of less massive, with more compact objects within this bin resulting in an over estimation of surface densities as low surface brightness disks are not detected. When statistics is low these galaxies dominate the average gas fraction measurement and the point should not be considered reliable.

Similarly, in Figure 4.3 we split the sample by NUV-r colour instead of stellar surface density. The blue, green and red lines denote galaxies within the NUV-r colour bins - chosen to approximately correspond to traditional blue cloud, green valley and red sequence classifications - and stellar mass bins - chosen to span the transition mass of $\sim 3 \times 10^{10} M_\odot$ (Kauffmann et al., 2003b), where there is an observed shift in the abundance

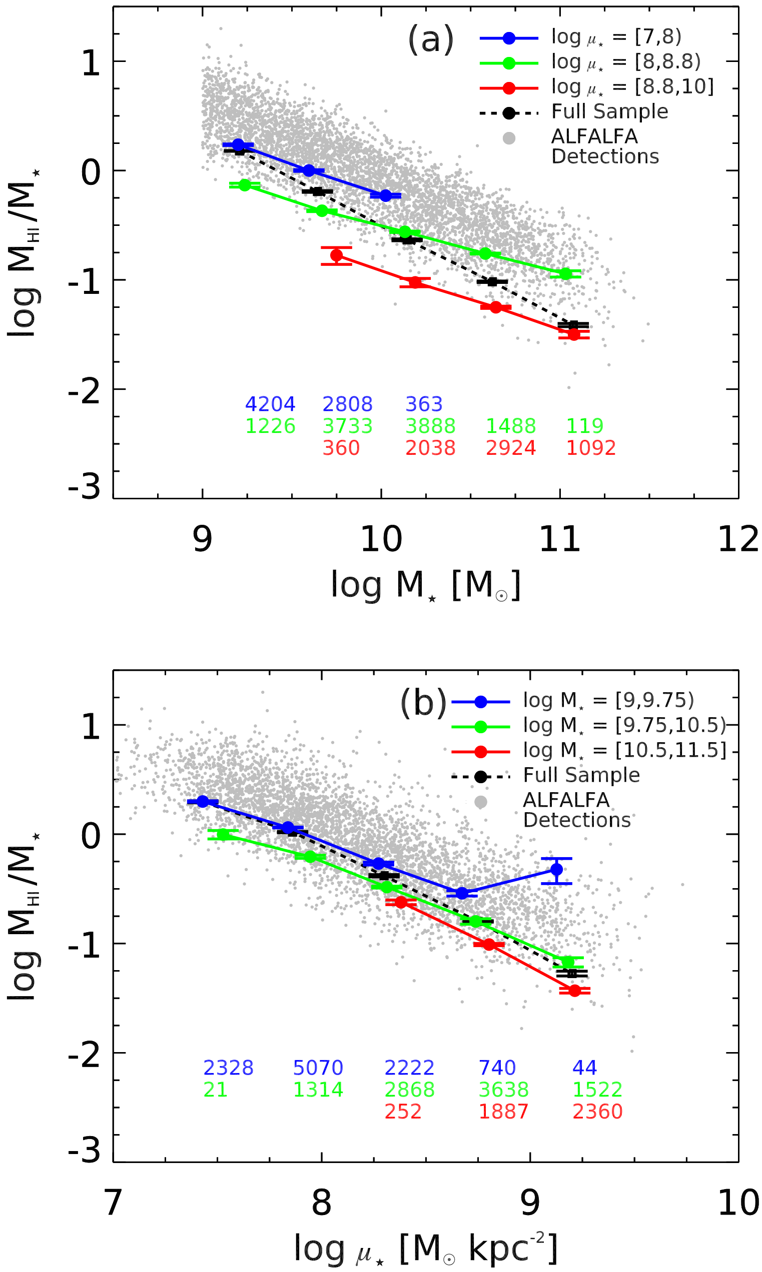

Figure 4.2: Top: H I gas fraction as a function of stellar mass, separated into low (blue), intermediate (green) and high (red) stellar surface density bins, as indicated on the top right. Bottom: H I gas fraction as a function of stellar surface density, separated into bins of stellar mass. In both panels, grey points indicate ALFALFA detections and the numbers below the relations indicate the number of galaxies co-added in each bin. The dashed black lines show the scaling relations for the whole sample from Figures 4.1a and 4.1b respectively. The gas fraction data points from each scaling relation is given in Appendix Table A.1.

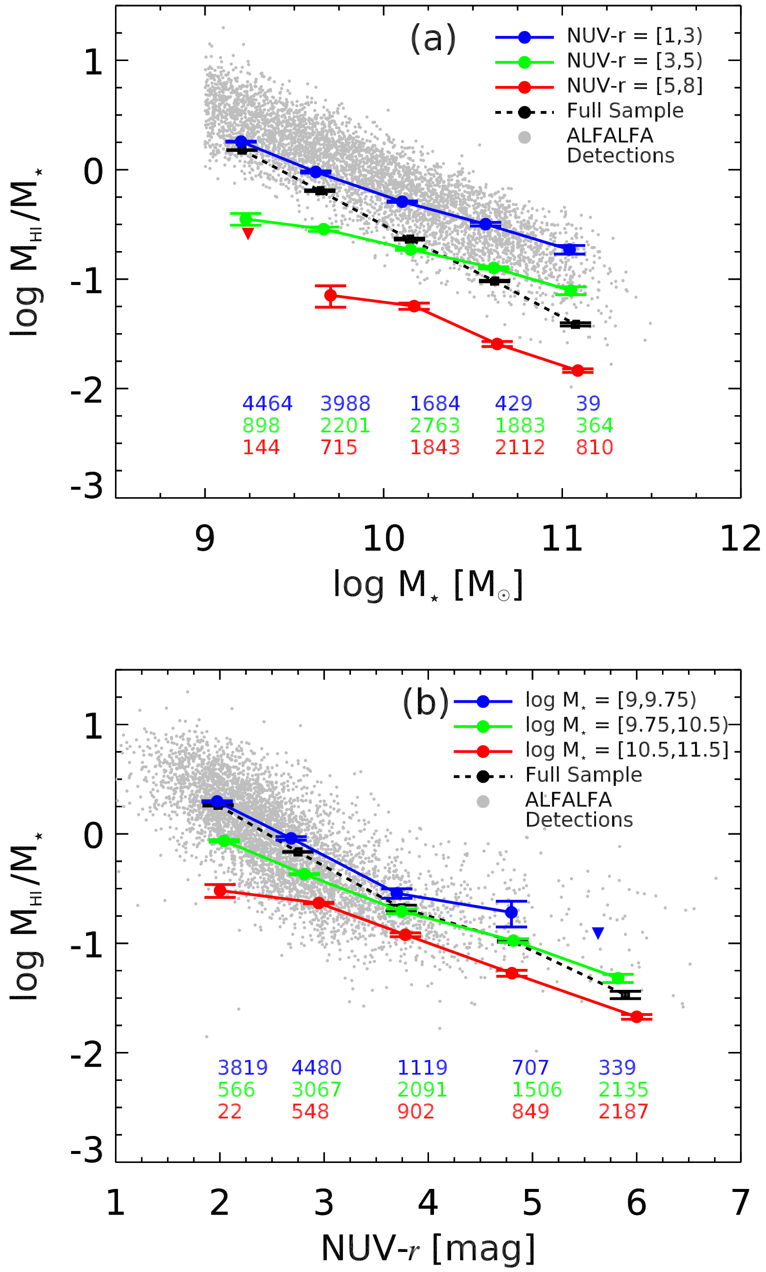

Figure 4.3: Top: H I gas fraction as a function of stellar mass, separated into bins of NUV-r colour, as indicated on the top right. Bottom: H I gas fraction as a function of NUV-r colour; separated into bins of stellar mass. Symbols and numbers as in Figure 4.2. Upper limits obtained for the bins where the stacked spectrum is a non-detection are shown as upside-down triangles. The dashed black lines reproduce gas fraction scaling relations for the whole sample from Figures 4.1a and 4.1c respectively. The average gas fraction values for each relation are given in Appendix A (Table A.2).

of late, star forming galaxies to earlier, quiescent systems. Stacked spectra that remain undetected (upside-down triangles) are found exclusively on the red sequence with NUV-r $>$ 4.5 mag. Non-detections are set to their upper limits (see Chapter 3).

Figure 4.3a shows that for a given stellar mass the gas fraction varies significantly ($\sim$1.0 dex) across the NUV-r colour range, while in Figure 4.3b the difference between average gas fraction at fixed NUV-r colour for low and high mass systems is significantly less ($\sim$0.5 dex). Even in this case, these differences are larger than the typical standard deviation of the scaling relation as directly measured from detections (e.g. $\sigma_{NUV-r} = 0.3$ dex, Catinella et al., 2013). The implication is that, once the trend of high stellar mass bins being dominated by redder galaxies is removed, the average H I fraction for galaxies of given sSFR is only weakly dependent on mass.

When splitting the gas fraction-stellar mass scaling relation in terms of either surface density or NUV-r we find that the slope of the linear fit to the relation flattens considerably ($\nabla_{\mu_\star} = -0.45 \pm 0.01$; $\nabla_{\text{NUV-}r} = -0.35 \pm 0.02$) with respect to the whole sample ($\nabla_{\text{full}} = -0.85 \pm 0.01$) shown by the dashed black line in Figures 4.2 and 4.3. This clearly shows that the steep slope of the gas fraction-stellar mass relation arises as a result of preferentially stacking blue, gas-rich galaxy populations in low stellar mass bins and red, gas-poor systems in high mass bins.

Having confirmed that stellar surface density and NUV-r colour dominate over stellar mass as tracers of atomic gas content, the next step is to test which one of these is the principal parameter in the determination of average H I fraction. The best method to address this is presented in Figure 4.4, the main result of this chapter. We remove the stellar mass constraints and show how the gas content varies when we fix surface density or NUV-r colour while binning in terms of the other. As with Figures 4.2 and 4.3, the non-detections occur exclusively in bins of NUV-$r >$ 4.5, where the stacking of large numbers of galaxies is required to reduce the rms noise sufficiently for the signal to be detected. Figure 4.4a demonstrates that galaxies at a given stellar surface density exhibit a spread ($\sim$1.0 dex) in H I content across the NUV-r colour range (solid lines) from blue cloud to red sequence. The difference of 1.0 dex is statistically significant when compared to the scatter in the gas fraction-stellar surface density relation found by Catinella et al. (2013), $\sigma_{\mu_\star} = 0.4$ dex. In contrast, we fix the colour in Figure 4.4b while separating galaxies according to low, intermediate or high stellar surface density (solid lines). In this case the difference in gas fraction between the highest and lowest $\mu_\star$ bins, i.e. the bulge- and disk-dominated systems, decreases to $\sim$0.5 dex on average.

This implies that the most important quantity in tracing of neutral atomic hydrogen content is NUV-r colour. Making additional cuts in surface density, already a mass dependent quantity, does not significantly alter the values of gas fraction. In other words, galaxies of a similar colour are, on average, likely to contain similar ratios of cold gas to stellar mass, showing only a small dependence on size (e.g. mass) or morphology (e.g. surface density).

While the strong relation between gas and NUV-r colour is evident, Figures 4.3b and 4.4b do demonstrate a smaller, residual dependence of gas fraction on stellar mass and density at fixed NUV-r colour. This result raises interesting questions surrounding the impact of stellar mass and surface density on gas consumption in addition to star formation, which will be discussed in Section 4.4.

In addition, the slope of the gas fraction-surface density relation is flattened when binning galaxies by NUV-r colour. The linear fit to the full sample relation has a gradient of $\nabla_{\rm full} = -0.89 \pm 0.04$, while selection by NUV-r reduces the slope to $\nabla_{\text{NUV-}r} = -0.33 \pm 0.06$. This reaffirms that selection by only surface density yields a mixed population of galaxies and shows that the surface density scaling relation is driven by the underlying correlation of gas content with NUV-r colour.

To confirm the low affinity of gas fraction with stellar mass, we also split the scaling relations of Figure 4.4 into lower ($\mathrm{M}_\star/\mathrm{M}_\odot < 10^{10}$) and higher stellar mass ($\mathrm{M}_\star/\mathrm{M}_\odot \geq 10^{10}$) bins (not shown). The difference between the relations of Figure 4.4 for low and high mass galaxies was not significant, $\sim$0.1 dex (a) and $\sim$0.2 dex (b). This reaffirms that, within our sample, the variation of gas content across the stellar mass range is small, once morphology and/or sSFR of galaxies has been fixed.

Finally, we stress that by looking at the effect of secondary parameters on the H I scaling relations we have been able to put important constraints on the underlying gas fraction distribution of the stacked population. In other words, we have been able to overcome one of the main limitations of H I stacking, thus enhancing the scientific potential of this technique.

4.4 Discussion and Conclusions

In this work we have applied H I spectral stacking to a large sample of 24,337 galaxies. Each galaxy is selected according to redshift and stellar mass from the SDSS DR7, with H I spectra and NUV data from the ALFALFA blind H I-survey and GALEX catalogues respectively. The goal of this study is to investigate the dependence of gas content on the galaxy properties of stellar mass, stellar surface density and NUV-r colour, extending the previous work of Fabello et al. (2011a) down to lower stellar masses ($\mathrm{M}_\star/\mathrm{M}_\odot \geq 10^9$) and

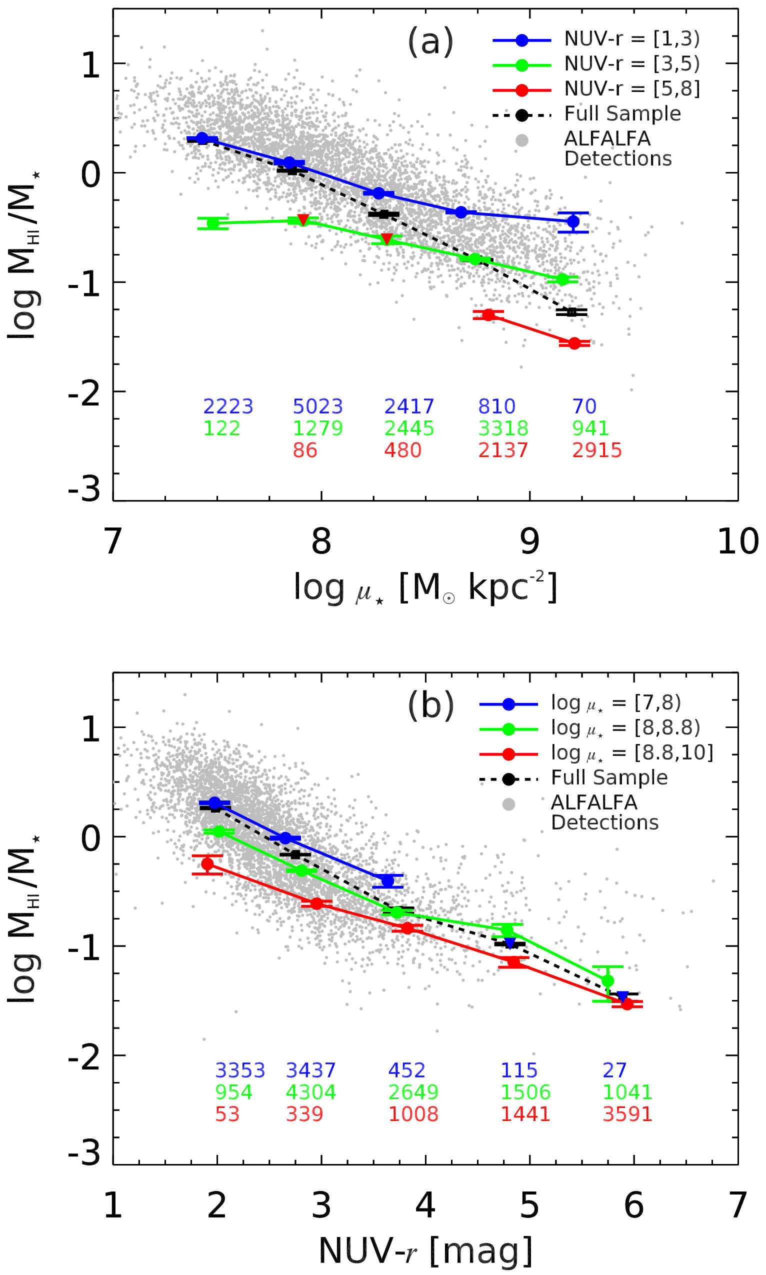

Figure 4.4: Top: Gas fraction as a function of stellar surface density, separated into blue, green and red sequence galaxies according to their NUV-r colour (solid lines) as indicated on the top right. Bottom: Gas fraction as a function of NUV-r colour, separating into bins of stellar surface density. Symbols and numbers as in Figure 4.2. The full sample scaling relations from Figures 4.1b and 4.1c are plotted as black dashed lines. Triangles denote non-detections, set to their upper limits. For the data points, refer to Table A.3.

significantly increasing the number of galaxies.

The key conclusions of this chapter can be can be summarised as follows:

i) We confirm that NUV-r colour excels over stellar mass and stellar surface density as a tracer of galaxy gas content, as previously noted by Cortese et al. (2011), Fabello et al. (2011a) and Catinella et al. (2013). Additionally, we quantify the strong decrease of gas fraction with increasing NUV-r colour at fixed stellar mass and stellar surface density.

ii) We show for the first time that the gas fraction-stellar mass and, to first order, the gas fraction-surface density scaling relations are driven by the primary correlation of gas content with NUV-r colour.

iii) At fixed NUV-r colour we find a small residual dependence on stellar mass and stellar surface density. This suggests a residual effect of mass and morphology on gas consumption at fixed sSFR, as discussed below.

As already mentioned, one may regard NUV-r colour as a proxy for sSFR and thus, under the simple assumption that gas fraction and colour are derived over the same surface area, it is easy to show that the gas fraction-NUV-r relation can be interpreted as an integrated H I Kennicutt-Schmidt (KS) law (see Figure 1.4; Schmidt, 1959; Kennicutt, 1998b) relating the atomic gas content of galaxies to their star formation activity. In this context, our findings not only confirm that star formation is the property of galaxies most closely related to their H I gas content, but also highlight that the main scaling relations of gas fraction-stellar mass, stellar surface density and NUV-r colour can be understood as a combination of the underlying bimodality in specific star formation and the KS relation. Low mass or disk galaxies are preferentially blue, star formers, whereas massive or bulge-dominated systems are, on average, more quiescent. This means that the gas fraction-stellar mass and -stellar surface density relations are simply driven by the variation in sSFR between the two populations.

Given that NUV-r colour is the principle driver behind the main integrated gas fraction scaling relations, it is interesting that we find a residual dependence on stellar mass and surface density in the gas fraction-NUV-r plane (see Figures 4.3b and 4.4b). High mass and surface density galaxies have lower H I fraction than low surface density objects at fixed NUV-r colour. If, as above, we take NUV-r as equivalent to sSFR, this shows that massive and bulge-dominated galaxies have a lower gas fraction than low mass and disk-dominated systems respectively for the same sSFR. It follows from this that the timescales over which

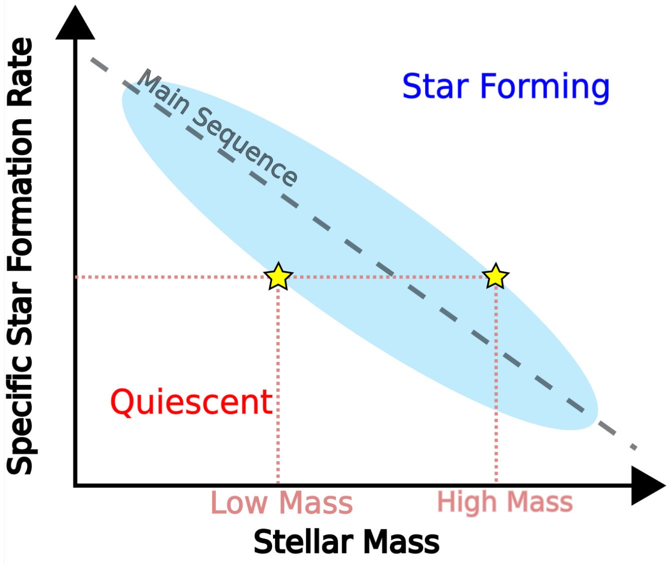

Figure 4.5: Diagram outlining the dependence of sSFR on stellar mass. The scatter in the sSFR-$M_\star$ relation is represented by the blue ellipse and the grey-blue dashed line denotes the star forming 'main sequence'. Galaxies classified as quiescent reside below the main sequence while star forming galaxies are found above. The pink dotted line shows a constant value of sSFR; the galaxies lying 'below' the main sequence, and thus deemed quiescent, are low mass systems while the high mass galaxies deviate 'above' the main sequence in their role as star formers.

gas is consumed by star formation (hereafter depletion time, M_{HI}/SFR) are shorter and therefore star formation efficiency (SFE $\equiv$ SFR/M_{HI}) is higher in more massive or bulge-dominated galaxies than their low mass or disk-dominated counterparts, assuming that star formation continues at its current rate. Our evidence contributes to a physical picture where the gas content of galaxies is strongly regulated by star formation and, albeit to a lesser extent, influenced by both mass and structure. This is tentative evidence that the KS relation is to some degree dependent upon the morphological properties of the galaxy.

Assuming that our results might extend to the molecular gas (H_2) component, the secondary dependence of gas content on stellar mass and surface density is in qualitative agreement with Huang & Kauffmann (2014) who examined the variation in H_2 depletion times in bulges, spirals, bars and rings, at fixed sSFR, for a sample of massive ($M_\star/M_\odot > 10^{10}$) local galaxies. Their results show that, at fixed sSFR, the H_2 depletion times are shorter for bulge-dominated galaxies. In their discussion they invoke the conclusions of Helfer & Blitz (1993) as a possible explanation, whereby gravitational potential and the density of molecular clouds is increased in the presence of a stellar bulge.

Conversely, Saintonge et al. (2012) find that for their sample of galaxies, a subset of which is analysed in Huang & Kauffmann (2014), the shortest depletion times and thus the highest SFEs are found in disk-dominated galaxies. This conclusion is reached by examining SFE as a function of distance from the star formation (SF) main sequence, showing that weakly star forming galaxies have low gas content and long depletion times, and combining this with the derived H_2 KS law where high surface density systems lie systematically below the mean relation.

At face value, our result showing variation in gas content with surface density seems contradictory to the conclusion of Saintonge et al. (2012). However, it is easy to show that the inconsistency is only apparent. Both Huang & Kauffmann (2014) and this work examine depletion time at fixed sSFR while, on the other hand, Saintonge et al. (2012) leave sSFR unconstrained and investigate depletion time as a function of distance from the SF main sequence. Once one accounts for the difference in method our findings are entirely consistent with those of Saintonge et al. (2012). To illustrate this, the cartoon in Figure 4.5 shows that for fixed sSFR, high mass galaxies are clearly offset above the specific SF main sequence (see Salim et al., 2007) while low mass galaxies are found in the quiescent region below the mean relation. The increasing prominence of bulges in high mass galaxies leads to bulge-dominated galaxies being deemed *stronger star formers* in comparison with disks at fixed sSFR. Accepting this, Figure 4.5 illustrates how bulge-dominated galaxies lie above the specific SF main sequence and thus must have lower depletion time in comparison to disks at fixed sSFR, a result that is in agreement with the works of both Huang & Kauffmann (2014) and Saintonge et al. (2012).

Of course it is important to offer more conclusive arguments on the relationship between gas and star formation in galaxies and to do so we must use the physical star formation rates and efficiencies, rather than employing the proxy of NUV-r colour. For this reason we switch to the star formation indicators described in Chapter 2 for the rest of the work in this thesis.

This chapter introduced a more complete description of the relationships between galaxy properties and gas content, an area that provides strong constraints for galaxy formation and evolution models by disentangling the influence of these properties on gas content. We demonstrated the importance and power of H I spectral stacking, a technique with great potential for investigating the physical mechanisms that drive the evolution of galaxies by probing the cold gas content of huge galaxy samples in regimes that would otherwise be inaccessible until the full capability of the Square Kilometre Array is realised.

5

Gas Stripping in Satellite Galaxies: from Pairs to Clusters

5.1 Introduction

Since observations first demonstrated that morphological fraction changes dramatically with the density of galaxies (Oemler, 1974; Dressler, 1980; Postman & Geller, 1984), galaxy properties have been known to reflect their environment. More recent studies have built upon this early work to show the strong environmental dependencies of morphology (Whitmore & Gilmore, 1991; Poggianti et al., 2008; Wilman & Erwin, 2012) and star formation (Balogh et al., 1999; Gómez et al., 2003; Kauffmann et al., 2004; Hogg et al., 2004; Blanton et al., 2005; Baldry et al., 2006; Cooper et al., 2008).

Studies of cold gas ($\sim 10^2$ K) show that reservoirs are adversely affected in the highest density environments such as galaxy clusters (Giovanelli & Haynes, 1985a,b; Chung et al., 2009; Cortese et al., 2011; Serra et al., 2012), and that gas processing begins to occur within the group environment (Kilborn et al., 2009; Rasmussen et al., 2012; Catinella et al., 2013; Hess & Wilcots, 2013). Throughout the literature, depletion of H I content due to environment is attributed to several different processes: the interaction between the interstellar-medium (ISM) and intergalactic-medium (IGM) known as ram-pressure stripping (Gunn & Gott, 1972; Hester, 2006); heating and stripping of hot gas in the DM halo preventing replenishment (strangulation; Larson et al., 1980); high (harassment; Moore et al., 1998) and low (tidal stripping; Moore et al., 1999) velocity gravitational interactions with neighbours.

Despite this progress, disentangling the primary, secondary and, in some cases, tertiary connections between internal galaxy properties, environment and gas content is a topic of much current interest. In pursuit of this, recent works have begun to separate and classify galaxies based upon their status as a central (most-massive and/or luminous) or satellite galaxy within the halo. The physical motivation is that satellites infalling into the halo have undergone a distinct evolutionary history from that of their central, as well as being the bulk of the group and cluster populations (e.g. van den Bosch et al., 2008; Kimm et al., 2009; Peng et al., 2012; Woo et al., 2013). Following this, van den Bosch et al. (2008) and Wetzel et al. (2012) argue that the environmental relationships and build-up of the red sequence are primarily driven by the quenching of satellites rather than their central and it is likely that the transformations are caused by removal or consumption of the cold gas content.

Despite such studies, there remains a paucity of works investigating the gas content of satellite galaxies from a statistical perspective. The extent to which H I loss may be attributed to environment, what processes are at work and in which regimes are questions that have not yet been answered. In this work we use the largest representative sample of H I to date, coupled with the spectral stacking technique, to address these questions. We look to provide the very first large-scale, statistical census of cold gas and environment for satellite galaxies in the local Universe.

Section 5.2 contains an overview of the sample selection, environmental measures and stacking technique used in this chapter. Section 5.3 studies the main H I-to-stellar mass ratio scaling relations as a function of halo mass. We look at the effect of mergers on our results in Section 5.4. In Section 5.5 we characterise environment using nearest neighbour and fixed aperture densities, investigating their impact on gas fraction. Lastly, we discuss our conclusion in Section 5.6 and consider the physical mechanisms at play.

5.2 Sample and Stacking

The sample used in this work contains 10,567 satellite galaxies (35 per cent of the parent sample) for which there are observed atomic hydrogen and optical data along with subsequently derived stellar masses and star formation rates. We refer to this sample as Sample B. Each galaxy has an assigned halo mass, as well as calculated nearest neighbour and fixed aperture densities. It is a sample built to be representative of the local Universe and therefore an ideal resource for studying environment driven evolution in the gas content of satellites from groups to clusters.

The focus of this work is to probe the relationship between H I and environment in satellite galaxies. As discussed in the introduction, sensitivity limitations of current H I surveys make it infeasible to obtain detections for very large, representative samples such as ours. This effect is particularly pronounced in studies of environment because of the H I deficiency of galaxies found in the large group and cluster regimes (Giovanelli & Haynes, 1985b; Cortese et al., 2011). Using H I stacking, we are able to quantify the gas content for the entire gas-poor to -rich regime and obtain the average atomic hydrogen content for each selection of co-added galaxies, regardless of whether the objects are formal detections in emission. Our sample consists of 1627 satellite galaxies detected in H I (15 per cent) and 8940 non-detections (85 per cent).

As discussed in Chapter 3, errors presented in this work are calculated using the statistical delete-a-group jackknife routine which iteratively discards a random 20 per cent of the stack selection, recomputing the average gas fraction. The jackknifed uncertainty is essentially the standard error on the mean gas fraction calculated by stacking and depends most strongly on the number of galaxies in each stack.

The rate of spectroscopic confusion with a sample is an important concern with all single-dish radio observations and in particular when stacking. Chapter 3 also contains a thorough discussion of the level of confusion present within our sample. Briefly, across the redshift range of our parent sample, the rate of confusion within the ALFALFA dataset is no more than ten per cent (Jones et al., 2015). This is an acceptably low rate and unlikely to heavily bias the stacked average results. Further to this, although the number of sources blended within the Arecibo beam (3.5 arcmin) clearly increases in crowded environments, the impact of confusion is to increase stacked H I mass and, thus, will not contribute to any observed trends of decreasing gas content with environment.

5.3 The Influence of Halo Mass upon Gas Fraction

In this section we present the main gas fraction scaling relations of H I-to-stellar mass ratio versus stellar mass, sSFR and stellar surface density. This work disentangles, for the first time, the effects of mass, morphology and star formation on the gas content of satellite galaxies as a function of DM halo mass.

In Figure 5.1, we plot the stacked average H I fraction as a function of stellar mass (a), sSFR (b) and stellar surface density (c) for satellite galaxies. The scaling relations for all satellites, not binned by environment, are denoted by the black dashed lines. Coloured lines show the stacked gas fraction relations in each of the halo mass bins given in the legend. At the bottom of each plot we provide the corresponding numbers of galaxies in each bin. Our chosen halo mass intervals divide the galaxies among the environments that are outlined in Table 5.1. In the bin containing galaxy pairs, the question of whether such systems can be considered bound is a valid one. While interlopers and chance superpositions do occur, comparisons of the Yang et al. (hereafter Y07; 2007) catalogue with detailed mock galaxy

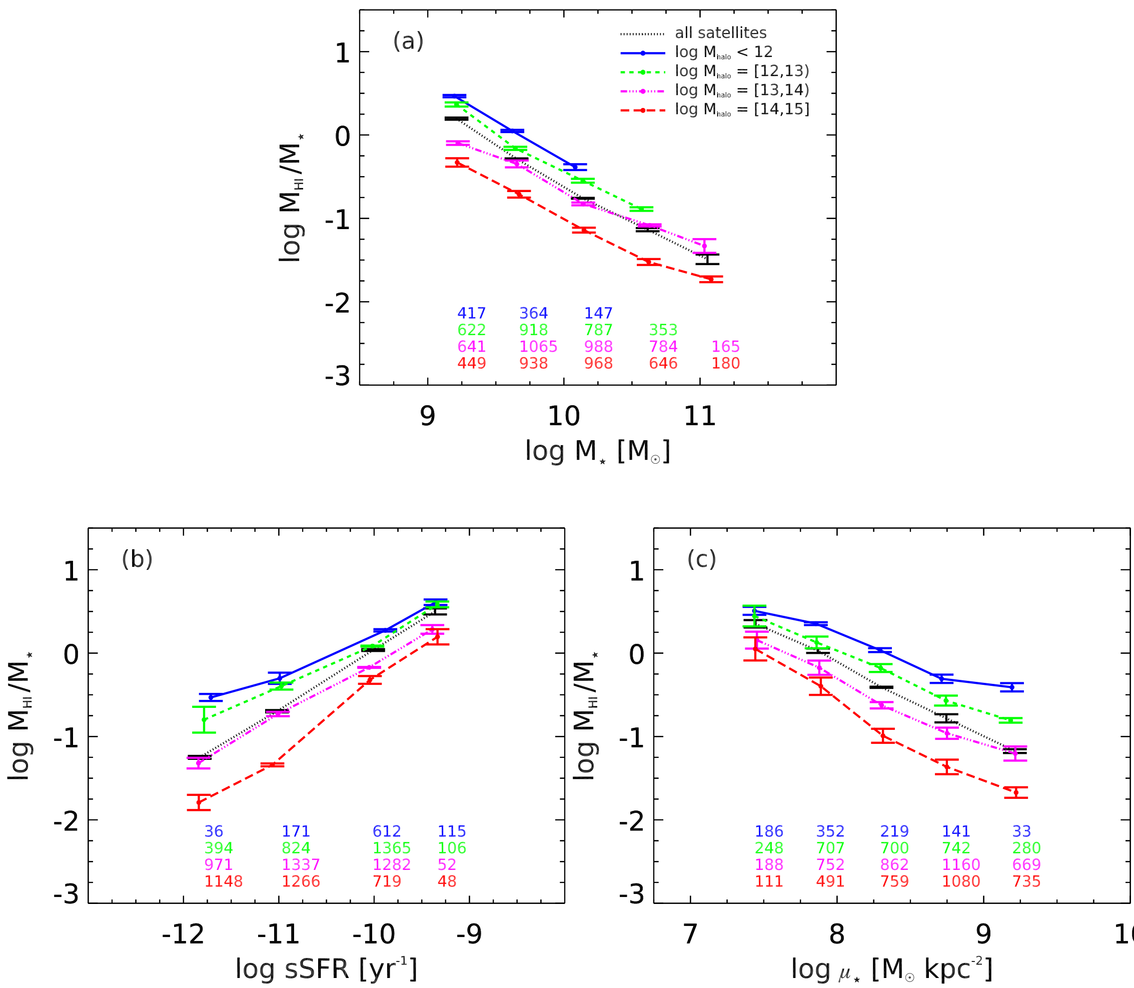

Figure 5.1: Average H I gas fractions as a function of stellar mass (a), specific star formation rate (b) and stellar surface density (c) for ∼10,600 satellite galaxies. Black dotted lines show each scaling relation for the full sample of satellites. Solid blue, dashed green, dot-dashed magenta and long dashed red lines denote the relations when binned according to the halo mass of each satellite's host central. Halo mass limits are given in the legend and numbers along the bottom correspond to the sample statistics in each bin.

Table 5.1: Environments and their equivalent halo mass interval used throughout Section 5.3. The upper and lower bin bounds are given in the first column. $\overline{\mathrm{N}}_{\mathrm{gal}}$ is the mean number of group members found in each environment, while $\widetilde{\mathrm{N}}_{\mathrm{gal}}$ is the median value.

$\mathrm{x} = \log \mathrm{M_h/M_\odot}$	Environment	$\overline{\mathrm{N}}_{\mathrm{gal}}$	$\widetilde{\mathrm{N}}_{\mathrm{gal}}$
$\mathrm{x} < 12$	Pairs/small groups	2	2
$12 \leq \mathrm{x} < 13$	Medium groups	5	4
$13 \leq \mathrm{x} < 14$	Large groups	26	21
$\mathrm{x} \geq 14$	Clusters	242	169

redshift surveys shows that the group finding algorithm performs remarkably well in this regime (>95 per cent completeness; see Y07). Note that this work does not investigate the effect of environment in the small group regime, instead using it simply as the 'zero-point' against which we compare the larger groups.

Figure 5.1a shows gas fraction versus stellar mass, separating the sample into the four environment bins. Note that there are no high stellar mass satellites in small groups, this is due to the abundance matching technique used to assign halo masses (see Section 2.6.1). When the sample is split by halo mass we find that *at fixed stellar mass* there is a smooth and systematic reduction of satellite H I content as halo mass increases. A satellite of $\mathrm{M_\star/M_\odot} = 10^{10}$ which resides in a pair or small group is, on average, 0.2 - 0.5 dex more gas-rich than its stellar mass equivalent in a medium or large group, and has a gas fraction 0.8 dex larger than its cluster counterpart in a halo of $\mathrm{M_h/M_\odot} \geq 10^{14}$.

In Figure 5.1b sSFR is held constant and the sample is again separated by halo mass. In the two bins where $\mathrm{M_h/M_\odot} < 10^{13}$, small to medium sized groups, the difference between average gas fraction across the range of sSFR is not significant (0.1 dex). *At fixed sSFR*, the galaxies in these halos are statistically comparable in their average gas content. However, we do see large decreases (0.5 dex) in the average H I fraction as a function of environment for galaxies with equivalent sSFR in halos of $\mathrm{M_h/M_\odot} \geq 10^{13}$. Not surprisingly, it is in the cluster regime where halo masses exceed $\mathrm{M_h/M_\odot} = 10^{14}$ that we see the greatest impact (0.8 dex) on the H I of satellites.

We examine how H I content varies as a function of halo mass at fixed stellar surface density in Figure 5.1c. Galaxies of a given surface density exhibit a large spread (0.9 dex) in gas fraction, with a smooth progression to lower H I fraction across the range of environments. Note that the dispersion in gas content with halo mass at fixed surface density increases from small (0.5 dex) at disk-dominated, low densities to large (1.3 dex) at bulge-dominated, high densities. This is the result of the increasing contributions from the bulge in the measurement of $\mathrm{R}_{50,z}$, meaning that the high surface density regime includes a fraction of galaxies that, while bulge dominated, still have a disk component.

Chapter 4 established that, despite a residual dependence, stellar mass is not in fact an ideal tracer of neutral atomic hydrogen content and that sSFR is more closely related to the H I-to-stellar mass ratio. Along with H I, these two properties in particular correlate strongly with halo mass (see Wetzel et al., 2012, 2013), thus when taking this analysis further one must check if decreases in H I content as a function of halo mass are due to the sensitivity of gas to the external environment, or a consequence of stellar mass or sSFR properties. The large number of galaxies available allow us to disentangle these dual effects by controlling halo mass, sSFR and satellite mass simultaneously.

Figure 5.2 shows H I fraction versus stellar mass (top panel) and sSFR (bottom panel) where the dashed coloured lines represent the average stacked gas fraction in each of the halo mass bins indicated. Note that, in order to increase statistics when selecting by many properties, it is necessary to reduce the number of halo mass bins from four to three. Again, we provide the number of galaxies in the corresponding bin along the bottom of each plot.

In the top panel, we plot the gas fraction-stellar mass relation as function of halo mass, shading between bins of sSFR for each environment. The upper bounds of the shaded regions trace the average gas fraction as a function of stellar and halo mass for galaxies with sSFR/yr $\geq 10^{-10.7}$, whereas those systems with sSFR/yr $< 10^{-10.7}$ form the lower bound. This clearly shows that the residual dependence (0.5 dex on average) upon sSFR remains even when controlling for stellar and halo mass. Non-detections from stacking are plotted at their upper limit and marked with an open inverted triangle. By looking at the either the upper or lower bounds of each shaded region in Figure 5.2a we are comparing satellites at fixed stellar mass and sSFR as a function of environment. For a given satellite mass in the blue cloud and red sequence we still see a decrease (on average 0.2 dex and 0.5 dex respectively) in gas fraction between each environment.

We apply a similar analysis to Figure 5.2b. The dotted lines trace the gas fraction-sSFR relationship in each environment and we split each halo mass interval into low ($M_\star/M_\odot < 10^{10.5}$) and high ($M_\star/M_\odot \geq 10^{10.5}$) stellar mass satellites. The high and low mass relations form the outline of each shaded region on the top and bottom edges respectively. There is no blue shaded polygon because there are no galaxies in our sample that have a host halo mass of $M_h/M_\odot < 10^{12}$ and a stellar mass of $M_\star/M_\odot \geq 10^{10.5}$ (see Figure 5.1a). There is a scatter (0.4 dex on average) introduced to the gas fraction-sSFR relation in halos above $M_h/M_\odot = 10^{12}$ by the residual effect of stellar mass on gas content. Despite the dependency on stellar mass, the second-order effect of environment on H I shown in

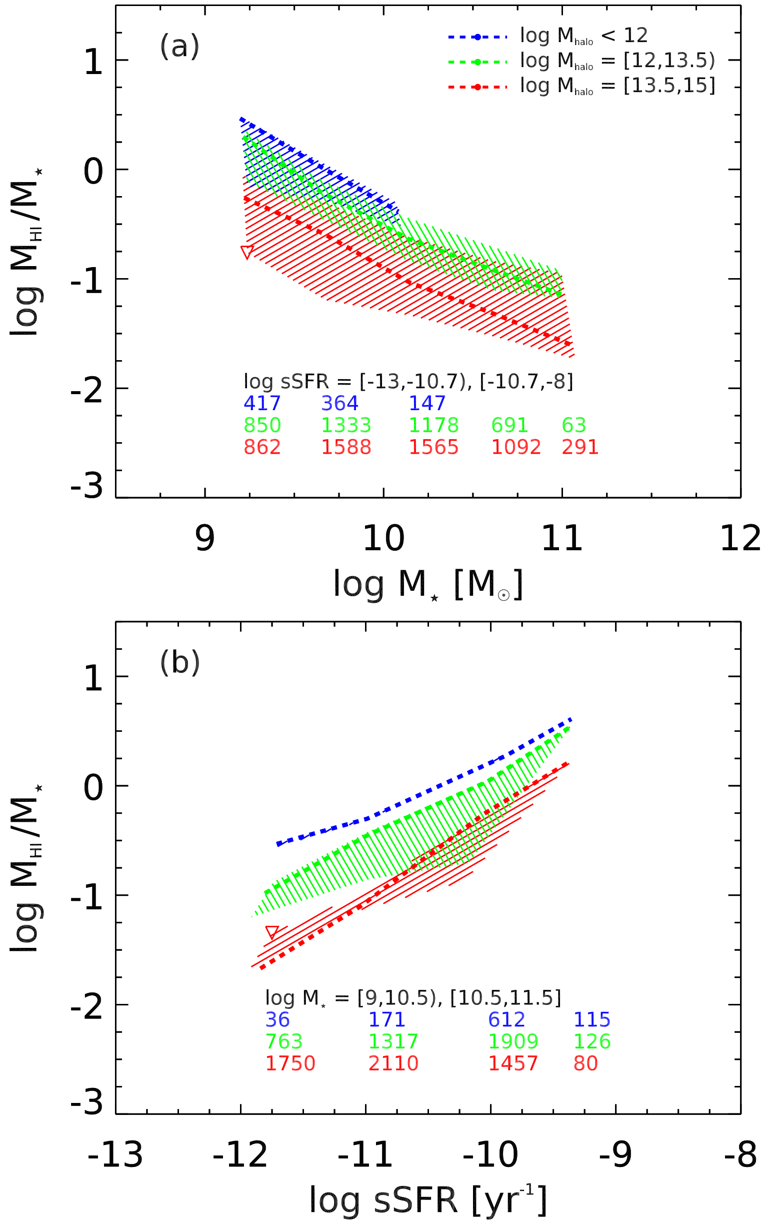

Figure 5.2: Log $M_{HI}/M_\star$ versus log $M_\star$ (a) and log sSFR (b). The sample is binned according to the halo mass limits provided in the legend and plotted by the thick dashed lines. In addition, each halo mass cut in the top and bottom figures is also binned by sSFR and stellar mass respectively. On the top, the upper bound of each shaded region is given by the blue, star forming population (sSFR/yr $\geq 10^{-10.7}$) while the red, quiescent galaxies (sSFR/yr $< 10^{-10.7}$) provide the lower bounds. The bottom panel shows shaded regions for each halo mass bin bounded on the upper edge by galaxies in the low stellar mass bin ($M_\star/M_\odot < 10^{10.5}$) and lower edge by high mass galaxies ($M_\star/M_\odot \geq 10^{10.5}$). Non-detections from stacking are included and are denoted by empty triangles. The numbers shown correspond to galaxies in each halo mass interval only, we do not show the number of galaxies in the stellar mass (b) and sSFR (a) envelopes. In order to ensure good statistics, we do not plot bins containing less than twenty satellites.

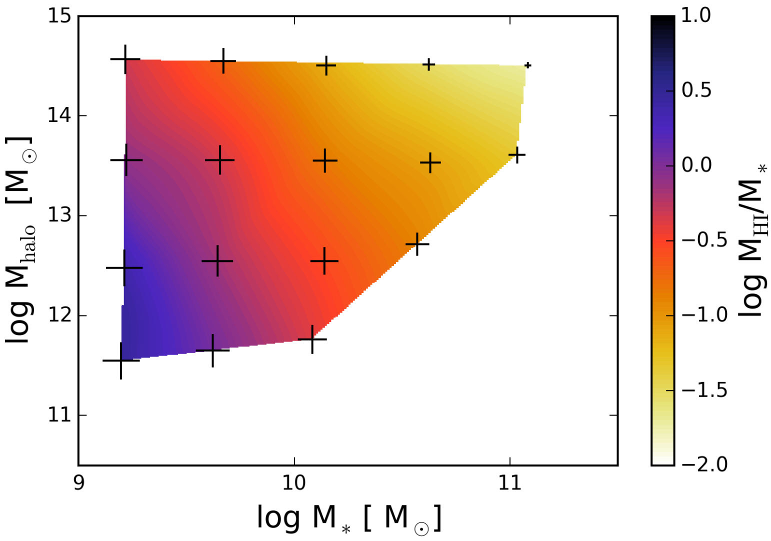

Figure 5.3: Halo mass as a function of stellar mass for satellite galaxies, colour-coded by average gas fraction. There is a co-dependence of gas fraction on both environment and stellar mass. Black crosses are plotted at the mean halo and stellar mass values within each bin. The size of each marker is scaled to the gas fraction at that point.

Figure 5.1b remains. Low and high mass satellites in more massive halos are gas poor (on average 0.3 dex and 0.5 dex respectively) at fixed sSFR compared to their counterparts in less massive halos.

We neatly show the simultaneous effects of stellar mass and environment on gas content in Figure 5.3, plotting the average gas fraction as a function of both host halo mass and satellite stellar mass. For illustration, we interpolate values between the bin centres marked by the black crosses. Contour colours reflect the average H I fraction in that region of parameter space within our sample. The diagonal gradient of the colour change from gas-rich (purple) to -poor (yellow) shows the differential effects of stellar mass and halo mass upon the gas reservoirs of satellite galaxies.

Taken together, these results display the effect of halo mass upon the gas content of satellite galaxies as a function of stellar mass, sSFR and surface density. By holding constant mass and sSFR in Figure 5.2, two galaxy properties shown to have a direct relationship with H I content, we show that there is a strong secondary dependence of gas fraction upon environment across the group regime and into the cluster. In determining this, we break the degeneracy between internal processes that consume gas reservoirs, and external mechanisms which hinder the replenishment or encourage removal of gas from satellites. If satellite gas content is subject to physical mechanisms of an external origin (e.g. hydrodynamical pressure within the halo, gravitational interaction), the offset to lower gas fractions at fixed sSFR for halo masses above $M_h/M_\odot = 10^{13}$ suggests that H I loss is occurring more quickly than the resulting quenching of star formation in these environments.

5.4 The Effect of Mergers on Gas Fraction Scaling Relations

Satellites with a mass similar to that of their group are subject to short dynamical friction timescales and thus increased merger rates (see Chandrasekhar, 1943; Weinberg, 1998; Colpi et al., 1999; Taffoni et al., 2003). In order to verify our definition of a 'satellite' and discriminate between the stripping scenario (as outlined above) and galaxy mergers in driving gas content, we check our results excluding satellites with a ratio between group total stellar mass and satellite mass ($M_{\star,grp}/M_{\star,sat}$) less than ten. This is shown in Figure 5.4, where we plot the gas fraction-stellar mass and sSFR scaling relations, binned by halo mass, for this subset of Sample B (7353 satellites). Despite approximately 30 per cent of satellites residing below this limit, the trend of gas depletion at fixed stellar mass and sSFR remains once these galaxies are removed from the sample. We also note that the rate of confusion within our sample is less than 10 per cent and satellites with a short dynamical friction timescale would most likely be flagged as confused. We therefore rule out mergers as the main driver of gas depletion due to environment at fixed stellar mass and sSFR.

5.5 The Influence of Local Density upon Gas Fraction

Another way to characterise environments within a galaxy population is to use local density metrics. In this section we employ nearest neighbour and fixed aperture estimators in order to understand if our 'definition' of environment is important when determining which environments and processes are the main culprits of the gas depletion seen in satellite galaxies.

Differences in the methods and distributions of halo mass, nearest neighbour and fixed aperture techniques can make direct comparison difficult. Therefore we convert the indicators for each galaxy to a *percentage rank*. To do so we rank the satellites in terms of each metric and assign percentages based upon the orders. For example, a galaxy with a nearest neighbour percentage rank of seventy-five will reside in an environment more dense than 75 per cent of other satellites and less dense than 25 per cent. This method enables us to compare the relative rather than absolute values of each environment metric.

The two gas fraction scaling relations with stellar mass and sSFR, as function of local

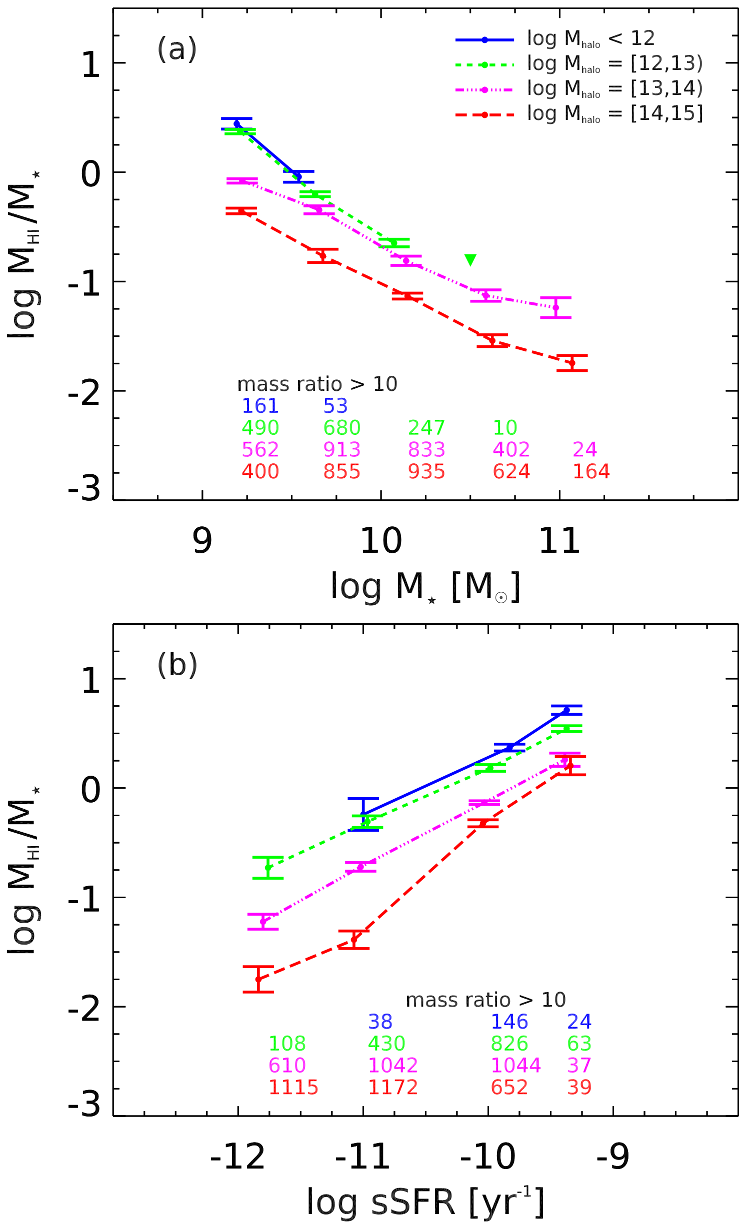

Figure 5.4: Log $M_{HI}/M_\star$ versus log $M_\star$ (a) and log sSFR (b) for Sample B satellite galaxies where $M_{\star,grp}/M_{\star,sat} > 10$. This subset is binned according to the halo mass limits given in the legend. Stacked non-detections are denoted by triangles.

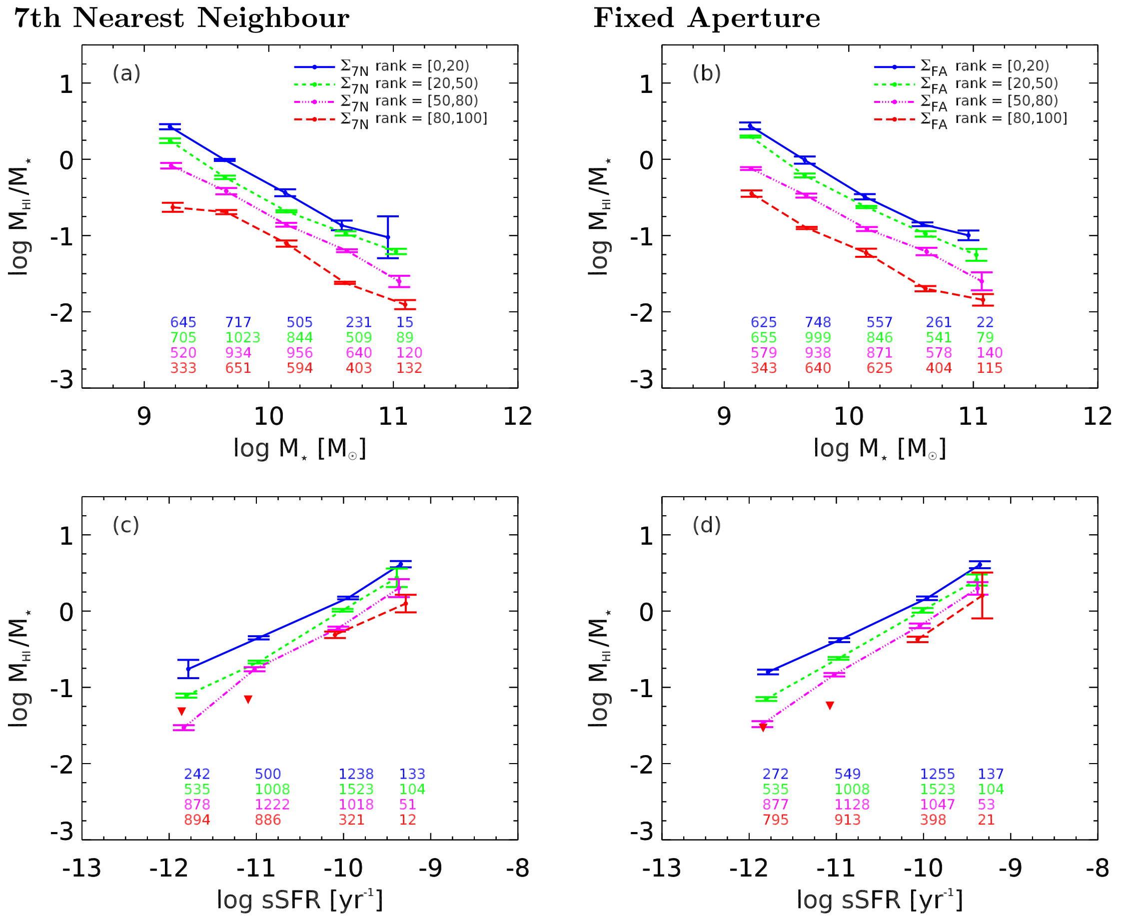

Figure 5.5: On the top, (a) and (b) show log $M_{HI}/M_\star$ versus log $M_\star$ for satellite galaxies. On the bottom, (c) and (d) show log $M_{HI}/M_\star$ versus log sSFR. In the left panels (a, c), data are binned by the percentage rank of their 7th nearest neighbour density. The right panels (b, d) show satellites binned according to their fixed aperture percentage rank. Bin limits are provided in the legends and numbers denote the statistics in each bin and non-detections from stacking are shown as upside-down triangles.

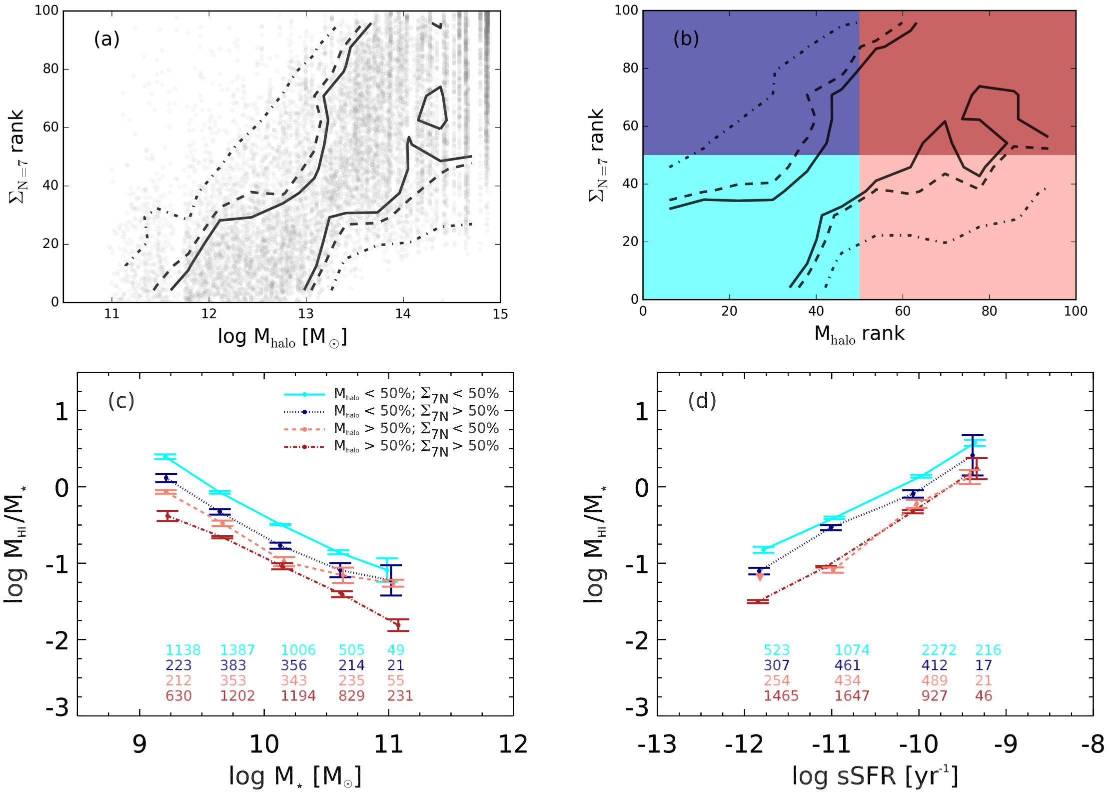

Figure 5.6: 7th nearest neighbour density percentage rank of satellite galaxies as a function of halo mass (a) and 7th nearest neighbour density percentage rank versus log M_h percentage rank (b). Individual satellites in (a) are plotted in grey and contours are plotted at one (solid), two (dashed) and three (dotted) sigma levels in both panels. Log $M_{HI}/M_\star$ versus log $M_\star$, divided simultaneously into bins of halo mass percentage rank and 7th nearest neighbour percentage rank (c) and log $M_{HI}/M_\star$ versus log sSFR in the same bins (d). In these panels, the bounds of each bin are provided in the legend and the parameter space is illustrated by the corresponding colour panel in (b). Non-detections from stacked averages are denoted using upside-down triangles and the number of galaxies in each bin is shown along the bottom.

density percentage rank, are shown in Figure 5.5. On the left, the coloured lines denote the sample galaxies separated according to their 7th nearest neighbour density percentage rank. On the right, the sample density is calculated using to a fixed aperture of radius = 1 Mpc (± 1000 km s^{-1}). Again, numbers shown provide the count of galaxies in the relative bins and stacked averages resulting in non-detections are denoted by upside down triangles.

Figures 5.5a and 5.5b show that, in both cases, satellites that reside in denser environments are significantly more gas poor at a given stellar mass than galaxies that are found in less dense regions. Between the sparsest and densest regions there is a steady progression (0.6 and 0.7 dex for nearest neighbour and fixed aperture respectively) from high to low average gas fractions. In Figures 5.5c and 5.5d we show that gas fraction also varies as a function of nearest neighbour and fixed aperture densities at fixed sSFR. For galaxies at a given sSFR, there is a significant decrease (0.6 dex) in H I content with increasing bins of both nearest neighbour and fixed aperture density.

We now look to determine whether the suppression of gas in the denser regions occurs because of the increase in galaxy number density and therefore chance of interaction, or is the consequence of the correlation between local density and halo mass. Figure 5.6a is a contour density plot showing percentage rank of the 7th nearest neighbour density versus halo mass for the sample of satellite galaxies. Grey points are individual galaxies and contours are set at the one, two and three sigma levels. There is a clear correlation between local density rank and halo mass with denser regions preferentially populating higher halo masses and visa versa. The interdependency is shown further once halo mass is ranked in the same manner by Figure 5.6b. For reference, the 50th percentile of the ranked halo masses corresponds to $M_h/M_\odot \sim 10^{13.5}$ in absolute value, while the 20th and 80th percentile correspond to $M_h/M_\odot \sim 10^{12.5}$ and $M_h/M_\odot \sim 10^{14.5}$ respectively.

We divide the parameter space of Figure 5.6b into colour shaded quadrants that correspond to the coloured relations of gas fraction as function of stellar mass and sSFR in Figures 5.6c and 5.6d respectively. For example, galaxies that are found in the dark red quadrant are included in the dark red, dot-dashed scaling relations, residing in the densest half of the sample and above the 50th percentile for halo mass. Numbers along the bottom in Figures 5.6c and 5.6d are the statistics in each bin.

Figure 5.6c shows that for a given stellar mass, gas reduction goes with both changes in local density and halo mass (0.6 dex). However, when one fixes density and alters the halo mass, comparing the cyan relation with the salmon and navy with the dark red, differences are larger (0.4 dex) than when density is changed at fixed halo mass (0.25 dex, cyan-navy/salmon-dark red). Interestingly, in Figure 5.6d, the effect is more noticeable. *At fixed sSFR, it is clear that gas fraction preferentially decreases with halo mass rather than density.* The differences in gas content between bins of varying halo mass (cyan-salmon/navy-dark red) are large (0.5 dex), while there is less of a difference ($<$ 0.2 dex) when only density changed. Figure 5.6 clearly shows that the observed environmental suppression of gas in galaxies at fixed stellar mass and sSFR is dominated by the halo mass in which they reside and not the density of neighbours.

As previously mentioned, we check the validity of these results and the sensitivity to the aperture used by performing the same study out to the 3rd, 5th and 10th nearest neighbour, and with fixed apertures of 1 Mpc ($\pm$ 500 km s^{-1}), 2 Mpc ($\pm$ 1000 km s^{-1}) and 2 Mpc ($\pm$ 500 km s^{-1}). Changing the aperture size does not significantly affect our results, because of this we do not show the additional figures.

5.6 Discussion and Conclusions

In this chapter we applied the H I spectral stacking technique to study the effect of environment on the gas content of 10,567 satellite galaxies, selected by redshift and stellar mass from the intersection of SDSS and ALFALFA. We quantified environment using FoF, nearest neighbour and fixed aperture metrics. The FoF-based DM halo masses and group association are provided by the Y07 galaxy group catalogue, while 7th nearest neighbour and fixed aperture densities are computed separately. The main conclusions of this work can be summarised as follows:

i) Satellite galaxies in more massive halos have, on average, lower H I-to-stellar mass ratios at fixed stellar mass, surface density and sSFR than those in smaller halos. The significant and systematic decrease in the gas content of satellites as a function of halo mass occurs across the entire group regime as well as the cluster environment.

ii) Following this, we suggest that the average timescale for H I loss from satellites in halos with masses above $M_h/M_\odot \geq 10^{13}$ is considerably faster than the subsequent quenching of star formation.

iii) Gas content is also depleted with increasing nearest neighbour and fixed aperture densities. However, halo mass is the dominant environmental driver of H I removal in satellites.

We show that significant and continuous suppression of satellite H I content due to environment is present in halos more massive than $M_h/M_\odot \sim 10^{12}$ at fixed mass and morphology, and above $M_h/M_\odot \sim 10^{13}$ at fixed sSFR. By controlling for influences of mass, morphology and star formation, and separating those from the effect of environment, we present a scenario whereby environment driven processes are directly acting upon the H I reservoirs of satellites. Observations have previously shown mechanisms such as strangulation, interaction and stripping to be prevalent in large galaxy groups and clusters. In order to explain our result of decreasing gas fractions in halos of $M_h/M_\odot \geq 10^{12}$ we suggest that one or more of these processes becomes efficient in small to mid-size groups, well before the large group and cluster regimes. Encouragingly, this is entirely consistent with the results of Stark et al. (2016) who show a decrease in satellite gas content in halos above $M_h/M_\odot \sim 10^{12}$ at fixed stellar mass using individual detections.

From the evidence presented it is possible to take this analysis a step further and speculate upon the prominence of these different mechanisms, a subject that is still very much up for debate. To do so we divide the processes into two categories: slow acting, such as strangulation or starvation, on the one hand; and fast acting, which primarily refers to ram-pressure stripping, on the other. The path that galaxies trace through the gas fraction-sSFR plane (Figure 5.1b) when under the influence of these two categories differs considerably. A slow reduction in atomic gas naturally eventuates in a reduction of the molecular phase as H_2 is consumed by star formation and not replenished. This results in a steady decline in star formation ($\gtrsim$ 1 Gyr, Balogh et al., 2000) as the available fuel reservoirs diminish and galaxies transition simultaneously to low gas fractions and onto the red sequence. In Figure 5.1b, this is seen in halos with $M_h/M_\odot < 10^{13}$ where differences in gas content as a function of *both* halo mass and sSFR are not significant.

In contrast, one would expect that when removal of H I is (nearly) instantaneous ($\gtrsim$ several 10 Myr, Vollmer et al., 2012), there is a necessary time lag before star formation is quenched accordingly and the chosen star formation indicator registers a change (Obreschkow & Rawlings, 2009). Galaxies that undergo such an event as they move into larger halos will exhibit lower gas fractions at fixed sSFR between halo mass bins. It is for this reason that ram-pressure stripping is our preferred explanation for the lower gas fractions seen in halos above $M_h/M_\odot = 10^{13}$ at fixed sSFR.

We caution the reader that, as stated in Chapter 2, the SFR indicator used in this chapter (Brinchmann et al., 2004) is a combination of fits to the Hα emission lines from star forming galaxies and, where emission in Hα is low, the break in galaxy spectra around 4000 Å. While the Hα emission traces star formation over the last few tens of Myr, the D4000 measurement is sensitive on longer timescales of $\sim$1 Gyr. This means that, where the D4000 is used, we may simply conclude that the gas depletion occurs at a faster rate than

1 Gyr at fixed sSFR. However, for star forming galaxies, depletion of H I with environment at fixed sSFR is indicative of an even more rapid removal of gas and ram-pressure stripping is favoured. This effect contributes to the broadening of the gas fraction-sSFR relation seen between the blue and red sequence in Figure 5.1b.

In addition to FoF based halo masses, we also conducted our analysis using 7th nearest neighbour and fixed aperture densities. We have shown that, for the satellite population within our sample, there is a tight correlation between halo mass and local density metrics and, in Figure 5.5, confirm that galaxies in denser environments tend to be more gas poor. We test whether gas is preferentially depleted by processes pertaining to either the halo mass (hydrodynamical) or the proximity of neighbours (gravitational) in Figure 5.6, showing that, at fixed stellar mass and sSFR, the reduction in H I follows an increase in halo mass and over local density. This paints a physical picture where halo mass is the dominant factor in environment driven cold gas depletion and agrees with our hypothesis above - ram-pressure stripping is the likely candidate for H I removal in massive halos ($\mathrm{M_h/M_\odot} \geq 10^{13}$).

Having said this, it is important to check for the possibility of mergers driving the observed gas depletion with halo mass. In particular, for systems where the group-to-satellite mass ratio is low, dynamical friction timescales are short and mergers can occur rapidly. We discussed the impact of this effect on our results in Section 5.4. At both fixed stellar mass and fixed sSFR, the trend of depletion with halo mass remains and we also note that the rate of confusion in our sample is <10%. It is likely that, in the majority of cases, systems with short dynamical friction timescales would be flagged as confused.

To conclude, this chapter has provided evidence that the gas content of satellites is depleted by external processes as they transition into higher mass halos and denser environments. Using observations, we discuss the likely processes at work, suggesting that fast acting hydrodynamical mechanisms such as ram-pressure stripping are efficient in the group environment as well as the high density clusters. The results also exemplify the gains to be made by using the H I spectral stacking technique, especially as we look to pave the way for the next generation of radio telescopes that will address these important questions.

6

Comparing Observations with Theoretical Predictions

6.1 Introduction

Using key H I-to-stellar mass scaling relations, Chapter 5 has demonstrated that systematic environmental suppression of gas content at both fixed stellar mass and fixed specific star formation rate (sSFR) in satellite galaxies begins in halo masses typical of the group regime ($M_h/M_\odot < 10^{13}$), well before galaxies reach the cluster environment. Further to this, in Section 5.5, we determined that there is also a trend to lower gas fractions with increasing nearest neighbour and fixed aperture local density, at fixed stellar mass and sSFR, however, we also show that environment driven gas depletion is more closely associated to halo mass than local density.

These results extend findings from previous studies, showing that lower H I fractions are found in the cluster environment (e.g. Giovanelli & Haynes, 1985b; Chung et al., 2009; Cortese et al., 2011), adding to the picture by demonstrating the systematic transition from gas-rich to -poor that occurs within groups, well before galaxies reach the cluster. Other observational (Cortese et al., 2011; Catinella et al., 2013) and theoretical (McCarthy et al., 2008; Rafieferantsoa et al., 2015) efforts support this picture, with both camps generally invoking external processes that are distinguishable based upon the time scales over which they act: (i) Those that act swiftly and directly upon the gas to remove it from the galaxy via interaction of the ISM and intergalactic-medium ($\sim$ 10s Myr; i.e. ram-pressure stripping; Gunn et al., 1998), or (ii) those that regulate the rate at which gas is able to accrete onto the galaxy from its dark matter halo over more lengthy time scales ($\gtrsim$1 Gyr; i.e. strangulation; Larson et al., 1980). Using this distinction, Chapter 5 provides strong evidence for the gas loss in massive halos ($M_h/M_\odot > 10^{13}$) is considerably faster than the subsequent quenching of star formation.

Open questions remain surrounding the dominant processes that are responsible for the observed differences in gas fraction between low, intermediate and high mass halos. We now look to identify the 'fundamental' physics that is driving the environmental dependence seen in the scaling relations by connecting our empirical results to theoretical work.

The two most common approaches for conducting such a comparison are semi-analytic models and hydrodynamical simulations. In both cases, correctly modelling the influence of environment on the H I content of galaxies is an extremely complex problem. Until relatively recently, much of the success in this area came from non-cosmological, high resolution simulations of well resolved galaxies (e.g. Marcolini et al., 2003; Mayer et al., 2006, 2007; Bekki, 2009; Tonnesen & Bryan, 2009, 2010). In recent years, the ability of cosmological models to reproduce observed trends in the global gas content of galaxies has improved significantly and successful comparisons have been made between theory and observations for general gas properties using both semi-analytic (e.g. Obreschkow et al., 2009; Power et al., 2010; Popping et al., 2014; Lagos et al., 2014) and hydrodynamical simulations (e.g. Davé et al., 2013; Rafieferantsoa et al., 2015; Crain et al., 2017). This success in replicating the global trends of gas content with key galaxy properties means it is important that we attempt understand how the models perform at reproducing the effect of environment upon the gas content of galaxies. In a recent study, Marasco et al. (2016) investigate the role of environment in dictating the H I content of galaxies within the EAGLE (Schaye et al., 2015) suite of cosmological simulations. Their study successfully reproduces the gas fraction scaling relations of Fabello et al. (2012) and Catinella et al. (2013), demonstrating that EAGLE galaxies in more massive halos tend to have lower gas content at fixed stellar mass. However, rather than a continuous trend of H I depletion as a function of environment found in our work, their investigation finds that environment effects in EAGLE drives a bimodal distribution in H I mass.

For our comparison, we choose one cosmological model from each camp; the semi-analytic model GALFORM (Gonzalez-Perez et al., 2014) and the hydrodynamic simulations of Davé et al. (2013), both have previously published successful comparisons with observational HI scaling relations (see Lagos et al., 2011b; Davé et al., 2013; Rafieferantsoa et al., 2015). These are introduced in Sections 6.2 and 6.3 respectively, summarising the methodology of each and how they evolve the gas content in galaxies.. Section 6.4 discusses our results in the context of this theoretical framework, showing that more work is needed if models are to reproduce our observations.

6.2 Semi-analytic Models

Semi-analytic models of galaxy formation typically treat each galaxy as a single object, using integrated properties and prescriptions to describe the baryonic physics governing their evolution. The primary advantage of this technique is its computational efficiency, allowing the production of statistical samples of galaxies that cover representative volumes and a large parameter space. One caveat is that bulge and disk properties (gas, stars, SFR etc.) are described by a single number for each component, meaning that the internal dynamics and physics are not resolved.

In this subsection we compare our results with the semi-analytic simulation GALFORM (Gonzalez-Perez et al., 2014). The motivation for choosing this model is its environmental treatment of the hot gas and tracking of the cold gas as galaxies evolve.

6.2.1 The GALFORM Model

The GALFORM semi-analytic model includes the main physical processes that are considered to shape the formation and evolution of galaxies (Cole et al., 2000). These are: (i) the collapse and merging of DM halos, (ii) the shock-heating and radiative cooling of gas inside DM halos (which lead to the formation of galactic disks), (iii) star formation in disks, (iv) feedback from supernovae, from active galactic nuclei and from photo-ionization of the inter-galactic medium, (v) chemical enrichment of stars and gas due to stellar evolution, (vi) galaxy mergers driven by dynamical friction (which trigger starbursts and lead to the formation of bulges), (vii) global disk instabilities (which also lead to the formation of bulges), and (viii) ram-pressure stripping of the hot gas. For this chapter we focus on the published version of GALFORM of Gonzalez-Perez et al. (2014), hereafter GP14.

In GP14, the halo merger trees are extracted from the updated version of the Millennium N-body simulation (Springel, 2005) using WMAP7 (Komatsu et al., 2011). Gonzalez-Perez et al. (2014) also includes the explicit tracking of the atomic and molecular cold gas component in galaxies (by using the hydrostatic midplane pressure of disks and bulges as a proxy of the atomic-to-molecular gas surface density ratio; Blitz & Rosolowsky, 2006; Leroy et al., 2008) as well as the hot gas phase. Star formation is characterised following Blitz & Rosolowsky (2006), who use an empirical formulation of the Kennicutt-Schmidt law:

$$\Sigma_{\rm SFR} = \tau_{\rm SF} \Sigma_{\rm H_2} \tag{6.1}$$

where $\Sigma_{\rm SFR}$ is the star formation rate surface density in units of $\rm M_{\odot}\ pc^{-2}\ Gyr^{-1}$, $\tau_{\rm SF}$ is the inverse star formation timescale in Gyr^{-1} and $\Sigma_{\mathrm{H_2}}$ is the surface density of molecular gas in $\mathrm{M_\odot\ pc^{-2}}$. The current implementation was developed by Lagos et al. (2011b), further details can be found in that paper.

Here, we present results from two variants of the GP14 model, which differ in their treatment of the hot gas of satellite galaxies once they cross the virial radius of the larger halo. The first variant (which we simply refer to as GP14 as it is the default implementation) assumes *instantaneous* stripping of the hot gas; once a galaxy becomes a satellite, its hot gas is removed and transferred to the gas reservoir of the main halo. By removing satellite hot gas, GP14 halts the replenishment of cold gas reserves through accretion, forcing each galaxy into a state of 'strangulation'. The second variant, which we refer to as GP14+GRP, assumes instead *gradual* ram-pressure stripping of the hot gas, which in practice leads to satellite galaxies having accretion rates that continuously decay in time once they become satellites (as opposed to a sharp shut off of the accretion rate in the first variant). GP14+GRP defines the 'stripping radius' as the point at which ram-pressure and gravitational pressure are balanced, gas that resides beyond this radius is removed from the satellite. The ram and gravitational pressure prescriptions are detailed in Lagos et al. (2014) and follow the respective functional forms

$$\mathrm{P_{ram}} = \rho_{\mathrm{gas,p}} v_{\mathrm{sat}}^2 \tag{6.2}$$

and

$$\mathrm{P_{grav}} = \alpha_{\mathrm{rp}} \frac{G\mathrm{M_{tot,sat}}(r_{\mathrm{str}})\rho_{\mathrm{gas,sat}}(r_{\mathrm{str}})}{r_{\mathrm{str}}} \tag{6.3}$$

where $\mathrm{P_{ram}}$ is the ram-pressure, $\rho_{\mathrm{gas,p}}$ is the gas surface density of the parent halo in $\mathrm{M_\odot\ pc^{-2}}$ and v_{sat} is the satellite's velocity with respect to its parent halo. $\mathrm{P_{grav}}$ is the gravitational pressure, α_{rp} is a geometric constant of order unity and set to 2, G is the gravitational constant, $\mathrm{M_{tot,sat}}(r_{\mathrm{str}})$ is satellite's total mass (including stars, gas and DM) within the stripping radius r_{str} and $\rho_{\mathrm{gas,sat}}(r_{\mathrm{str}})$ is the satellite's hot gas density at r_{str}.

It is important to note that neither implementation of GP14 used in this work accounts for ram-pressure stripping of the cold gas, a process known to dominate in larger halos (i.e. galaxy clusters; Boselli & Gavazzi, 2006; Tonnesen et al., 2007; McCarthy et al., 2008; Chung et al., 2009; Cortese et al., 2010, 2011). Other processes expected to drive mass loss in satellite galaxies (e.g. tidal stripping of stars, heating due to tidal shocks, harassment) are not included in either of the variants. In cases where the dynamical mass of the satellite is significantly below that of the group, such mechanisms have been shown to remove cold gas in significant amounts over many Gyrs, however, such effects are expected to be dominant in dwarf galaxies, i.e. those with circular velocities $\lesssim$ 30 km s^{-1}, which are smaller than the galaxies we are studying here (see Mastropietro et al., 2005; Mayer et al., 2006; Tomozeiu et al., 2016).

The GP14+RP variant has been shown to produce both atomic and molecular gas fractions of early-type galaxies, and fractions of passive galaxies as a function of stellar mass that are in better agreement with the observations (Lagos et al., 2014; Guo et al., 2016).

6.3 Hydrodynamical Simulations

Compared to semi-analytics, hydrodynamical simulations reproduce the processes that govern galaxy evolution at much higher resolution, solving the equations of gravity, dynamics and radiative transfer for up to 10^6 particles in each galaxy. With the caveat that mechanisms such as star formation, stellar feedback, black hole accretion still occur below the resolution limit, this approach has the advantage of galaxies being resolved into several elements and no imposed assumptions of how gas accretion takes place or the influence of DM on galaxy properties (e.g. sizes and accretion rates). These models have been successful in their reproduction of realistic objects, however, modelling such detailed physics on small scales is computationally very expensive, therefore sample size and parameter space explored is smaller than that of semi-analytics.

6.3.1 Davé et al. (2013) "ezw" Simulation

We compare to the cosmological hydrodynamic simulations of Davé et al. (2013), who used a modified version of Gadget-2 (Springel, 2005) to study the H I content of galaxies. The simulation uses a slightly different ΛCDM cosmology to the one we assume in the observations, with $\Omega_{\rm M} = 0.28$, $\Omega_\Lambda = 0.072$, $\Omega_{\rm baryons} = 0.046$, $\sigma_8 = 0.81$, $n_s = 0.96$ and $h = 0.70$. At $z = 0$ the difference in galaxy properties between these two cosmologies is negligible. There are 512^3 DM particles and 512^3 gas particles in a cubical, comoving volume of 32 h^{-1} Mpc on each side. The DM and gas particle resolution is 2.3×10^7 M$_\odot$ and 4.5×10^6 M$_\odot$ respectively. In the model halos and galaxies grow self-consistently from DM and gas particles. Here we briefly summarise the processes of galaxy formation and evolution modelled in the simulation: (i) primordial and metal line cooling based upon the photo-ionization equilibrium of Wiersma et al. (2009), (ii) star formation following a Schmidt law (Schmidt, 1959) where SFR is proportional to gas density, applied using the sub-grid recipe of Springel & Hernquist (2003), (iii) stellar and supernovae feedback provides metal enrichment following the prescriptions of Oppenheimer & Davé (2008), (iv) quenching energy comparable to AGN feedback is imparted on massive galaxies by heating infalling gas (to fifty times the virial temperature) once the halo mass, estimated from the individual galaxy mass, exceeds around $M_h/M_\odot = 10^{12}$, (v) finally, strong galactic outflows (their "ezw" model), driven through a hybrid of momentum flux from young stars and energy from supernovae, are assumed to kinetically eject gas from the ISM.

Each halo is identified using a spherical over-density based approach, while galaxies are identified using Spline Kernel Interpolative Denman (SKID; see Davé et al., 2013). For halos with multiple resolved galaxies (which is the majority for galaxies with $M_\star/M_\odot \geq 10^9$), the largest stellar mass galaxy is identified as the central, and the others are satellites. H I is computed within the model by determining the optically thin, neutral fraction of each gas particle. Davé et al. (2013) then separate the neutral gas into its atomic and molecular phases based upon the ISM pressure prescriptions of The H I Nearby Galaxy Survey (THINGS; Leroy et al., 2008).

Once a satellite enters the halo of another galaxy, the H I may be influenced by the following environmental processes, each of which are modelled self-consistently within the simulation: ram-pressure and viscous stripping (Marcolini et al., 2003), tidal interaction and harassment, and strangulation of inflowing gas. For further details on how these mechanisms are implemented and their dependence on halo properties see Rafieferantsoa et al. (2015). An important caveat to note is that Davé et al. (2013) employ entropy-conserving smoothed particle hydrodynamics (SPH) (Springel, 2005) that has been shown to handle surface instabilities such as those that occur during gas stripping poorly compared to more recent hydrodynamics methods (see Agertz et al., 2007; Hopkins, 2015; Schaller et al., 2015). This issue and its possible implications are discussed further in Section 6.4.

6.4 Comparison with Observations

In Figure 6.1 we compare our observations of gas content at fixed stellar mass (top panels) and sSFR (bottom panels) with the models. The H I gas content of satellite galaxies in the simulations is calculated in a way that is identical to our stacking procedure (i.e. $\log \langle M_{HI}/M_\star \rangle$). We split satellites according to identical bins of halo mass for both observations (blue, green, red) and theory (cyan, dark green, dark red). The number of galaxies in each observational bin is plotted along the bottom.

We see that the GP14 model predicts satellites that are too gas poor at fixed stellar mass (6.1a) and too gas rich at fixed sSFR (6.1d). While this seems in contradiction, the

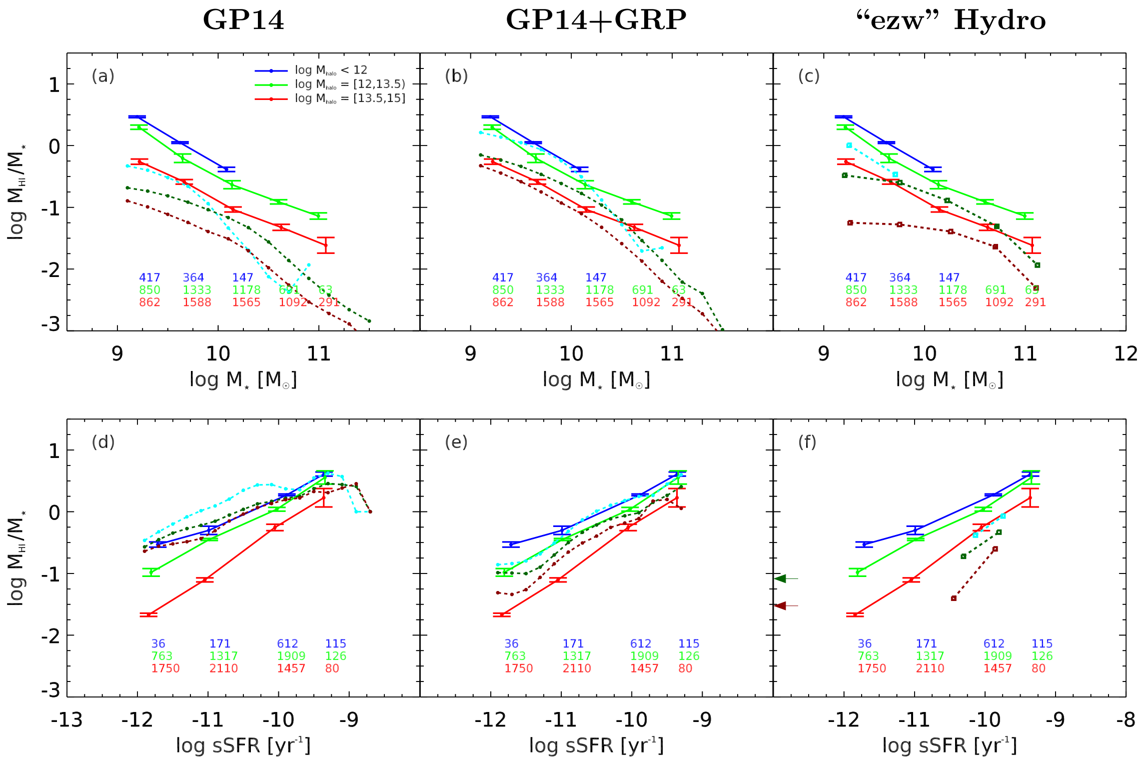

Figure 6.1: H I fraction versus stellar mass ($M_\star$, top) and specific star formation rate (sSFR, bottom) for satellite galaxies separated by halo mass. Solid lines are the same observations as the dashed lines shown in Figure 5.2. Each column over plots a different galaxy formation model with cyan, dark green and dark red dashed lines corresponding to the same halo mass bins as the blue, green and red observational relations. Left: GP14 semi-analytic model, which assumes that once galaxies cross the virial radius of a larger halo they instantaneously lose their hot gas content and prevents further gas accretion. Middle: GP14+GRP model, which adopts gradual ram-pressure stripping of the hot gas and continued accretion of gas onto the galaxy. Right: Dashed lines show the "ezw" hydrodynamical simulation of Davé et al. (2013), this models the stripping of the ISM from the disk. Arrows show upper limits on sSFRs set to the observational limit.

process of instantaneously removing a galaxy's hot gas from its subhalo once it becomes a satellite leads to an overly quenched satellite population. Naturally, these systems have low or negligible amounts of cold gas which means that, where they are included in the scaling relations (i.e. 6.1a), average gas fractions are artificially low. On the other hand, extremely quenched systems are removed from the gas fraction-sSFR plane by construction. Therefore they do not contribute to the average H I content and one is not comparing the same GP14 satellite population between Figures 6.1a and 6.1d. Once this is added to the picture, the two plots are in agreement and it becomes clear that the stripping implemented by GP14 is too strong and too rapid to match observations.

In Figure 6.1b we find that the GP14+GRP model, which assumes the hot gas gets stripped gradually as satellite galaxies travel through their host halos, is in better agreement with the observations for stellar masses of $M_{\star}/M_{\odot} \lesssim 10^{10.3}$ in the highest and lowest halo mass bins studied. This is because in the GP14+GRP model, satellite galaxies are able to retain their hot gas for a longer timescales, which in turn means that they are able to replenish their ISM for longer. Nevertheless, at these stellar masses, the GP14+GRP model is still on average $\sim$0.2 dex too low compared to observations at fixed stellar mass. This is because the models predict sSFRs that are usually lower than the observed sSFRs for fixed stellar masses at $z \approx 0$ (see for instance Mitchell et al. 2014). The effect is seen clearly in the two largest halo mass bins and at stellar masses above $M_{\star}/M_{\odot} \sim 10^{10.5}$. We expect that the low gas fractions are due to hot gas stripping being too severe, leading to starvation and quenching that is too strong. Note that GP14+GRP is too gas poor even without the inclusion of ram-pressure stripping of the cold gas which, as stated previously, has been shown to be an important driver of gas removal in this regime (i.e. cluster scales). At high stellar masses, depletion of gas is caused by AGN feedback being too strong, not allowing further cooling and replenishment of the ISM.

When we study H I gas fraction as a function of sSFR we find GP14+GRP delivers a better overall agreement with our measurements than gas fraction as a function of stellar mass. However, the predicted population of galaxies with very low sSFRs - and likely low gas fractions - is not visible in the parameter space shown. The predictions for halos of $M_h/M_{\odot} < 10^{13.5}$ (blue and green lines) are particularly successful when we compare with our observations. At higher halo masses, the model predicts H I gas fractions at fixed sSFR that are slightly too high compared to our observations. This is likely due to GP14+GRP not accounting for the ram-pressure stripping of H I, which is expected to significantly drive down gas fractions in this regime at both fixed sSFR and stellar mass. However, the inclusion of this effect would potentially increase tension between GP14+GRP and our observations.

For the comparison between observations and GP14+GRP, we tested the effect of using halo masses assigned via the abundance matching method rather than using the halo masses from the simulation on the scaling relations presented here. We follow the method of Yang et al. (2007), calculating the total stellar mass from all the galaxies in a halo that have absolute r-band magnitude M_r - 5 10^{12}(h) $\leq$ -19.5, ranking the groups using their integrated stellar mass and assigning a halo mass under the assumption that there is a one-to-one correspondence between the integrated stellar mass and halo mass. We found that the abundance matching method slightly under-predicts ($\sim$0.1 dex) gas fractions at fixed stellar mass for halo masses below $M_h/M_\odot = 10^{13.5}$, increasing the disagreement with observations. However, gas fractions at fixed sSFR show no significant differences between the samples using the two different halo masses.

Figures 6.1c and 6.1f show a comparison between the H I fraction of satellites, as a function of halo mass, between the simulations of Davé et al. (2013) and observations, against stellar mass (6.1c) and sSFR (6.1f). The strongly bimodal distribution of satellite sSFRs in the Davé et al. (2013) model means that quenched galaxies lie off the parameter space of Figure 6.1f to lower sSFRs. We compute the average gas fraction of these galaxies (coloured arrows) and set them to our observational sSFR limit of sSFR/yr = 10^{-13}. H I fractions of satellites in the simulation are systematically low ($\sim$0.6 dex) when compared to observations and, while some qualitative agreement exists, there is limited reproduction of the general trends with stellar mass and sSFR as a function of halo mass.

The fact that disagreement is present even in low mass halos where stripping is inefficient in these simulations (Rafieferantsoa et al., 2015) suggests that the origin of this deficit is likely endemic to the satellite population in the Davé et al. (2013) simulations. However, the explanation for this deficit is not straightforward, thus, below we briefly outline various possible causes, both physical and numerical.

Although the H I content is greatly underestimated, the depletion of gas as a function of halo mass in the simulation somewhat echoes the observed trend, suggesting that key H I removal processes are roughly followed. As shown in Rafieferantsoa et al. (2015), at fixed halo mass, satellite H I masses deviate from their stellar mass-matched central already at $M_h/M_\odot \sim 10^{11.5}$, while the data seems to suggest that such deviations do not begin until higher halo masses. This discrepancy may owe to overly-aggressive stripping or starvation within fairly low-mass halos, which then propagates to higher masses. Further to this, previous work using high resolution, non-cosmological simulations has shown stripping scenarios to be strongly dependent on both galaxy structure and the ISM model employed

(Mastropietro et al., 2005; Marcolini et al., 2003).

From a numerical perspective, replicating the hydrodynamical interaction between H I and the surrounding intracluster or intragroup medium is a notoriously difficult task. While remaining a dramatic improvement on the detail afforded by semi-analytics, the moderate resolution of Davé et al. (2013) simulations means that there is a possibility of dark matter particles being spuriously heated due to two-body interaction, this leads to artificial heating and momentum transfer to gas particles in the simulation (see Steinmetz & White, 1997; Abadi et al., 1999). In addition and as briefly mentioned in Section 6.3.1, the version GADGET used does not include newer SPH recipes that are required to correctly capture the fluid instabilities at this interface. While this is an undoubted shortcoming of the simulation, it results in the employed SPH underestimating the effect of ram-pressure as well as other suppression mechanisms (i.e. viscous stripping, Kelvin-Helmholtz instabilities), thus discrepancies are expected to worsen with the inclusion of new SPH, not improve. Schaller et al. (2015) compared the old SPH formulation with newer SPH in the EAGLE (Evolution and Assembly of GaLaxies and their Environments; Schaye et al., 2015; Crain et al., 2015) simulations and found that the amount of cold gas and therefore the star formation rates of galaxies are reduced in the new SPH formulation, evidencing that cold gas fractions would become even lower than those found in our comparison. However, Schaller et al. (2015) also show that other numerical aspects, such as the timestep limiter and how the sub-grid physics modules are implemented have significant effects on the properties of galaxies. The latter means that it is not straightforward to estimate how much the cold gas fractions may be affected in the Schaller et al. (2015) simulations, and instead direct testing is necessary in the future.

Having compared our observations to theory, we see that models and simulations are producing far too many gas poor galaxies. The results show that considerable modifications are required if we are to successfully characterise the impact of environment on the H I content of galaxies.

6.5 Discussion and Conclusions

The theoretical comparison presented in Figure 6.1 goes some way towards supporting the picture presented by observations. However, this work highlights the tendency for cosmological models to produce satellite populations that are too gas poor and therefore excessively quenched. Our main finding is that, at fixed stellar mass and sSFR, both semi-analytics and hydrodynamic produce too many gas poor satellites. There is, however, qualitative agreement that the gas content of satellite galaxies depends upon the mass of their DM halo and depletion, particularly at fixed sSFR, is caused by a stripping mechanism.

The semi-analytic model GP14+GRP performs best in reproducing the observed trends of gas fraction with stellar mass and sSFR. Having said this, if ram-pressure stripping of H I were to be included in the model we would expect the agreement to deteriorate as gas fractions are further reduced. In addition, at fixed sSFR, the separation in gas content between the bins of halo mass is not well recovered. We interpret this as evidence that a fast acting process acting directly upon the H I being required to deplete gas content as a function of halo mass at fixed sSFR. A final caveat is that other processes that drive mass loss in all components of satellites (i.e. stars, gas, DM) after infall are neglected in the model. While it is true that these processes may be responsible for gas removal in systems where the group-to-satellite mass ratio is low (see Mayer et al., 2006; Font et al., 2008; McCarthy et al., 2008), our check described above shows that this effect is not driving the observed trends. On top of this, the timescales over which these processes are expected to remove gas is significantly longer that of ram-pressure stripping, thus the effect is not likely to be present in the observations at fixed sSFR.

The advantage of the Davé et al. (2013) hydrodynamical simulations is that they include prescriptions for ram-pressure stripping of cold gas from the satellite disk. To first order, the gas depletion at fixed stellar mass and sSFR found in the observations is also present in the model. However, the large offset to low gas fractions is worrying as it is considerable compared to observed trend, even in the lowest mass halos where environmental processes are not expected to play a significant role.

Recently, Marasco et al. (2016) conducted an analysis of environmental processes that affect the H I content of galaxies in the EAGLE simulations that explicitly addresses some of the numerical concerns raised here. In their work, Marasco et al. (2016) find the fraction of H I poor galaxies increases with halo mass, predicting an environmentally-driven bimodal distribution in the H I-to-stellar mass ratio. The authors find that the most common environmental driver of gas in EAGLE satellites at $z = 0$ is ram-pressure stripping, with tidal forces and satellite-satellite interaction playing a secondary, yet significant role.

In summary, our comparison clearly illustrates that, while general trends of gas content with stellar mass and sSFR are grossly reproduced, there are fundamental inaccuracies in the way that both semi-analytic and hydrodynamical approaches deal with gas content. Significant improvements to are required if they are to match observations.

7

Gas as the Primary Regulator of the Mass-Metallicity Relation

7.1 Introduction

Theoretical work has long predicted that galaxy growth is to a large extent regulated by the balance of gas accretion (either pristine or previously ejected) against star formation, and the subsequent dilution or enrichment of metals (e.g. Tinsley & Larson, 1978; Köppen & Edmunds, 1999). In recent years, with the addition of outflows - the ejection of gas and metals from the ISM via energetic or momentum-driven winds - to this picture, we have begun to see the emergence of a framework for galaxy evolution where galaxies exist in a slowly evolving equilibrium between gas inflow, galaxy-scale outflows and star formation (Finlator & Davé, 2008; Oppenheimer & Davé, 2008; Oppenheimer et al., 2010; Davé et al., 2011a, 2012; Lilly et al., 2013; Christensen et al., 2016).

Observationally, the most common probe of this equilibrium is the relationship between stellar mass and gas-phase metallicity known as the mass-metallicity relation (MZR, see Figure 7.1; Lequeux et al., 1979; Garnett, 2002; Tremonti et al., 2004; Kewley & Ellison, 2008; Zahid et al., 2011). The general sense of the MZR is that metallicity, as traced by O/H, increases linearly with stellar mass up to $M_\star/M_\odot \sim 10^{10.5}$, after which the gradient flattens. There have been a number of mechanisms invoked to explain the observed slope and normalisation of the MZR, including: i) outflows driving enriched gas from galaxies with greater efficiency at low stellar masses, where the shallow depth of the potential well means that material is easily ejected (e.g. Tremonti et al., 2004, hereafter T04), ii) mass-dependent interplay between chemical enrichment and dilution, primarily driven by the correlation between outflow strength and stellar mass (e.g. Finlator & Davé, 2008), iii) increased metallicity of (recycled) accreted gas at higher stellar masses (e.g. Brook et al., 2014; Ma et al., 2016), iv) the redshift evolution of an empirical stellar mass limit, above which the abundance of metals begins to saturate (e.g. Zahid et al., 2013), and v) an effect of cosmic downsizing, where star formation and therefore chemical enrichment occurs preferentially in high mass galaxies at early-times (e.g. Maiolino et al., 2008; Zahid et al., 2011).

Although the MZR can be considered a tight scaling relation (σ ~0.05-0.2 dex), its scatter is generally found to be at least a factor of two larger than the uncertainties present in the metallicity estimates (Tremonti et al., 2004; Zahid et al., 2012). Even more importantly, this dispersion is strongly correlated with other galaxy properties (Cooper et al., 2008; Peeples et al., 2009; Davé et al., 2011a). These two statements are the motivation for a significant amount of theoretical and observational effort to understand the physical drivers of scatter in the MZR. The most commonly explored secondary dependency is the (anti-)correlation between metallicity and current SFR at fixed stellar mass, first observed in SDSS galaxies by Ellison et al. (2008) and since established, to varying extent, using a range of samples and methods at different redshifts (Hunt et al., 2012; Lara-López et al., 2013a; Stott et al., 2013; Cullen et al., 2014; Nakajima & Ouchi, 2014; Maier et al., 2014; de los Reyes et al., 2015; Salim et al., 2015). Since physical explanations for the origin of the MZR generally invoke a balance between enriched outflows and pristine or recycled gas inflow as an explanation for the mean relation, deviations from this equilibrium (i.e. due to star formation) can be used as a probe of these mechanisms.

Despite such studies, open questions remain as to the extent and physical nature of the metallicity-star formation rate dependency (Z-SFR). Early results by Mannucci et al. (2010, herafter M10) seemed to show an invariance of the mass-metallicity-SFR ($M_\star$-Z-SFR) relation with redshift, resulting in those authors dubbing it the "fundamental" metallicity relation, or FMR. However, more recent work suggests that, although the qualitative sense of the FMR persists beyond the local Universe, the normalisation and strength of the trend evolves with redshift (Brown et al., 2016; Ma et al., 2017). A dependence of the Z-SFR relationship on stellar mass was proposed by Yates et al. (2012), who find that star forming, low mass galaxies ($M_\star/M_\odot \leq 10^{10.2}$) are indeed metal-poor, however, above this threshold they see a reversal where systems with higher SFR are found to have higher metallicities. Although more work is needed to explain this result, Salim et al. (2014) find that selection biases in the abundance calibration used by Yates et al. (2012) are responsible. Sánchez et al. (2013) go so far as to suggest that the Z-SFR relationship is driven by the presence of observational biases in SDSS data (i.e. fibre aperture effects), although several works have for the most part ruled out this conclusion by performing their own detailed analysis of the SDSS MZR, showing that the relationship between metallicity and star formation persists even when these uncertainties are accounted for (Andrews & Martini, 2013; Salim et al., 2014; Telford et al., 2016).

In addition to the $M_\star$-Z-SFR relationship, it is reasonable to expect that there also exists an observable connection between the metallicity and gas content. Indeed, theoretical work generally supports this notion as a natural consequence of the equilibrium between inflows, outflows and gas processing (Dutton et al., 2010; Davé et al., 2011a, 2012; Lagos et al., 2016). Despite its theoretical importance, the intrinsic faintness of H I emission, poor statistics and selection biases have historically made accounting for the effect of gas on the MZR much more difficult than that of star formation and only recently have studies begun to produce observational evidence for the MZR dependency on gas. Hughes et al. (2013) use H I observations of $\sim$250 objects to show that the H I mass of metal-rich systems is typically lower than their metal-poor counterparts, attributing this to the increased efficiency of the star formation process in more massive galaxies. Using an increased sample of $\sim$4000 H I selected galaxies, Bothwell et al. (2013) also find an anti-correlation between gas mass and metallicity at fixed stellar mass, using this to argue that the H I-MZR relationship is more fundamental than the dependence observed with star formation. Lara-López et al. (2013b) also find that galaxies with high gas fractions are metal-poor compared to their gas-poor counterparts.

Based upon the work outlined above, it appears that a physical connection between the stellar mass of a galaxy, its gas-phase metallicity, gas content and current SFR exists. However, the exact character of the relationship remains elusive and establishing whether gas content or star formation is the most important factor in this picture has so far not been possible. If we are to properly interpret theoretical predictions and develop a comprehensive understanding of chemical enrichment and star formation, it is clear that we must establish the most physically motivated dependence of the MZR. To do so, we now stack atomic gas spectra of $\sim$10,000 star forming galaxies - the largest sample for which data are available - and quantify the *relative* importance of gas content and star formation as drivers of scatter in the MZR.

This chapter is organised as follows: Section 7.2 introduces the MZR for our sample. In Section 7.3 we examine the secondary drivers of the MZR, quantifying the dependencies on gas and star formation for a variety of SFR and metallicity indicators. Section 7.4 considers potential biases that may be present in our work and how, if possible, we account for them. Finally, Section 7.5 outlines our conclusions and discusses them in the context of previous work.

7.2 The Mass-Metallicity Relation

In this section, we stack atomic gas spectra of ∼10,000 nearby galaxies along the MZR. These galaxies are selected from our parent sample to be star forming using the standard Baldwin, Phillips, & Terlevich (1981, Figure 2.4) diagram and according to the Kauffmann et al. (2003c) cut. We require that each galaxy has valid M10 and T04 metallicity estimates as well as B04, K98 and S16 SFR indicators (see Chapter 2; Mannucci et al., 2010; Tremonti et al., 2004; Brinchmann et al., 2004; Salim et al., 2016; Kennicutt, 1998a). We also only select galaxies with stellar masses $M_\star/M_\odot \leq 10^{11}$. This mass cut removes 8 galaxies while reducing the range of stellar masses in our most massive bin by 0.4 dex. These selections leave a final sample of 9720 objects with H I, metallicity and SFR information that we refer to in this thesis as Sample C.

Our justification for choosing the M10 calibration is that it has previously been used to show a significant dependence of the MZR upon SFR and H I content (M10, Bothwell et al., 2013; Salim et al., 2014). It is an important check of our analysis that we are able to qualitatively reproduce the $M_\star$-Z-SFR and, in the case of Bothwell et al. (2013), $M_\star$-Z-H I relationships found in these works. Similarly, the T04 estimate is a commonly used diagnostic of the SFR dependence of the MZR within the literature, however, the results are often in apparent conflict with those that use the M10 calibration (Yates et al., 2012; Lara-López et al., 2013a). Furthermore, the Bayesian SED fitting approach used in the estimation of T04 metallicity is completely distinct from the methodologies used to calculate M10 (and other strong line) metallicities. See Chapter 2 for description of both techniques. These two calibrations therefore characterise the uncertain nature of the MZR and its secondary dependencies, allowing for an effective comparison and estimation of bias in the context of previous work.

The black points in Figure 7.1 show two MZRs for the 9720 star forming galaxies of Sample C. In panel (a) we show the M10 metallicity calibration as a function of stellar mass (M10-MZR) while in panel (b) we use the T04 estimate (T04-MZR). The M10-MZR for our sample is well fit by a fourth order polynomial (cyan line) of the form:

$$[\log\,(\mathrm{O/H}) + 12]_{\mathrm{M10}} = 3.341950 + 0.41401x + 0.03477x^2 + 0.00113x^3 - 0.00031x^4 \quad (7.1)$$

where x is log $M_\star/M_\odot$ and the polynomial is valid over the stellar mass range $10^9 \leq M_\star/M_\odot \leq 10^{11}$. For comparison, the green dashed line denotes the polynomial given in the original M10 study. Differences in selection mean that the relation for our sample plotted in Figure 7.1a lies above the original M10 relation at low stellar mass. This is

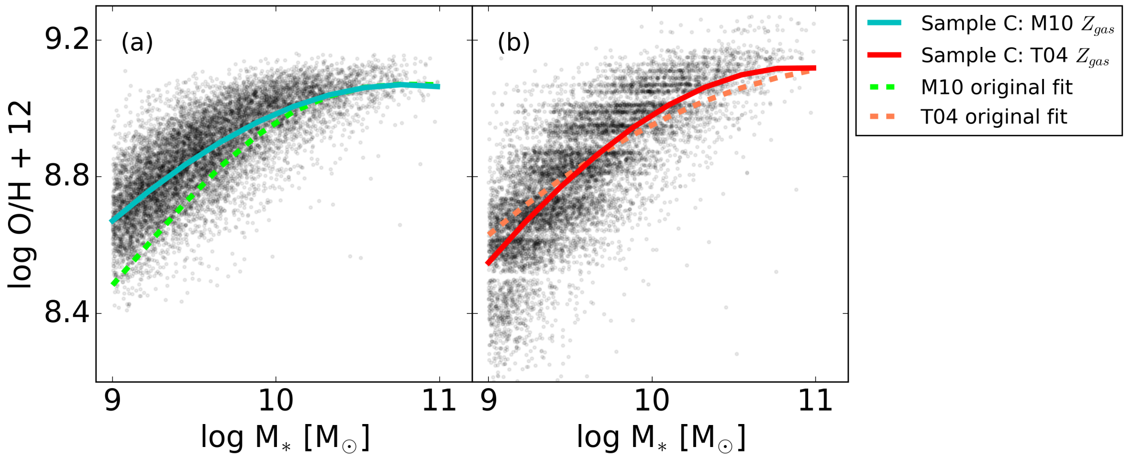

Figure 7.1: Stellar mass ($M_\star$) versus gas-phase metallicity (log (O/H) + 12) using the M10 (a) and T04 (b) calibrations. Individual galaxies in Sample C (9720 objects) are plotted as black points in both panels. The best fit to each relation is shown by the solid cyan (a) and red (b) lines. The polynomial fit quoted in the original M10 and T04 papers are shown by the green and orange dashed lines respectively.

driven by the low redshift range of our sample ($0.02 \leq z \leq 0.05$) compared to the original M10 study ($0.1 \leq z \leq 0.3$) and the subsequent aperture affects that arise from the 3 arcsec SDSS fibre. We discuss the impact of this bias on our results in Section 7.4.1. The average 1σ scatter of our M10-MZR is $\sim$0.1 dex, which matches the value quoted in the original work.

Figure 7.1b demonstrates the much better agreement between our T04-MZR (red line) and the functional form provided in the original study (dashed orange line). This is most likely because Sample C is a subset of the original T04 sample, which spans the redshift range $0.005 \leq z \leq 0.3$. As we are using their estimates, we recover identical scatter in the MZR to T04 ($\sim$0.1 dex). The horizontal banding that is apparent in the scatter of the T04 relation is due to the discrete sampling of metallicities in the grid of photoionisation models that those authors use to assign abundances (see Section 2.5.3 and T04). In the original paper, T04 smooth this effect out using a narrow Gaussian of 0.2 dex, however, for the sake of clarity we choose not to follow this approach, instead displaying the log (O/H) + 12 values obtained from the MPA-JHU catalogue. The polynomial best fit to the T04-MZR (red line) for our sample is described as:

$$[\log\ (\mathrm{O/H})\ +\ 12]_{\mathrm{T04}} = 1.94603 + 0.4594x + 0.04726x^2 + 0.00191x^3 - 0.00042x^4 \quad (7.2)$$

where x is again log $M_\star/M_\odot$ and valid over the range $10^9 \leq M_\star/M_\odot \leq 10^{11}$.

In both panels, we recover the steep correlation between stellar mass and metallicity for galaxies with stellar mass below $\sim 10^{10.5}$ $\mathrm{M}_{\odot}$. Above this stellar mass, the gradient flattens until the correlation with mass disappears. As outlined in the introduction to this chapter, this result has been noted by many observational and theoretical studies using a variety of samples and measurements of metallicity (e.g. T04 Tremonti et al., 2004; De Lucia et al., 2004; de Rossi et al., 2007; Kewley & Ellison, 2008; Ellison et al., 2008; Finlator & Davé, 2008, M10).

7.3 H I and Star Formation as Drivers of the Mass-Metallicity Relation

In this section we look to identify the strongest driver of scatter in the MZR. To do so, this work quantifies average H I mass and SFR as a function of both M10 and T04 metallicity calibrations, at fixed stellar mass.

In Figure 7.2, the black points in the background show the M10 (top) and T04 (bottom) MZR for our sample. Their respective best fits, given in Equations 7.1 and 7.2, are shown by the white dashed lines. So that we may select subsamples for stacking, we adopt two dimensional binning approach that is chosen to fully sample the dispersion in the MZR while still having sufficient galaxies to stack and detect H I . This involves dividing the sample first into 5 bins of stellar mass (limits are given in Figure 7.2 caption) and then again according to the metallicity percentiles (6.7%, 30.9%, 69.1%, 93.3%) corresponding to $\pm 0.5\sigma, \pm 1.5\sigma$ and $> 1.5\sigma$. In this way, we are able to probe the average H I mass and SFR of galaxy populations in each bin at fixed stellar mass. As the MZR changes depending on metallicity calibration used, this method allows us to bin in a consistent manner using the M10 (a, b, c) and T04 (d, e, f) calibrations. The bins are illustrated by the boxes in each plot. We discuss the effect of the size and redshift distributions of galaxies within each bin in Section 7.4.1. Note that for consistency with the stacked H I averages, we calculate the linear average of SFR in a bin and then take the logarithm.

In order to quantify star formation on a physical scale that is comparable to the H I measurement, we use the total SFRs of Sample C galaxies from the MPA-JHU catalogue that are derived via optical SED fitting. These SFRs are known to be a robust indicator of global star formation properties (Brinchmann et al., 2004; Salim et al., 2007; da Cunha et al., 2008; Salim et al., 2016) and, as such, have been used extensively within the field. On the other hand, we also want to measure the star formation over the same spatial scale as where the metallicity is estimated. We therefore follow M10 and use the Hα fibre

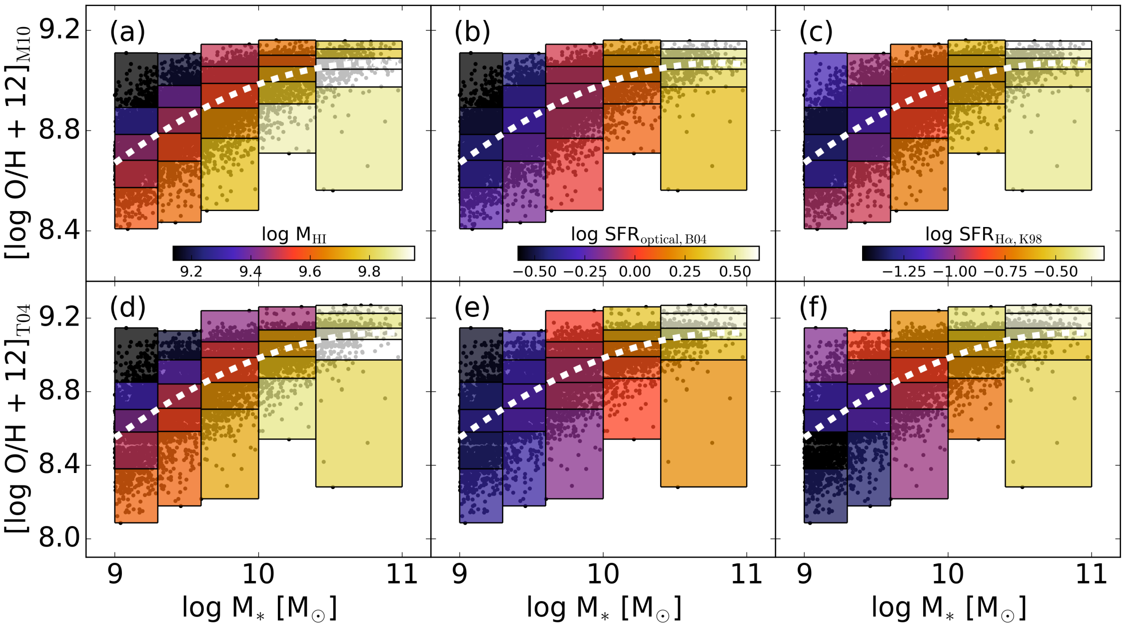

Figure 7.2: The M10 (top row) and T04 (bottom row) MZRs for galaxies in our sample. The 2D stellar mass-metallicity bins are shown by the boxes and are coloured according to their average H I mass measured in $M_\odot$ (left column), B04 total SFR (middle column) and K98 fibre SFRs (right column), both given in $M_\odot$ yr^{-1}. The bin edges along the x-axis are $\log M_\star/M_\odot = 9, 9.3, 9.6, 10, 10.4, 11$. For each stellar mass bin, we also divide log (O/H) + 12 into bins that are $\pm 0.5\sigma, \pm 1.5\sigma$ and $> 1.5\sigma$ from the median MZR. The corresponding colour bar for each column is given in the top panel. The white dotted lines are the best fit polynomials to the M10 and T04 relations (Equation 7.1 and 7.2 respectively).

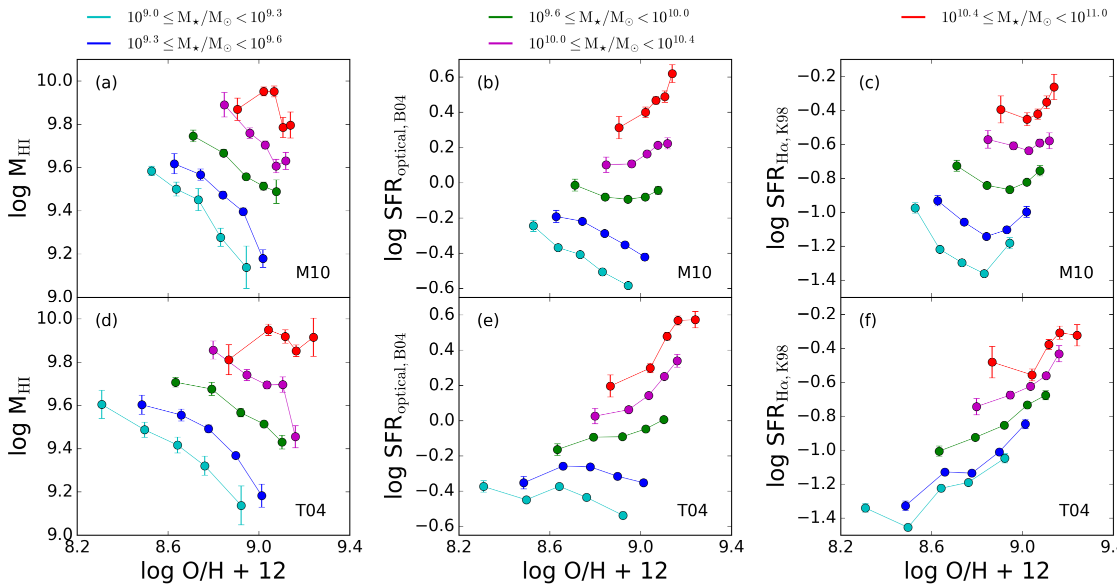

Figure 7.3: H I mass (left column), B04 total SFR (middle column) and K98 fibre SFR (right column) as a function of log (O/H) + 12 for the M10 (top row) and T04 (bottom row) metallicity calibrations. H I mass in units of $M_\odot$ and SFRs are given in $M_\odot$ yr^{-1}. Y-axis values are the average quantity in each bin of metallicity at fixed stellar mass and are identical those used to colour the corresponding panels of Figure 7.2. Stellar mass bins are given in the legend. Errors on average log M_{HI} are calculated using the Delete-a-Group Jackknife routine (DAGJK) described in Section 3.4, while the standard error on the mean is used for both SFR estimates.

SFRs calculated using the K98 prescription, a reliable indicator of star formation within the fibre aperture.

The left column of Figure 7.2 (a, d) shows the MZRs coloured according to the average H I mass of galaxies within each 2D bin. For both the M10 (a) and T04 (d) calibrations, there is a strong anti-correlation of H I mass with metallicity at fixed stellar mass, where galaxies below the MZR have, on average, up to $\sim$0.6 dex less H I mass than their high metallicity counterparts.

In the middle panels, boxes along the M10 (b) and T04 (e) relations are coloured according to their mean B04 SFR ($\mathrm{SFR_{optical,B04}}$, see Section 2.5.2). In the low mass bins ($\mathrm{M_\star/M_\odot} < 10^{10}$), the top panel demonstrates an anti-correlation between metallicity and total SFR at fixed stellar mass, where galaxies with higher SFRs lie below the M10-MZR while those with lower SFRs lie above. The trend appears reversed for the more massive galaxies, here star formers appear to be metal-rich and gas-poor. We see for the T04-MZR in panel (e), the low mass bins have a very weak correlation between metallicity and B04 total SFR, while the massive galaxies have a positive correlation.

On the right (c, f), we quantify the average K98 SFR in each box ($\mathrm{SFR_{H\alpha,K98}}$; see Section 2.5.2). In panel (c), the average fibre SFR echoes the anti-correlation with M10 metallicity exhibited by the total SFRs. However, in panel (f), when we look at the correlation between the average fibre SFR and T04 metallicities at fixed stellar mass we find a strong *positive* trend where galaxies that are star forming are also metal-rich. The relationship between metallicity and fibre SFRs is therefore contradictory for the two different abundance calibrations.

Figure 7.3 quantifies the three-way dependency between the stellar mass, metallicity and H I (or star formation) that are seen in the colour schemes of Figure 7.2 using a different projection. By plotting the average H I mass, total SFR and fibre SFR, in the same bins as before, as function of the mean metallicity in each bin we show the Z-$\mathrm{M_{HI}}$ and Z-SFR relationships at fixed stellar mass for both MZRs. As shown above, H I and metallicity are anti-correlated in each mass bin using both abundance calibrations while the unstable nature of the Z-SFR relation is clear. However, using Figures 7.2 and Figure 7.3 for a direct comparison of i) each parameter's influence on metallicity (comparing horizontally) and ii) the sensitivity of different abundance calibrations to each parameter (comparing vertically) is not a straightforward task as the MZRs and y-axis are changing between panels.

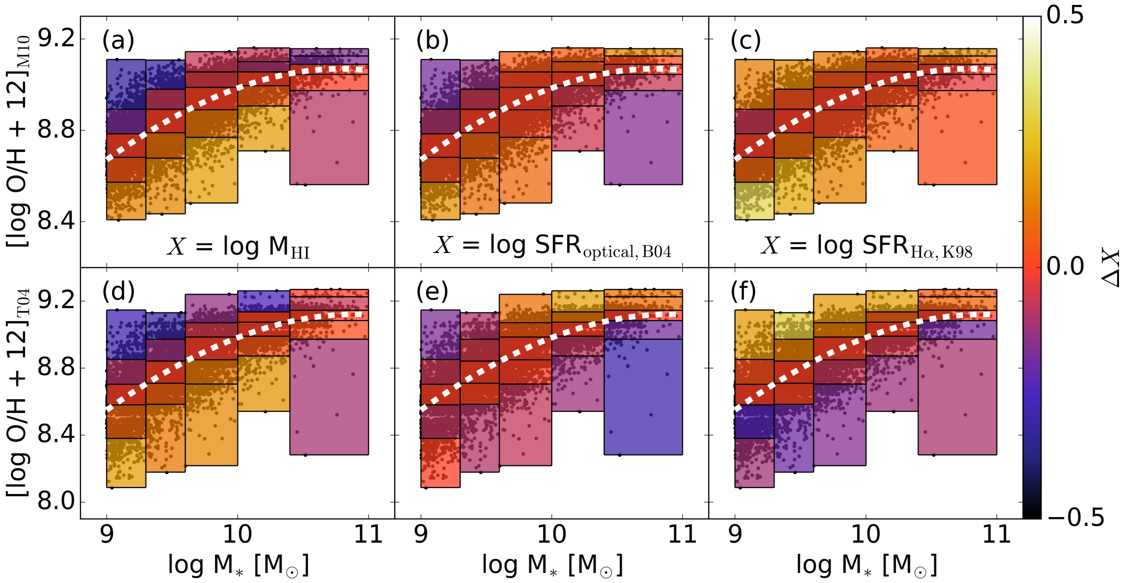

Figure 7.4: The M10 (top row) and T04 MZRs (bottom row). Each bin is coloured according to ΔX, where $\Delta X = X - X_{MZR}$ (see Equation 7.3.1) and, for each column, X is denoted in the top panel. The colour-coding is in dex and can be compared directly between panels. Black points are the individual galaxies of the M10 (top) and T04-MZRs (bottom) for Sample C and the white dashed lines are the polynomial fits to those relations. Both MZRs and their 2D bin limits are identical to Figure 7.2.

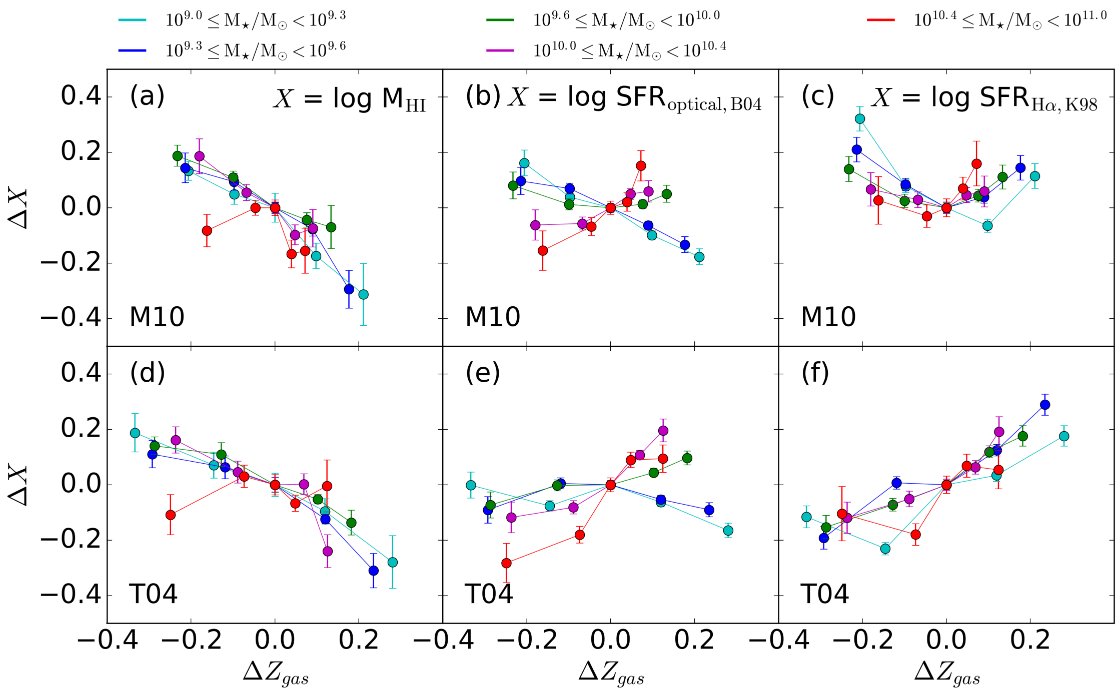

Figure 7.5: ΔX vs. ΔZ_{gas} for the M10 (top) and T04 (bottom) metallicity calibrations. Following the same pattern as Figures 7.2, 7.3 and 7.4, X = log M_{HI} in left-hand column, log $SFR_{optical,B04}$ in the middle column and log $SFR_{H\alpha,K98}$ in the right-hand column. The stellar mass bins are identical to those figures and provided in the legend. For the left column we plot DAGJK error bars while in the middle and right columns we use the standard error on the mean.

7.3.1 Identifying the Primary Driver of Scatter in the MZR

To investigate observed tensions and compare the *relative* strength of the MZR dependence upon H I mass and SFR we use Figures 7.4 and 7.5. Firstly, this allows us to quantify the balance between gas, star formation and metallicity as a function of stellar mass (Figure 7.4) and, secondly, compare the influence of H I mass, total SFR and fibre SFR in driving galaxies away from the equilibrium along the MZR (Figure 7.5).

In Figure 7.4, the top row shows the M10-MZR while the bottom row uses the T04 metallicities. This time each box is coloured by the difference of a given parameter ($\mathrm{M_{HI}}$, $\mathrm{SFR_{optical,B04}}$, $\mathrm{SFR_{H\alpha,K98}}$) from the value of that quantity on the MZR at fixed stellar mass. This relative quantity (ΔX) is therefore defined as:

$$\Delta X = \langle X \rangle - X_{MZR} \tag{7.3}$$

where $\langle X \rangle$ is the mean value of H I mass (left column), B04 total SFR (middle) or K98 fibre SFR (right) in that bin. X_{MZR} is the average value of the X quantity within $\pm 0.5\sigma$ of the MZR in the same stellar mass bin. By definition, the value of ΔX along the MZR is zero and, therefore, one can visualise these relative quantities as the offset in that property from the equilibrium population at fixed stellar mass. In other words, positive ΔX values are greater than and negative values less than the typical value of X on the MZR.

Since we are interested in the relative importance of H I mass and star formation in regulating the position of galaxies on the MZR, we suggest that ΔX is a more natural parameter for answering this question than the absolute value used in Figure 7.2. The colour-coding across Figure 7.4 is in dex, enabling direct comparison between all the panels.

At the same time, in Figure 7.5, we quantify the influence of ΔX on the mean distance from the MZR by plotting the relative H I mass ($\Delta\log\ \mathrm{M_{HI}}$) and SFR ($\Delta\log\ \mathrm{SFR_{optical,B04}}$, $\Delta\log\ \mathrm{SFR_{H\alpha,K98}}$) against relative metallicity (ΔZ_{gas}) for the M10 (top row) and T04 (bottom row) calibrations. Similarly to other relative quantities, the relative metallicity, ΔZ_{gas}, is defined as:

$$\Delta Z_{gas} = \langle \log(\mathrm{O/H}) + 12 \rangle - [\log(\mathrm{O/H}) + 12]_{\mathrm{MZR}} \tag{7.4}$$

where $\langle \log\ (\mathrm{O/H}) + 12 \rangle$ is the mean metallicity in a given box and $[\log\ (\mathrm{O/H}) + 12]_{\mathrm{MZR}}$ is the mean metallicity within $\pm 0.5\sigma$ of the MZR at fixed stellar mass. Thus, ΔZ_{gas} is the average vertical scatter above (positive) or below (negative) the MZR. The values of ΔX for each point are the same as those used to colour code the corresponding panels in Figure 7.4. Coloured lines show the $\Delta X - \Delta Z_{gas}$ relations in each of the stellar mass bins given in the legend. The relevant metallicity calibration is given in the bottom left corner of each panel.

Figure 7.4a shows the relative H I mass ($\Delta\log \mathrm{M_{HI}} = \langle\log \mathrm{M_{HI}}\rangle - \log \mathrm{M_{HI,MZR}}$) in each bin across the M10-MZR. We see that galaxies that lie vertically above the M10 relation are offset to lower gas masses than is typical for a galaxy on the M10-MZR at that mass. Correspondingly, objects that are below the M10-MZR are offset to higher H I mass. We quantify this anti-correlation seen in Figure 7.5a where galaxies that lie above the M10-MZR (positive ΔZ_{gas}) have up to $\sim$0.6 dex less H I (negative $\Delta\log \mathrm{M_{HI}}$) than their counterparts below the M10-MZR (negative ΔZ_{gas}). Importantly, this anti-correlation is almost entirely independent of stellar mass. The exception to this is at low metallicities in the highest mass bin, however, the large scatter and small number statistics in this bin mean that average H I masses measured for this regime are not deemed to be reliable. The dynamic range in metallicity is smaller at higher stellar masses but the slope remains remarkably consistent between bins. Having said this, the $\Delta\log \mathrm{M_{HI}}$-$\Delta Z_{gas}$ relationship is non-linear, with the dependence of the M10 calibration on H I steeper above the MZR than below.

Figure 7.4b presents the variations in total SFR across the same relation and, in the low mass bins ($\mathrm{M_\star/M_\odot} \leq 10^{10}$), the relationship between M10 metallicity and $\mathrm{SFR_{optical,B04}}$ agrees with the H I. Objects that are gas-poor and metal-rich are also, on average, more quiescent by $\sim$0.2 dex. However, it is only in the low mass regime that the anti-correlation of metallicity on total SFR holds true. At higher stellar mass, this trend reverses so that, on average, the more quiescent systems now lie $\sim$0.2 dex below the M10-MZR. Panel (b) in Figure 7.5 shows the relationship between relative total SFRs and MZR offset to be mass dependent. The three lowest mass bins have an anti-correlation between total SFR and M10 metallicities while the highest two mass bins exhibit a correlation.

The anti-correlation between average fibre SFR and M10 metallicity is again present in the lower stellar mass bins, however relative differences in fibre SFR between star forming and quiescent populations are smaller ($\sim$0.3 dex) than the dynamic range in H I mass (a, $\sim$0.6 dex) and total SFR (b, $\sim$0.4 dex). Figure 7.5c shows that the M10 metallicity trend for fibre SFRs differs above and below the MZR. Where ΔZ_{gas} is negative, we see an anti-correlation between star formation and metallicity, while above M10-MZR the relation is flat. At higher masses, the relationship between M10 metallicities and Hα SFR is not apparent.

The bottom row of Figure 7.4 shows the T04-MZR. In panel (d), the trend between T04 metallicities and H I content is found to be remarkably similar to the equivalent plot using M10 metallicities (a). Figure 7.5d shows that the anti-correlation between relative H I mass and T04-MZR is mass independent and also remains non-linear, being steeper above the T04-MZR than below. H I-rich galaxies that are metal-poor again have $\sim$0.6 dex more H I than gas-poor object, on average. In the low mass bins ($M_{\star}/M_{\odot} \leq 10^{10}$) of Figure 7.4e, we see a lack of dynamic range in total SFR across the T04-MZR ($\sim$0.2 dex) compared the M10-MZR ($\sim$0.4 dex, Figure 7.4b). At higher masses ($M_{\star}/M_{\odot} > 10^{10}$), there is an increase in B04 SFR as function T04 metallicity at fixed stellar mass.

Figure 7.5e shows the lack of trend between ΔZ_{gas} and total SFR for the low mass bins of the T04-MZR where the relationship is flat. A positive correlation between the two parameters is present at higher stellar masses. When examining the scatter in the T04-MZR using the fibre SFRs (7.4f) the picture changes dramatically from what is seen with the total SFRs and with M10-MZR. We find that, where in both panels (7.4c) and (7.4e) there is a weak negative correlation between metallicity and SFR, there is now a positive correlation that persists independently of stellar mass. Figure 7.5f quantifies the positive correlations between T04 metallicities and relative fibre SFR seen in Figure 7.4f. Interestingly, this relation is linear over the whole ΔZ_{gas} range and largely independent of stellar mass. This result appears inconsistent with the picture that is painted by the other metallicity-SFR relationships where there is an anti-correlation between star formation and metallicity at low stellar masses.

To summarise, comparing the variation of colour across all 6 panels in Figure 7.4 naturally leads to the conclusion that the $M_{\star}$-Z-SFR relationship is heavily reliant upon the combination of SFR indicator and metallicity calibration used as well as the stellar mass bin. On the other hand, the $M_{\star}$-Z-H I relation is stronger, relatively stable across M10 and T04 calibrations and independent of stellar mass. The quantification of this result in Figure 7.5 suggests that the H I content is a more reliable and physically motivated parameter than SFR for setting the metallicity at fixed stellar mass. In the next section we discuss if this result could be driven by observational biases present in the data.

7.4 A Discussion of Potential Biases

7.4.1 Aperture Effects

We now consider the role that the 3 arcsec aperture of the SDSS fibre plays in driving the secondary dependencies of the MZR seen in Section 7.3. It is well known that systematic gradients in the metallicity as a function of radius are present in most massive late-type galaxies with metallicities typically found to be higher in the centre than at the outskirts

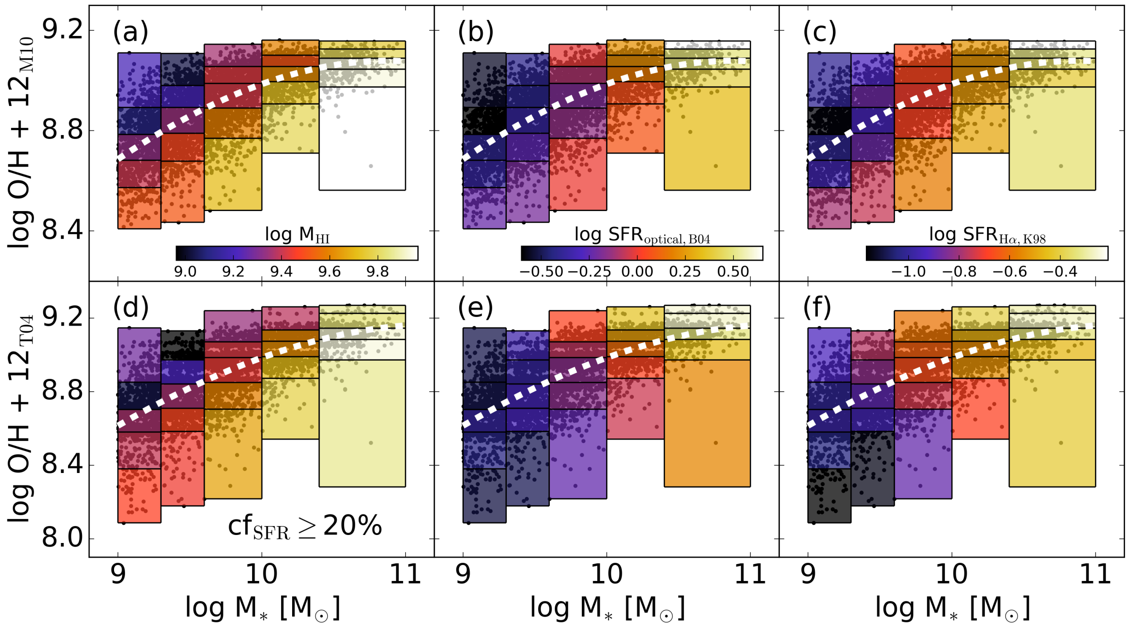

Figure 7.6: The same as Figure 7.2 using the sample of 5014 galaxies that have a star formation covering fraction (cf_{SFR}, the ratio of fibre-to-total SFR) ≥20%.

(Zaritsky et al., 1994). To avoid sampling only the central regions and therefore the abundance measurement being biased high, most studies using fibre spectroscopy tend to either set a lower redshift bound on their samples so that the fibre corresponds to a reasonable physical size (e.g. M10, Salim et al., 2014; Telford et al., 2016), or use a slightly more sophisticated cut to select galaxies for which the covering fraction of the SDSS fibre is large (Ellison et al., 2008; Kewley & Ellison, 2008). Previous work has found that a flux covering fraction (cf_{flux}, ratio of fibre-to-total flux) of 20% is suitable for recovering fibre derived abundances that agree well with estimates of global metallicity for galaxies with $M_\star/M_\odot < 10^{10}$ (Kewley et al., 2005; Kewley & Ellison, 2008). In their paper, Kewley & Ellison (2008) conclude that a covering fraction of >20% is not sufficient to avoid metallicity gradients in galaxies $M_\star/M_\odot > 10^{10}$. The 3 arcsec diameter of the SDSS fibre corresponds to between 1 and 3 kpc across the redshift range of our sample and is therefore typically smaller than the average galaxy size (the sample's mean radius is ∼3 kpc).

The effect of a redshift cut on the MZR can be seen in Figure 7.1, where our relation is systematically offset to higher metallicities than the M10 relation because their galaxies are more distant. Unfortunately, the requirement of H I data and large statistics for this work mean that such a cut is not an option for our analysis. We choose to check our results by ensuring they are not driven by fibre covering fraction. To do so, we repeat the analysis in Figure 7.6 using a reduced sample of 5014 galaxies for which the ratio of fibre-to-total SFR (cf_{SFR}) is more than 20%. We use a SFR ratio instead of a flux ratio for this criterion because only ∼30% of the sample has a cf_{flux} over 20%. Nevertheless, this cut is the same proxy used by Bothwell et al. (2013) and follows the same trends as the more conservative covering fraction estimates. In Figure 7.6 we plot the M10 and T04 relations for the ∼5000 galaxies in Sample C for which cf_{SFR} ≥20%. Encouragingly, each of the relationships between metallicity and H I (or SFR) remains consistent with what we find in Section 7.3. We use this cut for all our galaxies in order to ensure we achieve an H I detection in the high mass regime, increasing the cf_{SFR} cut higher than 20% would make this impossible.

While we cannot rule out that our results are in some part dependent on biases introduced by to metallicity gradients and the fibre aperture, particularly at larger masses ($M_\star \geq 10^{10}$ $M_\odot$), the fact that the trend persists after this cut suggests that the main conclusions of this chapter are not likely to be driven by the aperture effects. This is supported conclusion is supported in more detailed analysis of systematics within the MZR trends (Andrews & Martini, 2013; Salim et al., 2014; Telford et al., 2016).

7.4.2 Choice of Metallicity and Star Formation Rate Indicators

Given the significant practical and theoretical challenges present in the measurement of metallicities, determining the 'best' calibration is not a trivial task. As such, the different calibrations and their respective advantages remain a subject of much interest within the field. An in-depth discussion on this topic is well beyond the scope of this thesis and we therefore refer readers interested in a more thorough discourse to works of Kewley & Ellison (2008), Andrews & Martini (2013) and Salim et al. (2014), and the references contained therein. For this work, our objective is simply to ensure that key results are not driven to a significant extent by our choice(s) of metallicity calibration and it is for this reason that we select two calibrations that 'bracket' the range of disagreement found in the literature.

We choose the Kewley & Dopita (2002, hereafter KD02) metallicity calibration as a sanity check for our results. Successfully reproducing the mean MZR found in that work. However, the strict signal-to-noise emission line cuts in the KD02 method mean that there are only ∼5000 galaxies with valid KD02 metallicities within Sample C. Unfortunately, these statistics are not sufficient for recovering the stacked H I mass across the scatter of the MZR. We note that for it is very *difficult to imagine a scenario where either the choice of abundance calibration and/or aperture effects can result in a trend between gas and* *metallicity that is both strong and stable.* Observations searching for the dependence of the MZR on star formation have so far yielded vastly different results (e.g. M10, Yates et al., 2012; Sánchez et al., 2013). Considering this uncertainty, it is important to recognise that this is not the case for studies of the $M_{\star}$-Z-M_{HI} relationship. Using significantly different samples, each group investigating this dependence has found a qualitatively similar picture - gas content is anti-correlated with metallicity (Hughes et al., 2013; Bothwell et al., 2013; Lara-López et al., 2013b).

To check that the choice of B04 star formation rates is not driving some of the trends we find, we repeat the analysis in Figures 7.2 to 7.5 using the global star formation rates from the GALEX-SDSS-WISE Legacy Catalog (GSWLC; Salim et al., 2016, hereafter, S16). Since the dependency of the MZR on the S16 SFRs is very similar to the trend with the B04 SFRs we do not present the comparison figure in this chapter. Instead the comparison is shown in Figure B.1 of Appendix B. The similarity between the MZRs dependence on the two total SFR indicators is not surprising given the strong correlation between these two estimates for star forming galaxies (see Figure 2.6).

7.5 Discussion and Conclusions

In this chapter, we applied H I spectral stacking to a sample of 9,720 galaxies with available metallicity, star formation and H I information (Sample C) in order to determine whether atomic gas or star formation is the more physically motivated driver of scatter in the mass-metallicity relation.

The key conclusions can be can be summarised as follows:

i) We confirm that there is an anti-correlation between the atomic gas content and gas-phase metallicity of galaxies at fixed stellar mass as previously found by Hughes et al. (2013), Bothwell et al. (2013) and Lara-López et al. (2013b).

ii) The relationship between metallicity and gas content is consistent across the M10 and T04 MZRs as well as being largely independent of stellar mass. On the other hand, the dependency of metallicity on SFR is heavily reliant upon the choice of abundance calibration and star formation indicators used, and dependent upon the stellar mass of the system.

iii) Departures from the mean MZR are more strongly correlated with H I mass than either total or fibre SFR for both M10 and T04 abundance calibrations.

The analysis in Section 7.3 confirms that, at fixed stellar mass, increases in gas-phase metallicity above the equilibrium MZR are correlated with decreases in H I mass and, vice versa, deviations below the MZR depend on increased gas content. Furthermore, we show the anti-correlation between gas and metals is significantly stronger than the dependency of metallicity on either total or fibre SFR. The observed $M_{\star}$-Z-H I relationship also remains remarkably stable when using the M10 and T04 abundance calibrations while the character of the $M_{\star}$-Z-SFR relation changes depending on the metallicity and SFR indicators used. Discrepancies in the relationship between star formation and metals for the M10 and T04 calibrations have been reported by other studies (e.g. Yates et al., 2012; Salim et al., 2014), however, this is the first work to demonstrate the reliability of gas content in setting the gas-phase oxygen abundance using both calibrations. Following this, we suggest that it is gas content, not star formation, that should be considered the *de facto* third parameter of the MZR.

The anti-correlation between metallicity and H I content at fixed stellar mass found in our results can be explained by invoking changes in the rate at which gas is accreted. A rise in gas supply naturally leads to an increase in gas mass which, in turn, dilutes the metal abundance and boosts star formation. This is the situation we observe *below* the MZR where we find galaxies to be more gas-rich and star forming. In the opposite scenario, a slowdown in infall rate means that processed gas is not replenished while ongoing star formation is simultaneously enriching the ISM. This acts to drive galaxies *above* the MZR where, in the observations, we find them to be relatively gas-poor and quiescent. The fact that the $M_{\star}$-Z-H I relationship also appears to be insensitive to stellar mass, suggests that this process drives departures from the MZR effectively across the stellar mass range of our sample. The reduction in scatter as a function stellar mass is consistent with a scenario where massive, gas-poor galaxies return to equilibrium faster (Lagos et al., 2016). While we do not rule out the influence of outflows in the dispersion of the MZR, we note that it is difficult for this mechanism to produce the observed metallicity dependency on H I content. Outflows acting to decrease metallicity would also eject gas from the galaxy, making the observed trend of metal-poor galaxies being the most gas rich unlikely. Thus, a more physically motivated picture is one where fluctuations in the rate of gas accretion are the primary driver of scatter in the MZR at fixed stellar mass. The outflow rate, on the other hand, is strongly coupled to stellar mass of a galaxy and therefore it is possible this dictates the general form of the equilibrium MZR relation.

This scenario is supported by a number of theoretical efforts, all of which are able to qualitatively reproduce our observed dependence of metallicity upon gas and star formation. Davé et al. (2012) and Lilly et al. (2013) used equilibrium models to connect inflow, outflow and star formation to galaxy metallicities. In both models, the deviations from the equilibrium at fixed stellar mass are primarily governed by the rate of gas accretion while the MZR shape is governed by outflows. The analytic frameworks of Dayal et al. (2013) and Forbes et al. (2014) do not explicitly assume equilibrium yet their models also attribute the scatter in the MZR to fluctuations in accretion. It should be noted that, in this case, the word "accretion" does not distinguish between the mechanisms of gas inflow along large-scale cosmic filaments (e.g. Kereš et al., 2005; Dekel et al., 2009) and gas that is condensing onto the disk via the galactic fountain (e.g. Shapiro & Field, 1976; Joung & Mac Low, 2006). Lagos et al. (2016) used the EAGLE suite of cosmological hydrodynamical simulations (Schaye et al., 2015) to investigate the metallicity of galaxies along the $M_\star$-SFR-gas fraction plane. The use of EAGLE means that Lagos et al. (2016) also relax the requirement for equilibrium as well as relying on fewer assumptions about the processing of gas than analytic models. Using principle component analysis they determined that galaxy metallicity is most strongly correlated with gas fraction and, interestingly, recover an equilibrium as an output, suggesting that galaxies self-regulate along this plane. This supports the notion that, even in where no equilibrium is assumed, once models are tuned to the observations galaxies naturally tend toward this balance.

Lastly, we explore the non-linear nature of the ΔH I-ΔZ_{gas} relationship shown in Figures 7.5a and 7.5b. This result is intriguing because, if taken at face value, it suggests that regulation of metal content by H I mass is more effective below the MZR than above. The general sense of this statement is that the increase of gas supply above the equilibrium rate is more efficient at diluting metal content than a corresponding slowdown is at increasing metal concentration. The detailed modelling required to properly interpret such a result is considerably beyond the scope of this chapter and will likely be the focus of future work. Here we simply note that the expected timeframe on which the ISM reacts to changes in the infall rate qualitatively supports this picture. In driving galaxies below the mean MZR, the process of increasing gas supply, the subsequent decrease in metallicity and boost in star formation can be expected to occur on a timescale that is of order a dynamical time ($\sim$100 Myr; Mo et al., 1998; Davé et al., 2011b). Above the MZR, the slowdown of inflow, consumption (or expulsion) of gas reservoirs and ongoing chemical enrichment can only occur on a timescale that is equivalent to M_{HI}/SFR, or depletion time ($\sim$1-10 Gyr; Daddi et al., 2008). Naively, this scenario would be expected to produce a skewed distribution of metallicities about the mean MZR and, indeed, this is what we find in observations, where there is a tail in the log (O/H) + 12 distribution off to lower values at fixed stellar mass.

An alternative possibility is that the change in ΔH I-ΔZ_{gas} slope is due to the increasing contribution of H_2 to the total gas mass as function of metallicity. It is straightforward to show that a correction of the gradient above the MZR to match the gradient below requires an *increase* in H_2 mass equivalent to $\sim$40% of the H I mass. This scenario seems plausible given the role of metallicity in the conversion of atomic to molecular hydrogen (McKee & Krumholz, 2010), however, there are two significant caveats. Firstly, such a large increase in the H_2 mass at fixed stellar mass does not appear to be typical for star forming galaxies (e.g. Saintonge et al., 2016) and, secondly, the fact that we see more quiescent systems residing above the MZR goes against the idea that H_2 increases in this regime. For these reasons, we find this explanation unlikely although further observations are required in order to test it fully.

To summarise, we stack H I spectra in bins across the MZR, improving statistics and depth beyond what has previously been possible and investigating the balance between gas, metallicity and star formation as a function of stellar mass. The results presented can be understood within the context of a model in which galaxies tend to grow in an equilibrium between gas content, metallicity and star formation. The fact that deviations from this equilibrium are most strongly correlated with gas mass suggests that the scatter in the mass-metallicity relation is primarily driven by the fluctuations in the rate of gas supply, and the secondary dependence upon SFR arises because a source of gas is required to sustain star formation. The results presented provide new and important observational constraints for theoretical models of galaxy evolution.

8

Summary and Future Work

8.1 Overview

Despite galaxies in the local Universe exhibiting a huge variety of properties across many different environments, a global picture has emerged where the bulk of the local galaxy population falls into one of two categories: either they are gas-rich and star forming, or else they are gas-poor and quiescent. Understanding how the atomic gas regulates this picture via the formation of stars and build up of stellar mass as well as the role of the environment in quenching star formation is vital to disentangling the astrophysical mechanisms that drive the evolution of galaxies.

This thesis has gained new insights into these secular and external processes by quantifying the role of atomic hydrogen in regulating the stellar mass, star formation activity and metallicity of galaxies across different environments, and providing important constraints for theoretical models of galaxy formation and evolution.

The analysis presented has addressed three key questions of galaxy evolution; i) how does gas content regulate the star formation process and build up of stellar mass in galaxies?, ii) where and how precisely does external influence on H I reservoirs become important? and iii) how does the cycling of gas into and out of galaxies affect the star formation cycle?

To answer these questions, we applied the H I spectral stacking technique to subsets of a volume-limited, stellar mass-selected parent sample comprising over 30,000 galaxies and spanning the entire environment regime from the field to clusters. This large multi-wavelength dataset is purposefully designed to maximise scientific gains from stacking, combining the largest sample of H I, optical and UV data available with stellar mass, star formation and environment information.

One important aspect of our research is that we have used H I stacking to overcome the sensitivity limitations of current H I surveys, co-adding H I sources that are selected according to optical position and redshift regardless of their formal detection status. This thesis showcases the gains to be made by H I stacking, completing the most comprehensive study yet into the connections between gas, global galaxy properties and environment across the range of gas-rich to gas-poor regimes.

We summarise the main results of this thesis as follows:

8.2 The Links between Gas Content and Galaxy Properties

Much of our progress so far has relied upon the existence scaling relations that link the physical properties of galaxies together. The fact that there is significant galaxy-to-galaxy variation in such relations means that secondary correlations with their scatter are powerful probes for understanding the physical basis for the diversity of galaxy properties. They also provide simple, quantitative tests for the theoretical predictions of galaxy formation and evolution. Indeed, the success (or failure) of such models is to a large extent judged by their ability to reproduce a number of key scaling relations between global galaxy properties.

From an observational point of view, the gas fraction scaling relations with stellar mass, colour and morphology are well established (e.g. Catinella et al., 2010; Fabello et al., 2011a). However, the faintness of H I emission and the subsequent poor statistics have hampered investigations into the second-order trends that drive these relationships and their scatter. As such, the full extent of the interplay between the atomic gas reservoirs and stellar mass, star formation activity and morphology remain unclear.

To address this, we provided a framework of H I scaling relations deep enough to probe representative H I masses, and comprehensive enough that the independent influence of each variable on gas content is established. By stacking atomic gas spectra for a statistically-large, representative sample of $\sim$27,000 galaxies, we constrain the secular mechanisms that regulate the main gas fraction scaling relations with stellar mass, NUV-r colour and stellar surface density. This work demonstrated that the specific star formation rate (sSFR), not stellar mass, of nearby galaxies is the dominant parameter when it comes to gas content. Furthermore, we show that the scaling relations of H I fraction with stellar mass and morphology are second order effects driven by the bimodality of star forming and quiescent populations and the relationship between star formation activity and gas. This is not surprising given that the stellar mass, the product of many successive generations of ISM, is not fundamentally connected to the current mass of H I.

This finding has important ramifications for theory since it suggest that the gas fraction-stellar mass scaling relation should naturally be present in a model that successfully captures the more fundamental relationship between gas content and star formation, and the galaxy bimodality. It therefore interesting and encouraging that several cosmological models are now able to reproduce realistic H I properties of galaxies as function of mass and star formation activity (Davé et al., 2013; Lagos et al., 2014; Crain et al., 2017).

The key results of this work are:

i) We confirm the importance of NUV-r colour, a proxy for specific star formation rate (sSFR), over stellar mass when it comes to tracing the H I-to-stellar mass ratio of nearby galaxies (Cortese et al., 2011; Fabello et al., 2011a; Catinella et al., 2013). Galaxies of with similar sSFR are, on average, likely to exhibit similar H I fractions, showing only a small residual dependence on mass or morphology.

ii) The scaling relations of H I fraction with stellar mass and morphology are second order effects driven by the bimodality of star forming and quiescent galaxy populations and the relation between star formation and gas surface densities known as the Kennicutt-Schmidt law (Schmidt, 1959; Kennicutt, 1998b).

iii) The residual dependence of the gas fraction-NUV-r colour scaling relation on stellar mass and surface density is driven by increased star formation efficiency (SFE = SFR/M_{HI}) in more massive or bulge-dominated galaxies compared to low mass or disk-dominated systems.

8.3 The Impact of Environment on H I in Galaxies

Once the role of gas in driving the main scaling relations with galaxy properties was established, the obvious next step was to determine how external processes regulate the depletion of gas content as function of environment, and the direct consequences for star formation and stellar mass growth.

Historically, there is a large body of work showing the strong environmental dependencies of star formation (e.g. Balogh et al., 1999; Gómez et al., 2003) and structural properties (e.g. Dressler, 1980; Wilman & Erwin, 2012). At radio wavelengths, studies have shown that H I reservoirs are adversely affected in galaxy clusters (Giovanelli & Haynes, 1985a; Chung et al., 2009; Cortese et al., 2011) and that gas processing begins to occur within the group environment (e.g. Kilborn et al., 2009; Hess & Wilcots, 2013). However, there is a shortage of work investigating the impact of environment on the gas content of galaxies from a statistical perspective.

To address this gap, we selected ∼10,500 galaxies, based upon their "satellite" status within their parent dark matter halo, in order to determine the extent to provide the first statistical assessment of H I depletion as function of both galaxy properties and environment. We control for stellar mass and sSFR in order to disentangle the external processes responsible for the quenching of galaxies from the internal star formation activity and stellar mass growth. This work supports a picture where galaxy gas content is suppressed by external mechanisms across the groups regime and into the cluster. We also show that the H I loss in massive halos ($M_h/M_\odot > 10^{13}$) is considerably faster than the subsequent quenching of star formation. This result contributes to a picture whereby fast-acting stripping in groups and clusters drives gas from galaxies over very short timescales via interaction of the interstellar medium and intergalactic medium (Gunn & Gott, 1972).

In this work, we also compare our observations to theoretical predictions (Davé et al., 2013; Gonzalez-Perez et al., 2014), finding that the satellite galaxy populations in both semi-analytic models and hydrodynamical simulations are endemically gas poor. This demonstrates that the progress made by theory in understanding the gas fraction scaling relation, while promising, does not explain everything. Even in models where environment does not act directly on the cold gas reservoir (Gonzalez-Perez et al., 2014), the predicted H I content of satellites galaxies is lower than observations. Although there is qualitative agreement that the gas content of these galaxies depends upon the mass of their halo and depletion, particularly at fixed sSFR, is caused by a stripping mechanism, it is widely recognised that there is much work to do for future theoretical studies if they are to correctly reproduce environmental drivers of galaxy evolution.

Our conclusions are as follows:

i) Satellite galaxies in more massive dark matter halos have, on average, lower H I fractions at fixed stellar mass and fixed sSFR than those in less massive halos. The significant and systematic decrease in the gas content of satellites as a function of halo mass occurs across the entire group regime as well as in the cluster environment.

ii) Ram-pressure stripping is the most prevalent gas removal mechanism in halos above $\sim 10^{13}$ $M_\odot$. The lower gas fractions at fixed sSFR of galaxies in such massive halos demonstrates that the depletion of H I in this regime occurs on shorter timescales than the shut down of star formation.

iii) In support of conclusion ii), we find that the observed environmental depletion of gas content is primarily driven by hydrodynamical processes associated with the host halo mass rather than gravitational mechanisms that are related to the local density in which a galaxy resides.

iv) H I properties of satellite galaxies in the local Universe are not well reproduced by theory. There is a tendency for the models used in this thesis to produce satellite populations that are too gas poor and therefore excessively quenched. It is likely that this problem is not simply environment driven, instead being caused by one or more processes endemic to the satellite population in the theoretical models (e.g. feedback, cosmological accretion, H I-to-H_2 conversion).

8.4 Gas as a Driver of the Mass-Metallicity Relation

In the current concordance model of cosmology, the primary mechanism through which star forming galaxies of $M_\star/M_\odot < 10^{11}$ are predicted to grow is the smooth accretion of pristine and recycled gas and its subsequent conversion into stars (Nelson et al., 2013; Christensen et al., 2016; Finlator, 2016). The interplay between these inflows and the processing of gas via star formation and outflows leads to tightly coupled relations between stellar mass, gas-phase metallicity, star formation and gas content. Theoretical treatments have shown that the observed mass-metallicity relation (MZR) and its secondary dependence on SFR (Ellison et al., 2008; Mannucci et al., 2010; Salim et al., 2014) arise naturally due to an equilibrium where star formation and outflows are balanced against inflow and dilution is offset by enrichment (Davé et al., 2012; Forbes et al., 2014). The idea that the scatter in the MZR is driven by departures from this equilibrium has led to a significant amount of recent work, both theoretical and observational, aimed at understanding the processes that regulate the MZR.

The intrinsic faintness of H I emission, poor statistics and selection bias mean accounting for the effect of gas on the MZR has been much more difficult than that of star formation. However, recent observations have begun to provide evidence for the secondary dependence of the MZR on H I gas (Hughes et al., 2013; Bothwell et al., 2013; Lara-López et al., 2013a). To properly explore this scenario, we selected ∼10,000 galaxies for which SDSS spectroscopy and (total and fibre) SFR indicators are available and quantified both H I content and star formation along the MZR. By stacking in bins that maximise coverage across this parameter space, we are able to improve statistics and depth beyond what has previously been possible, investigating the balance between gas, metallicity and star formation as a function of stellar mass.

We find the anti-correlation of metallicity on gas content to be significantly stronger than the dependence on both total and fibre SFR as well as independent of stellar mass and the abundance calibration used. These trends are consistent with the model of galaxy evolution in which galaxies tend to exist in an evolving equilibrium between gas inflow/outflow, star formation and metal production. The fact that deviations from equilibrium are strongly anti-correlated with gas mass suggests that the scatter in the MZR is primarily driven by the fluctuations in the rate of gas inflow while enriched gas outflows determine the slope and normalisation of the MZR. The secondary dependence upon star formation rate arises simply because a source of gas is required to sustain star formation.

There is support for this scenario from the literature. Models that explicitly assume equilibrium have been used to infer changes in the rate of inflow across the MZR (Davé et al., 2011a, 2012; Lilly et al., 2013). Perhaps more significantly, analytic models and cosmological hydrodynamical simulations have recently been shown to recover the MZR and its gas content dependency self-consistently once parameters are tuned to produce sufficiently realistic galaxy populations (Forbes et al., 2014; Lagos et al., 2016).

Our results are in agreement with the notion that, as outlined above, the star forming galaxies grow in a quasi-steady state that is regulated by the smooth accretion of gas and reflected in the continual rebalancing between star formation, metallicity and gas content on relatively short timescales. Our results can be summarised as:

i) We confirm that the scatter in the mass-metallicity relation is anti-correlated with H I mass, as found by Bothwell et al. (2013), Hughes et al. (2013) and Lara-López et al. (2013a).

ii) H I mass is a more physically motivated second parameter for the mass-metallicity relation than star formation. The mass-metallicity relation has a stronger dependence on H I mass than either total or fibre star formation rate (SFR). Furthermore, the metallicity-gas mass relationship is independent of stellar mass while the trends between metallicity and star formation rate vary across the mass range.

iii) The choice of metallicity calibration and SFR indicator affect the strength and direction of the secondary dependence of the mass-metallicity relation on star formation, however, the secondary dependence upon gas content is remarkably robust to the choice of metallicity calibration.

8.5 Final Conclusion

The aim of this thesis was to use statistically powerful samples to disentangle the physical processes regulating when, where and how star formation is triggered or quenched, and how the secular and external mechanisms at play combine to shape the local galaxy population.

We presented new insights into the relationships between gas content, galaxy properties and environment in the local Universe, providing strong constraints for galaxy formation and evolution models. This work also demonstrated the importance of H I stacking as a versatile tool for statistical studies of H I content.

We employed well known galaxy scaling relations to establish that it is sSFR, not stellar mass, that is the most important quantity relating to the gas content. The large scatter that is present in the gas fraction-stellar mass and morphology relationships is driven by the bimodality in star formation activity and the strong connection between star formation and gas content. Externally, the ongoing influence of dark matter halos on gas reservoirs is shown to be a critical mechanism in driving galaxies from the blue cloud to the red sequence. Further more, environmentally driven gas depletion occurs across the group regime, well before galaxies enter the cluster. We demonstrated that, while a small residual dependence remains, gas content is not strongly dependent on the local density of objects, downplaying the importance of gravitational interactions in systematically affecting the star formation cycle. Lastly, we showed that H I mass is the primary driver of scatter in the relationship between galaxy stellar mass and gas-phase metallicity. This conclusion supports a scenario where changes in the rate of gas inflow continuously regulate the gas content, star formation and chemical evolution of galaxies.

In conclusion, this thesis begins to disentangle the complex relationships between galaxy gas content, star formation, metallicity and environment. In doing so, we have provided new and important constraints for models of galaxy evolution and shed light on the internal and external processes that combine to regulate the life cycle of nearby galaxies.

8.6 Future Work

8.6.1 The Star Formation Cycle across Cosmic Time

The star formation rate "main sequence" (SFR vs. stellar mass) shows a smooth evolution over cosmic time from $z \sim 2$ to $z \sim 0$, while over the same period, the star formation rate density of the Universe has dropped by more than an order of magnitude from its peak at $z \sim 2$ (Madau et al., 1998; Hopkins & Beacom, 2006). This evidence implies the systematic quenching of galaxies as function of cosmic time, however, the question *why do galaxies stop forming stars?* has not yet been answered.

Studies also show star formation and environment to be inextricably linked as a function of redshift (Balogh et al., 1999; Cooper et al., 2008; Poggianti et al., 2008), however, radio observations of the H I suppression in galaxies as they move to denser environments are primarily restricted to the local Universe (Giovanelli & Haynes, 1989; Chung et al., 2009; Cortese et al., 2011). This means that the interaction of gas, stars and environment in the galaxy-building process is poorly understood beyond $z \sim 0.1$, driving a need for inventories of more distant H I.

There are several ongoing or recently completed surveys designed to provide inventories of H I at higher redshift. The first two, the Blind, Ultra-Deep H I Environmental Survey (BUDHIES; Verheijen et al., 2007) and the HIGHz survey (Catinella & Cortese, 2015), provide a glimpse into the H I properties of galaxies at $z \sim 0.2$. More broadly, the COSMOS H I Large Extragalactic Survey (CHILES) is designed to be the first survey to probe H I content, galaxy properties and environment of galaxies as a function of redshift out $z = 0.5$.

In addition to programs underway at existing facilities, the pathfinders for the Square Kilometre Array (SKA) will soon survey the H I content of galaxies in the local Universe and beyond with unprecedented speed and sensitivity. As part of this advance, the Australian SKA Pathfinder (ASKAP) in Western Australia will embark upon two major H I emission surveys in the coming years: i) WALLABY, is a shallow all-sky census of H I mass for an unprecedented 6×10^5 galaxies out to $z \sim 0.26$; ii) DINGO, is a deep, small-area survey probing the H I properties of 10^5 galaxies as a function of cosmic time ($z < 0.43$).

H I stacking will be an invaluable tool for maximising scientific gains from these surveys. Although source confusion will always remain a limiting factor on studies of distant H I, high resolution observations H I surveys such as CHILES, WALLABY and DINGO mean that stacking is embarking on an era of remarkable discovery potential where it will yield average H I measurements for large numbers of galaxies at much higher redshifts than ever before.

Thus, the natural next step for this work would be to apply our methodology to disentangle the drivers of the star formation cycle, both as function of environment *and* of cosmic time.

8.6.2 Resolving the Star Formation Cycle

Recently, optical integral field spectroscopy (IFS) surveys have begun to reveal the mechanisms that drive the star formation cycle (e.g. gas accretion, feedback, chemical evolution) in extraordinary detail. For example, the SAMI Survey (Croom et al., 2012) is mapping the two dimensional kinematic, structural and chemical properties of stars and ionised gas for 3000 galaxies out to $z \sim 0.1$.

From the H I perspective, resolved maps of gas distributions are limited to only the closest systems and a small number ($<10^3$) of mostly gas-rich galaxies (e.g. Broeils & Rhee, 1997; Swaters et al., 2002; Walter et al., 2008; Hunter et al., 2012). As such, our understanding of H I in this picture has so far been driven by pairing single-dish H I information for $\sim10^4$ local galaxies (e.g. HIPASS, ALFALFA; Meyer et al., 2004; Giovanelli et al., 2005) with data from optical surveys (e.g. SDSS, GAMA; York et al., 2000; Driver et al., 2011). Fortunately, the WALLABY survey expects to resolve ∼5000 galaxies and will be well matched by the HECTOR survey, a planned IFS study that builds upon SAMI to exploit synergies with ASKAP H I surveys.

Along with other works, this thesis has demonstrated the importance of large groups and cluster environments in shutting down star formation via strong gas stripping. However, how the various phases of the gas (ionised, atomic, molecular) are stripped, the radii at which the truncation occurs and how it depends on galaxy and environmental properties remain open questions.

A study combining the ASKAP H I data with HECTOR IFS maps would be in a unique position to address this problem directly. By characterising the gas content of cluster members with respect to the spatial distribution of star formation, such a study would identify the "smoking gun" examples of ram pressure stripping. ASKAP's 30 arcsec beam makes it one of the few instruments capable of isolating H I reservoirs of cluster members at significant distance. The addition of high resolution CO observations to this investigation in order to trace the molecular gas component would create a sample of exceptional quality for probing the influence of environment on the star formation cycle of galaxies.

In summary, further study of the neutral atomic hydrogen in galaxies across different environments and as a function of cosmic time is required to improve our understanding of galaxy growth and evolution, but its broad role as a fundamental driver of the galaxy life cycle is now beyond dispute. Since the earliest days of radio astronomy, measurements of gas content have driven the advancement of this field and there is no doubt they will continue to do so. The future is bright.

Bibliography

Abadi, M. G., Moore, B., & Bower, R. G. 1999, MNRAS, 308, 947

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., & et al. 2009, ApJS, 182, 543

Abell, G. O. 1958, ApJS, 3, 211

Agertz, O., Moore, B., Stadel, J., et al. 2007, MNRAS, 380, 963

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140

Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, MNRAS, 373, 469

Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54

Balogh, M. L., Navarro, J. F., & Morris, S. L. 2000, ApJ, 540, 113

Balogh, M. L., Schade, D., Morris, S. L., et al. 1998, ApJ, 504, L75

Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57

Bekki, K. 2009, MNRAS, 399, 2221

Bianchi, L., Conti, A., & Shiao, B. 2014, Advances in Space Research, 53, 900

Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846

Blanton, M. R. 2006, ApJ, 648, 268

Blanton, M. R., Eisenstein, D., Hogg, D. W., Schlegel, D. J., & Brinkmann, J. 2005, ApJ, 629, 143

Blitz, L., & Rosolowsky, E. 2006, ApJ, 650, 933

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

Böhringer, H., Voges, W., Huchra, J. P., et al. 2000, ApJS, 129, 435

Böhringer, H., Schuecker, P., Guzzo, L., et al. 2004, A&A, 425, 367

Boselli, A., Boissier, S., Cortese, L., et al. 2006, ApJ, 651, 811

Boselli, A., Eales, S., Cortese, L., et al. 2010, PASP, 122, 261

Boselli, A., & Gavazzi, G. 2006, PASP, 118, 517

Bothun, G. D. 1984, ApJ, 277, 532

Bothwell, M. S., Maiolino, R., Kennicutt, R., et al. 2013, MNRAS, 433, 1425

Bravo-Alfaro, H., Cayatte, V., van Gorkom, J. H., & Balkowski, C. 2000, AJ, 119, 580

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151

Broeils, A. H., & Rhee, M.-H. 1997, A&A, 324, 877

Brook, C. B., Stinson, G., Gibson, B. K., et al. 2014, MNRAS, 443, 3809

Brown, J. S., Martini, P., & Andrews, B. H. 2016, MNRAS, 458, 1529

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

Burbidge, E. M., & Burbidge, G. R. 1961, AJ, 66, 541

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Catinella, B., & Cortese, L. 2015, MNRAS, 446, 3526

Catinella, B., Schiminovich, D., Kauffmann, G., et al. 2010, MNRAS, 403, 683

Catinella, B., Schiminovich, D., Cortese, L., et al. 2013, MNRAS, 436, 34

Chabrier, G. 2003, PASP, 115, 763

Chandrasekhar, S. 1943, ApJ, 97, 255

Chengalur, J. N., Braun, R., & Wieringa, M. 2001, A&A, 372, 768

Christensen, C. R., Davé, R., Governato, F., et al. 2016, ApJ, 824, 57

Chung, A., van Gorkom, J. H., Kenney, J. D. P., Crowl, H., & Vollmer, B. 2009, AJ, 138, 1741

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

Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039

Colpi, M., Mayer, L., & Governato, F. 1999, ApJ, 525, 720

Conroy, C. 2013, ARA&A, 51, 393

Conroy, C., Schiminovich, D., & Blanton, M. R. 2010, ApJ, 718, 184

Cooper, M. C., Tremonti, C. A., Newman, J. A., & Zabludoff, A. I. 2008, MNRAS, 390, 245

Cortese, L., Catinella, B., Boissier, S., Boselli, A., & Heinis, S. 2011, MNRAS, 415, 1797

Cortese, L., Davies, J. I., Pohlen, M., et al. 2010, A&A, 518, L49

Cowie, L. L., & McKee, C. F. 1977, ApJ, 211, 135

Crain, R. A., Schaye, J., & Bower, R. G. e. a. 2015, MNRAS, 450, 1937

Crain, R. A., Bahé, Y. M., Lagos, C. d. P., et al. 2017, MNRAS, 464, 4204

Croom, S. M., Lawrence, J. S., Bland-Hawthorn, J., et al. 2012, MNRAS, 421, 872

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Cullen, F., Cirasuolo, M., McLure, R. J., Dunlop, J. S., & Bowler, R. A. A. 2014, MNRAS, 440, 2300

da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595

Daddi, E., Dannerbauer, H., Elbaz, D., et al. 2008, ApJ, 673, L21

Davé, R., Finlator, K., & Oppenheimer, B. D. 2011a, MNRAS, 416, 1354

Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98

Davé, R., Katz, N., Oppenheimer, B. D., Kollmeier, J. A., & Weinberg, D. H. 2013, MNRAS, 434, 2645

Davé, R., Oppenheimer, B. D., & Finlator, K. 2011b, MNRAS, 415, 11

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

Dayal, P., Ferrara, A., & Dunlop, J. S. 2013, MNRAS, 430, 2891 de los Reyes, M. A., Ly, C., Lee, J. C., et al. 2015, AJ, 149, 79

De Lucia, G., Kauffmann, G., & White, S. D. M. 2004, MNRAS, 349, 1101

de Rossi, M. E., Tissera, P. B., & Scannapieco, C. 2007, MNRAS, 374, 323

de Vaucouleurs, G. 1975, Nearby Groups of Galaxies, ed. A. Sandage, M. Sandage, & J. Kristian (the University of Chicago Press), 557

Dekel, A., Birnboim, Y., Engel, G., et al. 2009, Nature, 457, 451

Delhaize, J., Meyer, M. J., Staveley-Smith, L., & Boyle, B. J. 2013, MNRAS, 433, 1398

Dressler, A. 1980, ApJ, 236, 351

Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971

Duarte, M., & Mamon, G. A. 2015, MNRAS, 453, 3848

Duffy, A. R., Kay, S. T., Battye, R. A., et al. 2012, MNRAS, 420, 2799

Dutton, A. A., van den Bosch, F. C., & Dekel, A. 2010, MNRAS, 405, 1690

Ebeling, H., Edge, A. C., Allen, S. W., et al. 2000, MNRAS, 318, 333

Ebeling, H., Edge, A. C., Bohringer, H., et al. 1998, MNRAS, 301, 881

Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, ApJ, 672, L107

Fabello, S., Catinella, B., Giovanelli, R., et al. 2011a, MNRAS, 411, 993

Fabello, S., Kauffmann, G., Catinella, B., et al. 2012, MNRAS, 427, 2841

Fabello, S., Kauffmann, G., Catinella, B., et al. 2011b, MNRAS, 416, 1739

Finlator, K. 2016, ArXiv e-prints

Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181

Font, A. S., Bower, R. G., McCarthy, I. G., et al. 2008, MNRAS, 389, 1619

Forbes, J. C., Krumholz, M. R., Burkert, A., & Dekel, A. 2014, MNRAS, 438, 1552

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Garnett, D. R. 2002, ApJ, 581, 1019

Gavazzi, G., O'Neil, K., Boselli, A., & van Driel, W. 2006, A&A, 449, 929

Gereb, K., Morganti, R., & Oosterloo, T. A. 2014, A&A, 569, A35

Gereb, K., Morganti, R., Oosterloo, T. A., Guglielmino, G., & Prandoni, I. 2013, A&A, 558, A54

Gereb, K., Morganti, R., Oosterloo, T. A., Hoppmann, L., & Staveley-Smith, L. 2015, A&A, 580, A43

Giovanelli, R., & Haynes, M. P. 1983, AJ, 88, 881

Giovanelli, R., & Haynes, M. P. 1985a, AJ, 90, 2445

Giovanelli, R., & Haynes, M. P. 1985b, ApJ, 292, 404

Giovanelli, R., & Haynes, M. P. 1989, AJ, 97, 633

Giovanelli, R., Haynes, M. P., & Chincarini, G. L. 1981, ApJ, 247, 383

Giovanelli, R., Haynes, M. P., & Chincarini, G. L. 1982, ApJ, 262, 442

Giovanelli, R., Haynes, M. P., Herter, T., et al. 1997, AJ, 113, 53

Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, AJ, 130, 2598

Gómez, P. L., Nichol, R. C., Miller, C. J., et al. 2003, ApJ, 584, 210

Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., et al. 2014, MNRAS, 439, 264

Grützbauch, R., Conselice, C. J., Varela, J., et al. 2011, MNRAS, 411, 929

Gunn, J. E., & Gott, III, J. R. 1972, ApJ, 176, 1

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040

Guo, Q., Gonzalez-Perez, V., Guo, Q., Schaller, M., & et al., F. 2015, ArXiv e-prints

Guo, Q., Gonzalez-Perez, V., Guo, Q., et al. 2016, MNRAS, 461, 3457

Haynes, M. P., & Giovanelli, R. 1986, ApJ, 306, 466

Haynes, M. P., Giovanelli, R., & Chincarini, G. L. 1984, ARA&A, 22, 445

Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, AJ, 142, 170

Helfer, T. T., & Blitz, L. 1993, ApJ, 419, 86

Hess, K. M., & Wilcots, E. M. 2013, AJ, 146, 124

Hester, J. A. 2006, ApJ, 647, 910

Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, AJ, 122, 2129

Hogg, D. W., Blanton, M. R., Brinchmann, J., et al. 2004, ApJ, 601, L29

Holmberg, E. 1941, ApJ, 94, 385

Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142

Hopkins, P. F. 2015, MNRAS, 450, 53

Huang, M.-L., & Kauffmann, G. 2014, ArXiv e-prints

Huang, S., Haynes, M. P., Giovanelli, R., & Brinchmann, J. 2012, ApJ, 756, 113

Huang, S., Haynes, M. P., Giovanelli, R., et al. 2014, ApJ, 793, 40

Hughes, T. M., Cortese, L., Boselli, A., Gavazzi, G., & Davies, J. I. 2013, A&A, 550, A115

Hunt, L., Magrini, L., Galli, D., et al. 2012, MNRAS, 427, 906

Hunter, D. A., Ficut-Vicas, D., Ashley, T., et al. 2012, AJ, 144, 134

Jones, M. G., Papastergis, E., Haynes, M. P., & Giovanelli, R. 2015, MNRAS, 449, 1856

Joung, M. K. R., & Mac Low, M.-M. 2006, ApJ, 653, 1266

Kanekar, N., Sethi, S., & Dwarakanath, K. S. 2016, ApJ, 818, L28

Kannappan, S. J. 2004, ApJ, 611, L89

Kauffmann, G., Heckman, T. M., De Lucia, G., et al. 2006, MNRAS, 367, 1394

Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, MNRAS, 353, 713

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003a, MNRAS, 341, 33

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003b, MNRAS, 341, 54

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003c, MNRAS, 346, 1055

Kenney, J. D. P., van Gorkom, J. H., & Vollmer, B. 2004, AJ, 127, 3361

Kennicutt, Jr., R. C. 1983, AJ, 88, 483

Kennicutt, Jr., R. C. 1998a, ARA&A, 36, 189

Kennicutt, Jr., R. C. 1998b, ApJ, 498, 541

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35

Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183

Kewley, L. J., Jansen, R. A., & Geller, M. J. 2005, PASP, 117, 227

Kilborn, V. A., Forbes, D. A., Barnes, D. G., et al. 2009, MNRAS, 400, 1962

Kimm, T., Somerville, R. S., Yi, S. K., et al. 2009, MNRAS, 394, 1131

Komatsu, E., Smith, K. M., Dunkley, J., & Bennett, C. L. e. a. 2011, ApJS, 192, 18

Köppen, J., & Edmunds, M. G. 1999, MNRAS, 306, 317

Kott, P. 2001, Journal of Official Statistics, 17, 521

Kreckel, K., Platen, E., Aragón-Calvo, M. A., et al. 2012, AJ, 144, 16

Lagos, C. D. P., Baugh, C. M., Lacey, C. G., & et al. 2011b, MNRAS, 418, 1649

Lagos, C. d. P., Davis, T. A., Lacey, C. G., et al. 2014, MNRAS, 443, 1002

Lagos, C. d. P., Theuns, T., Schaye, J., et al. 2016, MNRAS, 459, 2632

Lah, P., Chengalur, J. N., Briggs, F. H., et al. 2007, MNRAS, 376, 1357

Lah, P., Pracy, M. B., Chengalur, J. N., et al. 2009, MNRAS, 399, 1447

Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53

Lara-López, M. A., Hopkins, A. M., López-Sánchez, A. R., et al. 2013a, MNRAS, 434, 451

Lara-López, M. A., Hopkins, A. M., López-Sánchez, A. R., et al. 2013b, MNRAS, 433, L35

Larson, R. B., & Tinsley, B. M. 1978, ApJ, 219, 46

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692

Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155

Leroy, A. K., Walter, F., Brinks, E., et al. 2008, AJ, 136, 2782

Leroy, A. K., Walter, F., Sandstrom, K., et al. 2013, AJ, 146, 19

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119

Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140

Ma, X., Hopkins, P. F., Feldmann, R., et al. 2017, MNRAS

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415

Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106

Maier, C., Lilly, S. J., Ziegler, B. L., et al. 2014, ApJ, 792, 3

Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463

Makino, J., & Hut, P. 1997, ApJ, 481, 83

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115

Marasco, A., Crain, R. A., Schaye, J., et al. 2016, MNRAS, 461, 2630

Marcolini, A., Brighenti, F., & D'Ercole, A. 2003, MNRAS, 345, 1329

Martin, D. C., Fanson, J., Schiminovich, D., & et al. 2005, ApJ, 619, L1

Mastropietro, C., Moore, B., Mayer, L., et al. 2005, MNRAS, 364, 607

Mayer, L., Kazantzidis, S., Mastropietro, C., & Wadsley, J. 2007, Nature, 445, 738

Mayer, L., Mastropietro, C., Wadsley, J., Stadel, J., & Moore, B. 2006, MNRAS, 369, 1021

McCarthy, I. G., Frenk, C. S., Font, A. S., et al. 2008, MNRAS, 383, 593

McKee, C. F., & Krumholz, M. R. 2010, ApJ, 709, 308

Meyer, M. J., Zwaan, M. A., Webster, R. L., et al. 2004, MNRAS, 350, 1195

Meyer, S. A., Meyer, M., Obreschkow, D., & Staveley-Smith, L. 2016, MNRAS, 455, 3136

Mihos, J. C., & Hernquist, L. 1996, ApJ, 464, 641

Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2016, MNRAS, 456, 1459

Mitchell, P. D., Lacey, C. G., Cole, S., & Baugh, C. M. 2014, MNRAS, 444, 2637

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319

Moore, B., Lake, G., & Katz, N. 1998, ApJ, 495, 139

Moore, B., Lake, G., Quinn, T., & Stadel, J. 1999, MNRAS, 304, 465

Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682

Muldrew, S. I., Croton, D. J., Skibba, R. A., et al. 2012, MNRAS, 419, 2670

Nakajima, K., & Ouchi, M. 2014, MNRAS, 442, 900

Nelson, D., Vogelsberger, M., Genel, S., et al. 2013, MNRAS, 429, 3353

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43

Nulsen, P. E. J. 1982, MNRAS, 198, 1007

Obreschkow, D., Croton, D., De Lucia, G., Khochfar, S., & Rawlings, S. 2009, ApJ, 698, 1467

Obreschkow, D., & Rawlings, S. 2009, MNRAS, 394, 1857

Oemler, Jr., A. 1974, ApJ, 194, 1

Oh, S.-H., de Blok, W. J. G., Brinks, E., Walter, F., & Kennicutt, Jr., R. C. 2011, AJ, 141, 193

Oppenheimer, B. D., & Davé, R. 2008, MNRAS, 387, 577

Oppenheimer, B. D., Davé, R., Kereš, D., et al. 2010, MNRAS, 406, 2325

Osterbrock, D. E. 1989, Astrophysics of gaseous nebulae and active galactic nuclei

Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, MNRAS, 189, 95

Papastergis, E., Giovanelli, R., Haynes, M. P., & Shankar, F. 2015, A&A, 574, A113

Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, MNRAS, 412, 1123

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

Peeples, M. S., Pogge, R. W., & Stanek, K. Z. 2009, ApJ, 695, 259

Peng, Y.-j., Lilly, S. J., Renzini, A., & Carollo, M. 2012, ApJ, 757, 4

Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59

Pier, J. R., Munn, J. A., Hindsley, R. B., et al. 2003, AJ, 125, 1559

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Poggianti, B. M., Desai, V., Finn, R., & et al. 2008, ApJ, 684, 888

Popping, G., Somerville, R. S., & Trager, S. C. 2014, MNRAS, 442, 2398

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

Power, C., Baugh, C. M., & Lacey, C. G. 2010, MNRAS, 406, 43

Rafieferantsoa, M., Davé, R., Anglés-Alcázar, D., et al. 2015, MNRAS, 453, 3980

Rasmussen, J., Bai, X.-N., Mulchaey, J. S., et al. 2012, ApJ, 747, 31

Rhee, J., Lah, P., Chengalur, J. N., Briggs, F. H., & Colless, M. 2016, MNRAS, 460, 2675

Rhee, J., Zwaan, M. A., Briggs, F. H., et al. 2013, MNRAS, 435, 2693

Roberts, M. S. 1962, AJ, 67, 437

Saintonge, A., Tacconi, L. J., Fabello, S., et al. 2012, ApJ, 758, 73

Saintonge, A., Catinella, B., Cortese, L., et al. 2016, MNRAS, 462, 1749

Salim, S., Charlot, S., Rich, R. M., & et al. 2005, ApJ, 619, L39

Salim, S., Lee, J. C., Davé, R., & Dickinson, M. 2015, ApJ, 808, 25

Salim, S., Lee, J. C., Ly, C., et al. 2014, ApJ, 797, 126

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Salim, S., Lee, J. C., Janowiecki, S., et al. 2016, ApJS, 227, 2

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

Sánchez, S. F., Rosales-Ortega, F. F., Jungwiert, B., et al. 2013, A&A, 554, A58

Schaller, M., Dalla Vecchia, C., Schaye, J., et al. 2015, MNRAS, 454, 2277

Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889

Schaye, J., Crain, R. A., & Bower, R. G. e. a. 2015, MNRAS, 446, 521

Schiminovich, D., Wyder, T. K., Martin, D. C., Johnson, B. D., & et al. 2007, ApJS, 173, 315

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidt, M. 1959, ApJ, 129, 243

Seibert, M., Wyder, T., Neill, J., et al. 2012, in American Astronomical Society Meeting Abstracts, Vol. 219, American Astronomical Society Meeting Abstracts #219, 340.01

Serra, P., Oosterloo, T., Morganti, R., & et al. 2012, MNRAS, 422, 1835

Shapiro, P. R., & Field, G. B. 1976, ApJ, 205, 762

Solanes, J. M., Manrique, A., García-Gómez, C., et al. 2001, ApJ, 548, 97

Solanes, J. M., Sanchis, T., Salvador-Solé, E., Giovanelli, R., & Haynes, M. P. 2002, AJ, 124, 2440

Springel, V. 2005, MNRAS, 364, 1105

Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289

Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, ApJS, 160, 149

Stark, D. V., Kannappan, S. J., Eckert, K. D., et al. 2016, ApJ, 832, 126

Steinmetz, M., & White, S. D. M. 1997, MNRAS, 288, 545

Stevens, A. R. H., Croton, D. J., & Mutch, S. J. 2016, ArXiv e-prints

Storchi-Bergmann, T., Calzetti, D., & Kinney, A. L. 1994, ApJ, 429, 572

Stott, J. P., Sobral, D., Bower, R., et al. 2013, MNRAS, 436, 1130

Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810

Stroe, A., Oosterloo, T., Röttgering, H. J. A., et al. 2015, MNRAS, 452, 2731

Swaters, R. A., van Albada, T. S., van der Hulst, J. M., & Sancisi, R. 2002, A&A, 390, 829

Taffoni, G., Mayer, L., Colpi, M., & Governato, F. 2003, MNRAS, 341, 434

Telford, O. G., Dalcanton, J. J., Skillman, E. D., & Conroy, C. 2016, ApJ, 827, 35

Tinsley, B. M., & Larson, R. B. 1978, ApJ, 221, 554

Tomozeiu, M., Mayer, L., & Quinn, T. 2016, ApJ, 818, 193

Tonnesen, S., & Bryan, G. L. 2009, ApJ, 694, 789

Tonnesen, S., & Bryan, G. L. 2010, ApJ, 709, 1203

Tonnesen, S., Bryan, G. L., & van Gorkom, J. H. 2007, ApJ, 671, 1434

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Tukey, K. 1958, Ann. Math. Statist., 29, 614

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

van den Bosch, F. C., Aquino, D., Yang, X., et al. 2008, MNRAS, 387, 79

van der Wel, A. 2008, ApJ, 675, L13

Verheijen, M., van Gorkom, J. H., Szomoru, A., et al. 2007, ApJ, 668, L9

Vollmer, B., Wong, O. I., Braine, J., Chung, A., & Kenney, J. D. P. 2012, A&A, 543, A33

Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, Ap&SS, 331, 1

Walker, I. R., Mihos, J. C., & Hernquist, L. 1996, ApJ, 460, 121

Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563

Wang, L., Yang, X., Luo, W., et al. 2011, ArXiv pre-prints (arXiv:1110.1987)

Wang, L., Yang, X., Shen, S., et al. 2014, MNRAS, 439, 611

Warren, M. S., Abazajian, K., Holz, D. E., & Teodoro, L. 2006, ApJ, 646, 881

Weinberg, M. D. 1998, MNRAS, 299, 499

Wetzel, A. R., Cohn, J. D., & White, M. 2009, MNRAS, 395, 1376

Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, MNRAS, 424, 232

Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, MNRAS, 432, 336

White, M., Cohn, J. D., & Smit, R. 2010, MNRAS, 408, 1818

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

Whitmore, B. C., & Gilmore, D. M. 1991, ApJ, 367, 64

Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009, MNRAS, 393, 99

Wilman, D. J., & Erwin, P. 2012, ApJ, 746, 160

Wong, O. I., Ryan-Weber, E. V., Garcia-Appadoo, D. A., et al. 2006, MNRAS, 371, 1855

Wong, T., & Blitz, L. 2002, ApJ, 569, 157

Woo, J., Dekel, A., Faber, S. M., et al. 2013, MNRAS, 428, 3306

Wyder, T. K., Martin, D. C., Schiminovich, D., & et al. 2007, ApJS, 173, 293

Yang, X., Mo, H. J., & van den Bosch, F. C. 2008, ApJ, 676, 248

Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005, MNRAS, 356, 1293

Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2007, ApJ, 671, 153

Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215

Yoon, I., & Rosenberg, J. L. 2015, ApJ, 812, 4

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zahid, H. J., Bresolin, F., Kewley, L. J., Coil, A. L., & Davé, R. 2012, ApJ, 750, 120

Zahid, H. J., Geller, M. J., Kewley, L. J., et al. 2013, ApJ, 771, L19

Zahid, H. J., Kewley, L. J., & Bresolin, F. 2011, ApJ, 730, 137

Zaritsky, D., Kennicutt, Jr., R. C., & Huchra, J. P. 1994, ApJ, 420, 87

A

Chapter 4: Data Points for Gas Fraction Scaling Relations

Table A.1: Average gas fractions for the scaling relations shown in Figure 4.2. The column labelled y is the secondary property and limits by which we bin the sample.

x	y	$\langle x \rangle$	$\langle M_{HI}/M_\star \rangle$	N
$\log M_\star$	$7 \leq \log \mu_\star < 8$	9.20	1.722 ± 0.029	4204
		9.59	1.000 ± 0.017	2808
		10.02	0.589 ± 0.018	363
	$8 \leq \log \mu_\star < 8.8$	9.24	0.735 ± 0.030	1226
		9.67	0.429 ± 0.008	3733
		10.13	0.275 ± 0.004	3888
		10.58	0.174 ± 0.002	1488
		11.03	0.114 ± 0.008	119
	$8.8 \leq \log \mu_\star \leq 10$	9.75	0.168 ± 0.029	360
		10.19	0.095 ± 0.008	2038
		10.64	0.056 ± 0.001	2924
		11.08	0.032 ± 0.002	1092
$\log \mu_\star$	$9 \leq \log M_\star < 9.75$	7.43	1.990 ± 0.028	2328
		7.84	1.151 ± 0.010	5070
		8.27	0.538 ± 0.016	2222
		8.67	0.288 ± 0.016	740
		9.13	0.476 ± 0.120	44
	$9.75 \leq \log M_\star < 10.5$	7.53	0.993 ± 0.088	21
		7.95	0.622 ± 0.020	1314
		8.31	0.328 ± 0.007	2868
		8.74	0.159 ± 0.009	3638
		9.18	0.068 ± 0.007	1522
	$10.5 \leq \log M_\star \leq 11.5$	8.38	0.238 ± 0.012	252
		8.80	0.098 ± 0.002	1887
		9.21	0.037 ± 0.002	2360

Table A.2: Average gas fractions for the scaling relations shown in Figure 4.3. Numbers preceded by a "<" sign are upper limits.

x	y	$\langle x \rangle$	$\langle M_{HI}/M_\star \rangle$	N
log $M_\star$	$1 \leq$ NUV-$r <$ 3	9.20	1.805 ± 0.017	4464
		9.62	0.953 ± 0.021	3988
		10.10	0.510 ± 0.008	1684
		10.57	0.318 ± 0.012	429
		11.04	0.186 ± 0.017	39
	$3 \leq$ NUV-$r <$ 5	9.23	0.355 ± 0.040	898
		9.66	0.286 ± 0.013	2201
		10.15	0.186 ± 0.003	2763
		10.62	0.127 ± 0.003	1883
		11.05	0.079 ± 0.007	364
	$5 \leq$ NUV-$r \leq$ 8	9.24	< 0.258	144
		9.70	0.071 ± 0.016	715
		10.17	0.057 ± 0.004	1843
		10.64	0.026 ± 0.001	2112
		11.09	0.015 ± 0.001	810
NUV-r	$9 \leq$ log $M_\star <$ 9.75	1.97	1.979 ± 0.028	3819
		2.68	0.910 ± 0.033	4480
		3.70	0.286 ± 0.029	1119
		4.80	0.192 ± 0.051	707
		5.63	< 0.123	339
	$9.75 \leq$ log $M_\star <$ 10.5	2.04	0.865 ± 0.023	566
		2.81	0.429 ± 0.005	3067
		3.74	0.196 ± 0.010	2091
		4.82	0.106 ± 0.004	1506
		5.82	0.048 ± 0.004	2135
	$10.5 \leq$ log $M_\star \leq$ 11.5	2.00	0.304 ± 0.041	22
		2.95	0.234 ± 0.003	548
		3.78	0.120 ± 0.005	902
		4.80	0.053 ± 0.003	849
		6.00	0.021 ± 0.001	2187

Table A.3: Average gas fractions for the scaling relations shown in Figure 4.4. Upper limits are preceded by a "<" sign.

x	y	$\langle x \rangle$	$\langle \mathbf{M_{HI}/M_\star} \rangle$	N
$\log\mu_\star$	$1 \leq$ NUV-$r <$ 3	7.43	2.069 ± 0.026	2223
		7.85	1.238 ± 0.035	5023
		8.27	0.649 ± 0.012	2417
		8.67	0.435 ± 0.006	810
		9.21	0.358 ± 0.072	70
	$3 \leq$ NUV-$r <$ 5	7.48	0.346 ± 0.038	122
		7.91	0.366 ± 0.021	1279
		8.31	0.244 ± 0.020	2445
		8.74	0.162 ± 0.004	3318
		9.16	0.106 ± 0.005	941
	$5 \leq$ NUV-$r \leq$ 8	7.95	< 0.318	86
		8.35	< 0.089	480
		8.80	0.050 ± 0.004	2137
		9.21	0.028 ± 0.001	2915
NUV-r	$7 \leq \log\mu_\star <$ 8	1.97	2.040 ± 0.037	3353
		2.65	0.969 ± 0.021	3437
		3.64	0.394 ± 0.049	452
		4.71	< 0.294	115
		5.69	< 0.512	27
	$8 \leq \log\mu_\star <$ 8.8	2.02	1.114 ± 0.041	954
		2.81	0.489 ± 0.009	4304
		3.72	0.203 ± 0.009	2649
		4.78	0.140 ± 0.018	1506
		5.75	0.048 ± 0.017	1041
	$8.8 \leq \log\mu_\star \leq$ 10	1.91	0.562 ± 0.110	53
		2.95	0.244 ± 0.013	339
		3.83	0.146 ± 0.009	1008
		4.85	0.071 ± 0.007	1441
		5.93	0.030 ± 0.002	3591

B

Dependency of M-Z-SFR Relation on Choice of SFR Indicator

Previous work has shown the anti-correlation of metallicity (Z) with star formation rate (SFR) at fixed stellar mass (M) to be dependent on choice of SFR indicator used (Salim et al., 2014; Telford et al., 2016). In addition to the total and fibre star formation rates used in Chapter 7, here we show the correlation of the scatter in the mass-metallicity relation (MZR) with the total SFR estimate taken from the GALEX-SDSS-WISE Legacy Catalog (GSWLC; Salim et al., 2016). Figure B.1 shows that there is no significant difference between these two SFR indicators as drivers of scatter in the MZR.

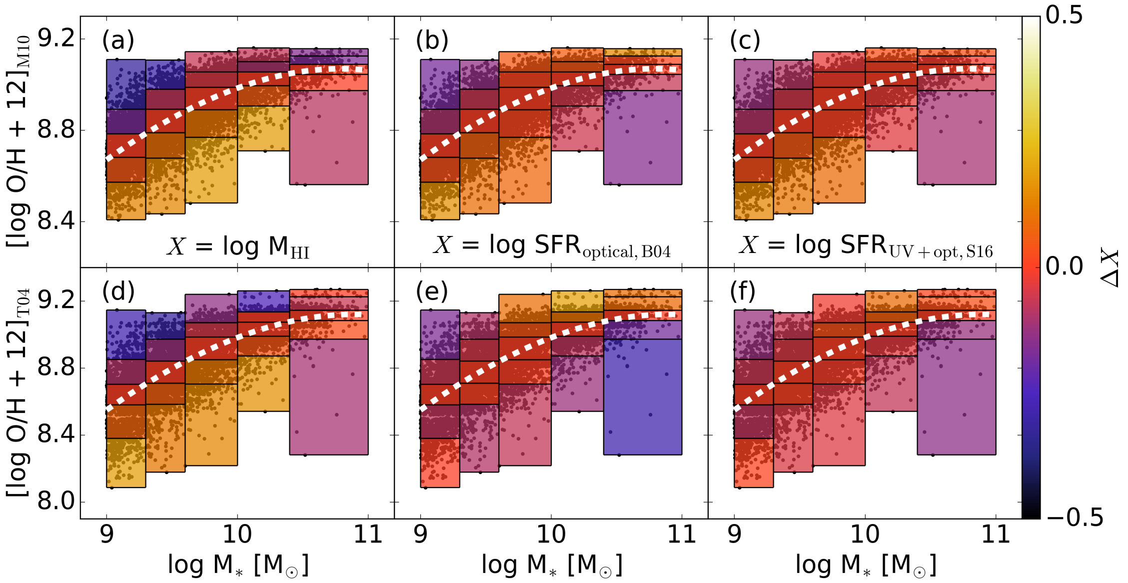

Figure B.1: Mannucci et al. (2010, top row) and Tremonti et al. (2004, bottom row) MZRs. The scatter in each relation is binned and coloured according to ΔX, where $\Delta X = X - X_{MZR}$ (see Equation 7.3.1). For each column, X is given in the top panel. H I mass measured in $M_\odot$ (left) while the units of total SFR from the MPA-JHU catalogue (middle; Brinchmann et al., 2004) and GSWLC total SFR (right; Salim et al., 2016) are $M_\odot$ yr^{-1}.