

Sample selection effects in LADUMA stacking experiments

A thesis submitted in partial fulfilment of the requirements

for the degree of

Magister Scientiae

in the

Department of Physics & Astronomy

The University of the Western Cape



Author: Lwandile Gwebushe

(Student Number: 3829533)

Supervisor: Prof. E. Elson

Co-Supervisor: Prof. R. Maartens

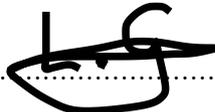
December 8, 2023

Declaration

I, Lwandile Gwebushe, solemnly declare that the thesis titled "**Sample selection effects in LADUMA stacking experiments**" and the work presented therein are solely my own original work. I affirm that:

- This research was undertaken entirely during my candidature for a research degree at this University.
- I have clearly indicated any prior submissions of parts of this thesis for a degree or any other qualification, whether at this University or any other institution.
- Whenever I have referred to the published work of others, I have consistently provided clear attribution.
- In cases where I have quoted from the work of others, I have diligently provided proper citation and acknowledged the original source. Apart from such explicit quotations, the entirety of this thesis represents my own independent work.
- I have duly recognized and acknowledged all significant sources of assistance received during the course of this research.
- In instances where the thesis is founded on collaborative work, jointly conducted with others, I have explicitly delineated the respective contributions of all individuals involved, distinguishing my own contributions from theirs.

Full Name: **Lwandile Gwebushe**

Signature: 

Date: **20 February 2024**

Acknowledgements

I would like to express my heartfelt appreciation to my supervisor, Prof. Ed Elson, for his unwavering guidance and support throughout this thesis. Your mentorship has been invaluable in shaping the direction of this work, and I am truly grateful for the opportunities you provided me to explore and grow in my research journey. Your patience and encouragement were particularly instrumental, especially during the moments when I sought a deeper understanding of the project. This thesis owes much of its success to your dedicated guidance, and I am sincerely thankful for the privilege of working under your supervision.

I would also like to extend my gratitude to my co-supervisor, Prof. Roy Maartens, for his valuable support and contributions to this thesis. Your expertise and insights have been instrumental in enriching the research process, and I am thankful for the guidance and encouragement you provided throughout this journey. Your involvement has been integral to the success of this work, and I deeply appreciate your assistance and encouragement.

I would like to express my sincere appreciation to the South African Radio Astronomy Observatory (SARAO), the National Research Foundation (NRF), the Centre for Radio Cosmology (CRC), and the Astronomy research group at the University of the Western Cape for their generous financial support. Their funding has been crucial in helping me carry out this meaningful and exciting research. I am also grateful to the Inter-University Institute for Data Intensive Astronomy (IDIA) for providing me access to the ilifu cloud computing facility, which has played a vital role in conducting my research efficiently. I would also like to thank the dedicated support team at IDIA for their invaluable assistance and technical guidance, which greatly contributed to the success of this project.

I would like to thank Xola Ndaliso, Siyambonga Matshawule, Junaid Townsend, and Narusha Isaacs for their unwavering support throughout my thesis. Their encouragement and assistance have been incredibly valuable in guiding me through this research journey, and I am truly grateful for their presence and contributions. Their support has played a significant role in shaping the success of this thesis, and I am thankful for their presence during this meaningful endeavor.

To Zandile Mpupa and my family, thank you for being there each and every day.

I appreciate your motivation and unwavering support throughout this work. Your presence and encouragement have been instrumental in helping me stay focused and determined during this journey. I am truly grateful for your love and support, which has been a source of strength and inspiration throughout this endeavor.

Lastly, I would like to commend myself for persevering and continuously moving forward, despite the challenges faced along this journey. It has not been an easy path, but my determination and resilience have kept me going. I am proud of the progress I have made and the growth I have experienced throughout this endeavor. Through all the ups and downs, I have remained steadfast in pursuing my goals, and I am grateful for the strength and determination that have brought me to this point. "Ithemba alibulali".



Abstract

Sample selection effects in LADUMA HI stacking experiments

L Gwebushe

M.Sc. Thesis

Department of Physics & Astronomy

The University of the Western Cape

This study explores the impact of selection effects introduced by input optical catalogues on high-redshift stacking experiments. Using the HI stacking technique, the HI content of LADUMA synthetic data cubes is analyzed within the redshift range of $0.7 < z < 0.758$. The focus is on galaxies with stellar masses of $M_{\star}/M_{\odot} \geq 10^{8.5}$.

The investigation reveals several key findings. Firstly, the input optical catalogues displays biases towards specific galaxy types, including those with lower stellar masses, lower star-formation rates, lower specific star-formation rates, galaxies in groups, and red galaxies. These biases can lead to higher confusion rates in stacked spectra, emphasizing the need for appropriate correction methods to avoid overestimation in subsequent quantitative analyses. Secondly, the environmental impact on HI gas properties is examined by comparing group galaxies and non-group galaxies. Non-group galaxies exhibit higher HI richness and HI gas fractions compared to group galaxies. Correlations between HI properties and star-formation/stellar mass properties are observed, indicating that lower mass galaxies are relatively inactive within their environments and are at an earlier stage of evolution.

The study also addresses the correction of confusion rates in stacked spectra. While a correction method is implemented, it is found to result in overestimated HI depletion timescales. Criticism is directed towards the use of this method, stressing the importance of employing accurate and robust approaches to correct for confusion rates and ensure precise estimation of galaxy properties. Furthermore, the derived scaling relations from the stacked spectrum, accounting for confusion, are consistent with previous stacking experiments, except for the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ relation. In this case, the results from simulations align with a study by [Rhee et al. \(2023\)](#), suggesting the contrast arises from the lack of precise confusion rate correction in their work and differences in redshift intervals. The study concludes by outlining future research directions, including refining the understanding and correction of selection effects, improving confusion rate correction methods, exploring the redshift

dependency of scaling relations, integrating multi-wavelength data for a comprehensive understanding of galaxy properties, and expanding the sample size and depth of surveys. Overall, this study highlights the significance of considering selection effects and accurately correcting for confusion rates in high-redshift stacking experiments. It emphasizes the need for cautious interpretation of stacking results and provides valuable insights and recommendations to enhance methodologies and advance our understanding of galaxy properties.

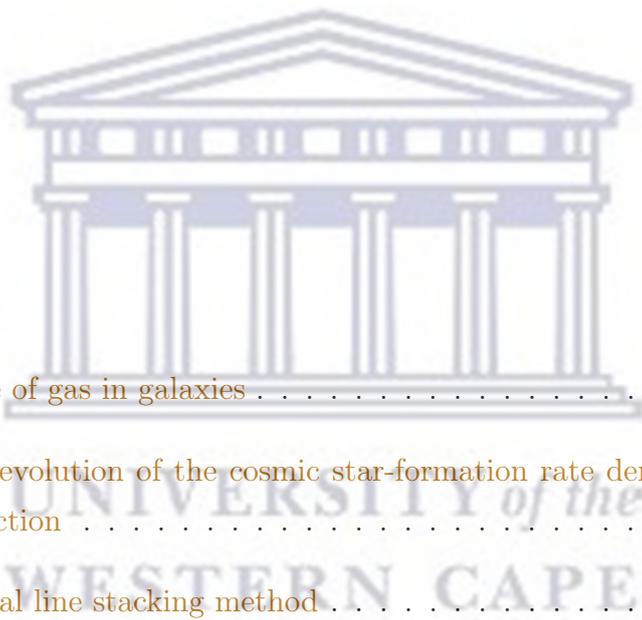
Keywords: galaxies:evolution – methods: data analysis – galaxies: ISM

December 8, 2023



Contents

Declaration	i
Acknowledgements	iii
Abstract	v
Contents	vi
List of Figures	ix
List of Tables	xiii
1 Introduction	1
1.1 Life-cycle of gas in galaxies	1
1.2 Redshift evolution of the cosmic star-formation rate density and HI mass function	4
1.3 HI spectral line stacking method	8
1.4 Sample selection effects in stacking experiments	10
1.5 The LADUMA MeerKAT Large Survey Project	12
1.6 Aims	14
2 Simulated data products	16
2.1 Semi-analytic models	16
2.2 HI data cubes of galaxies	20
2.2.1 HI mass profiles	22



2.2.2	Rotation curves	23
2.2.3	Three-dimensional modelling	24
2.3	LADUMA HI data cube	27
2.3.1	Intrinsic and observable properties of the LADUMA cubes.	28
2.3.2	The angular 2 point correlation function	34
3	HI stacking techniques	37
3.1	Spectral extraction	38
3.2	Spectrum de-redshifting and wrapping	41
3.3	Co-adding the spectra	44
4	HI stacking results	46
4.1	Stacking in all the galaxies	48
4.2	Stacking in stellar mass bins	51
4.3	Stacking in SFR bins	56
4.4	Stacking in sSFR bins	60
4.5	Stacking in blue and red galaxies	64
4.6	Comparing confusion rate for all sub-samples	69
5	HI content of galaxies	71
5.1	Group vs Non-Group galaxies	71
5.2	Confusion rate in Group and Non-Group galaxies	83
6	Correcting for confusion in the stacked spectrum	85

6.1	Correcting for confusion by fitting and subtracting the second-order polynomial on the stacks	85
6.2	HI depletion timescale	92
7	HI scaling relations	95
8	Summary and Future outlook	103
8.1	Summary	103
8.2	Future outlook	105
	Bibliography	107



List of Figures

1.1	Global scale Kennicutt-Schmidt law	2
1.2	Kilo-pacsec scales Kennicutt-Schmidt law	3
1.3	The cosmic neutral atomic hydrogen density from different surveys	6
1.4	The HI Mass Function at different redshifts	7
1.5	The stacking results from ALFALFA blind survey separated in stellar mass bins at redshift range $0.025 < z < 0.05$	9
1.6	The HI spectra for different mass components extracted from the mock LADUMA HI data cube in Elson et al. (2016)	11
1.7	The average contamination HI mass contributed to the stacked spectrum at different resolutions	12
1.8	The M_{HI} data from the S^3 -SAX catalogue	13
2.1	HI and H ₂ surface density profiles	18
2.2	The HI and H ₂ mass functions	19
2.3	The $M_{\text{HI}} - r_{\text{HI}}$ relation	20
2.4	The Tully-Fisher relation	22
2.5	The HI mass profiles for two galaxies from Obreschkow & Meyer (2014) catalogue produced by Elson et al. (2016)	24
2.6	The circular velocity profiles of galaxies	25
2.7	The channel maps for a single galaxy model	26
2.8	The HI data products for a single galaxy models	27
2.9	The LADUMA HI data cube channel maps	29
2.10	The total intensity map	30

2.11	The LADUMA HI data cube galaxy properties	32
2.12	The LADUMA scaling relations	33
2.13	Angular 2 point correlation function	35
3.1	The stacking results obtained from Chengalur et al. (2001)	38
3.2	Illustration of the HI line extraction	39
3.3	The extracted HI profiles	40
3.4	The wrapped HI profiles	43
3.5	The stacked Gaussian noise	45
4.1	Stacking results for all the galaxies in the sample	49
4.2	The confusion rate for the full-sample plotted against the number of stacked galaxies	50
4.3	The stacking results for the 1st stellar mass bin, with $8.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 9$	52
4.4	The stacking results for all the stellar mass bins	53
4.5	Confusion rate for each stellar mass bin.	54
4.6	The star-formation rate distribution for our sample	57
4.7	The stacking results for the 1st SFR bin, with $\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) \leq -0.9$	58
4.8	The stacking results for all the SFR bins	59
4.9	Confusion rate for each SFR bin.	60
4.10	The specific star-formation rate distribution for our sample	61
4.11	The stacking results for the 1st sSFR bin, with $\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$	62
4.12	The stacking results for all the sSFR bins	63

4.13	Confusion rate for each sSFR bin.	64
4.14	The red and blue galaxies on our sample	65
4.15	The distribution of the galaxy's colour in stellar mass	66
4.16	The stacking results for the blue and red samples	67
4.17	The confusion rate for red and blue samples as function of the number of stacked galaxies	68
5.1	The Friends-of-Friends algorithm	72
5.2	RA- and Dec-projected co-moving distance of the non-group and group sub-samples	73
5.3	The 3 dimensional cube of the non-group and group sub-samples	74
5.4	The properties of the non-group and group sub-samples	75
5.5	The stacking results for non-group and group sub-samples	76
5.6	The stellar mass to specific star-formation rate plane for non-group and group galaxies	79
5.7	The stacking results for the sub-classes of the non-group sub-sample	80
5.8	The stacking results for the sub-classes of the group sub-sample	81
6.1	The stacking results from Chowdhury et al. (2022)	87
6.2	The HI depletion timescale results from Chowdhury et al. (2022)	88
6.3	Red and blue galaxies with $M_*/M_\odot \geq 10^9$ in our sample	88
6.4	The stacking results for the first sub-sample, with $10^9 \leq M_*/M_\odot <$ 10^{10}	89
6.5	The stacking results for the first sub-sample, with $M_*/M_\odot > 10^{10}$	90
6.6	The average HI depletion timescale $\langle t_{\text{dep}} \rangle = \langle M_{\text{HI}} \rangle / \langle \text{SFR} \rangle$ for both stellar mass sub-samples.	93

7.1	Scaling relations from Maddox et al. (2021)	96
7.2	HI scaling relations for star-formation galaxies of our sample	97



List of Tables

2.1	Table show description of the columns of mock catalogue in ASCII format	21
2.2	Table summarising user specified inputs for LADUMA synthetic data cubes	28
4.1	The resulting parameters for the full sample	50
4.2	The resulting parameters for all the stellar mass bins	53
4.3	The resulting parameters for all the SFR bins	59
4.4	The resulting parameters for all the sSFR bins	64
4.5	The resulting parameters for red and blue samples	68
4.6	The resulting parameters for all the sub-samples	70
5.1	The resulting parameters for the non-group and the group sub-samples	78
5.2	The stacking results parameters for both sub-classes of the non-group and the group sub-samples	82
5.3	The $\langle \text{SFR} \rangle$ and $\langle t_{\text{dep}} \rangle$ for the sub-classes SQ and SFMS separated in stellar mass	83
6.1	The resulting parameters for the stacked sub-samples before and after fitting to and subtracting from the stacks the second and fourth order polynomial	91
6.2	The the average SFR for both sub-samples and the average HI depletion timescale for before and after correcting for the confusion rate.	94
7.1	The resulting parameters for the scaling relations	98

1 Introduction

1.1 Life-cycle of gas in galaxies

Gas is a critical component of galaxies as it supplies the fuel necessary for star formation. Understanding the structure and evolution of galaxies requires knowledge of their gas properties. The formation of stars serves as a reliable indicator of galaxy evolution, making star formation a central process in this regard. However, our current understanding of the conditions required for stars to form remains limited.

The star formation process not only leads to the creation of new stars but also enriches the interstellar medium (ISM) with metals through energy injection. Additionally, the ejection of gas during star formation contributes to the enrichment of the intergalactic medium (IGM) (de Blok et al., 2015). Another significant process in galaxy evolution is gas accretion, which transports gas from the outer regions of galaxies into the star-forming disk. This mechanism ensures that galaxies can continue forming stars over extended periods, spanning a Hubble time (de Blok et al., 2015). The presence of neutral atomic hydrogen (HI) serves as the raw fuel for star formation and can be detected through the 21-cm emission line.

The molecular hydrogen (H_2) within galaxies plays a crucial role in the formation of stars and thus influences galaxy evolution. The study of H_2 content in galaxies provides valuable insights into their properties. The Kennicutt-Schmidt law (Kennicutt Jr, 1998) established a direct link between the surface density of total gas, H_2 and the surface density of star formation rate (SFR) on both global scales (Figure 1.1) and kiloparsec scales (Figure 1.2). Recent studies have further analyzed the Kennicutt-Schmidt law and confirmed its validity for disk galaxies and late-type dwarf galaxies (Bigiel et al., 2008; Leroy et al., 2008, 2012; Elson et al., 2019a). However, the precise understanding of the star formation process in galaxies remains limited due to technological constraints (Healy et al., 2019).

N-body and hydrodynamic cosmological simulations have provided insights into two modes of gas accretion: the "Cold mode" and the "Hot mode" (Katz et al., 2003; Kereš et al., 2005). The "Cold mode" accretion refers to gas that does not experience significant shock heating, instead radiating its acquired gravitational energy. This mode dominates gas accretion consistently (Kereš et al., 2005). In the "Cold mode,"

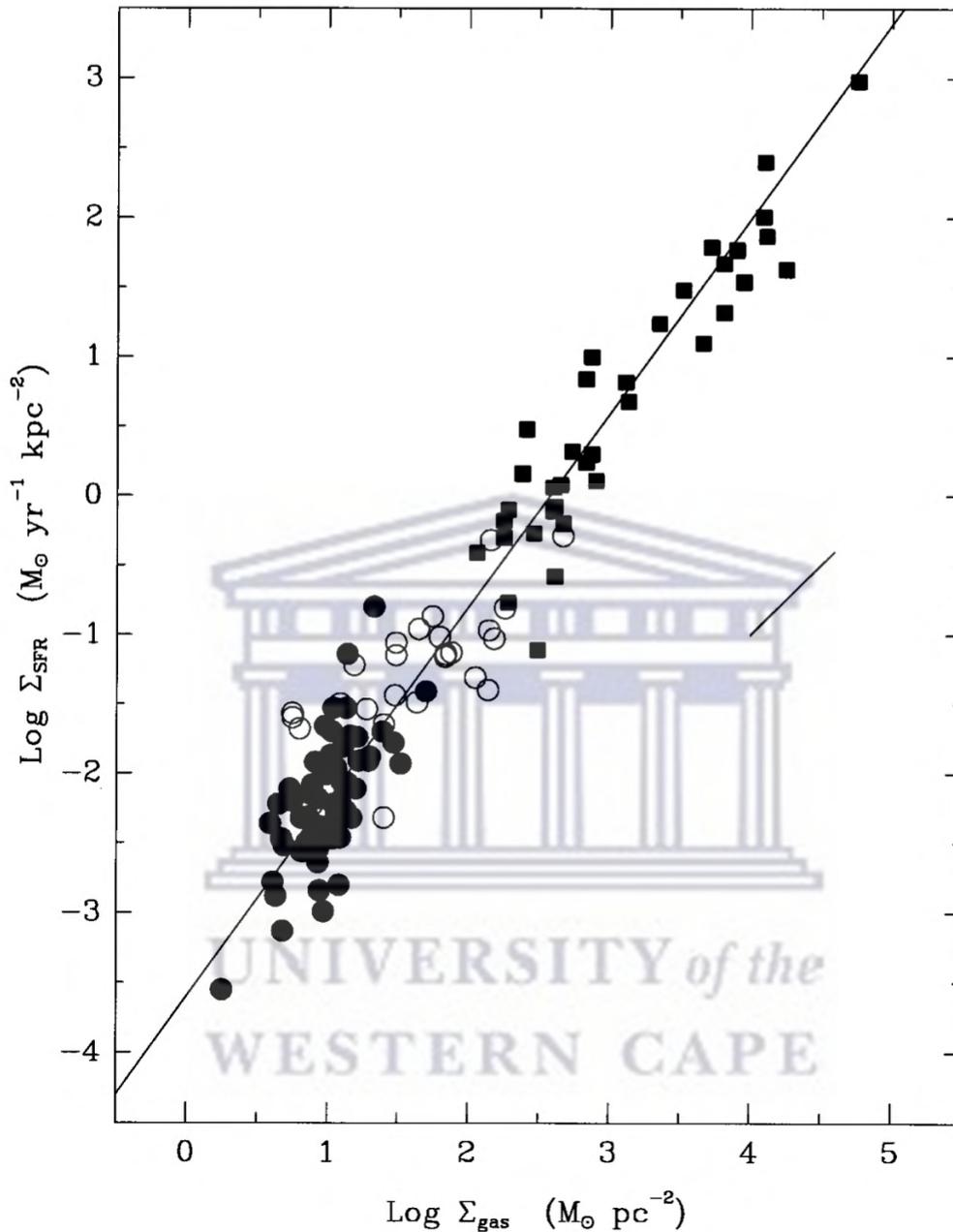


Figure 1.1: Show the Kennicutt-Schmidt law: relationship between gas surface density and star-formation rate surface density on global scale, for the spiral (or disk) galaxies represented by filled circles and starburst galaxies represented by filled squares (Kennicutt Jr, 1998).

gas flows into the dark matter halo along filaments of the cosmic web or IGM. On the other hand, "Hot mode" accretion follows the standard model, where gas is shock-heated to a temperature approximately equal to the virial temperature ($T \sim T_{\text{vir}}$). Gas accreted in the "Hot mode" enters the halo quasi-spherically and undergoes

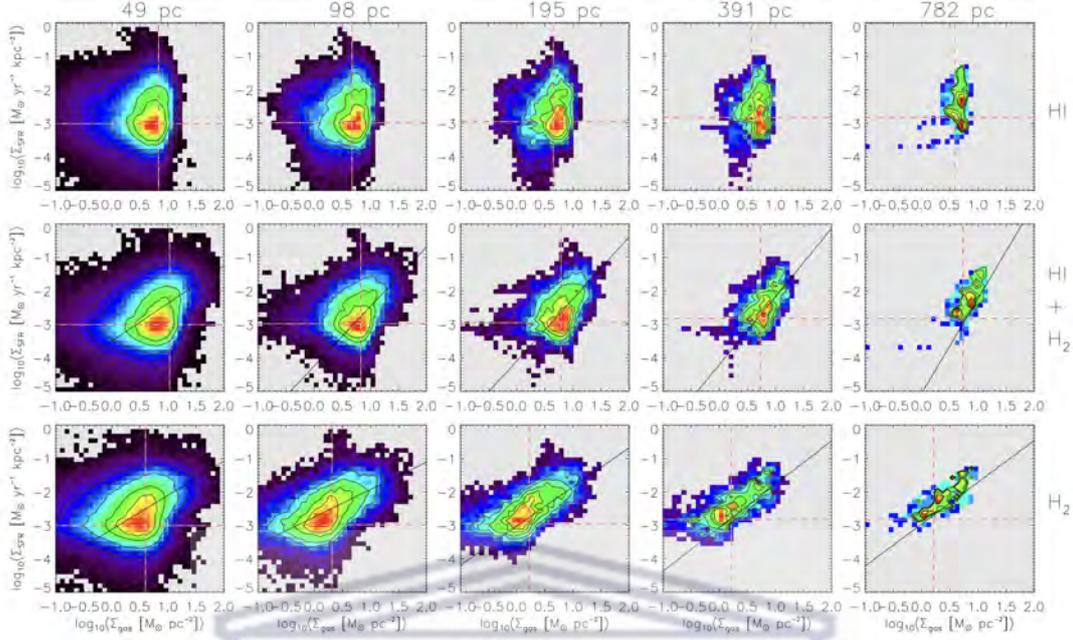


Figure 1.2: An example of a multi-scale study of the SF law for the nearby galaxy M33. It presents results adapted from Elson et al. (2019a) for the relation between SFR surface densities and gas surface densities at which the star formation law was measured. Shown above each panel is the spatial length-scale at which the star formation law was measured. Elson et al. (2019a) find the slope of the star formation law to be consistent with unity over a range of length-scale spanning ~ 50 to 780 pc for the relation between Σ_{H_2} and Σ_{SFR} in the bottom row. The top and middle rows show the relation between Σ_{HI} and Σ_{SFR} and $\Sigma_{\text{HI}+\text{H}_2}$ and Σ_{SFR} respectively.

shock heating before cooling and forming stars (Katz et al., 2003; Kereš et al., 2005). Kereš et al. (2005) found that the dominance of "Cold mode" accretion is observed in lower-mass galaxies, while the conventional "Hot mode" accretion dominates the growth of high-mass systems. Furthermore, they noted that "Cold mode" accretion prevails at high redshifts ($z \geq 3$), while "Hot mode" accretion dominates at low redshifts.

Understanding how the HI (neutral hydrogen) content of galaxies changes with redshift is of paramount importance in astrophysics and cosmology. The HI gas plays a fundamental role in galaxy evolution, star formation, and the overall structure of galaxies. Investigating the evolution of HI gas across different redshifts provides crucial insights into the formation and evolution of galaxies throughout cosmic history.

One of the key reasons why studying the HI content as a function of redshift is important lies in its role as a reservoir of gas for star formation. Galaxies undergo

significant transformations over time, and their ability to form stars is closely linked to the availability of gas. By understanding how the HI content changes with redshift, researchers can gain valuable information about the gas accretion processes and star formation rates in galaxies at different cosmic epochs (Lah et al., 2007).

Moreover, studying the HI content across different redshifts enables us to probe the processes of gas consumption and feedback mechanisms in galaxies. The amount of HI gas in galaxies affects their evolution and overall baryonic content, which, in turn, influences their stellar and dynamical properties. Investigating how the HI content changes with redshift provides insights into the balance between gas accretion and gas removal, shedding light on the complex interplay between various physical processes that shape galaxies (Davé et al., 2011).

Furthermore, the HI content of galaxies is crucial for understanding the Epoch of Reionization, a critical phase in the universe's history when the intergalactic medium transitioned from a neutral to an ionized state. Observations of HI at high redshifts help us to probe the contribution of galaxies to the reionization process and provide constraints on the ionizing photon budget during this epoch (Crain et al., 2016).

Additionally, measurements of the HI content as a function of redshift have profound implications for cosmological models and simulations. By comparing observational data with theoretical predictions, researchers can test and calibrate models of galaxy formation and evolution. The observed HI mass function and its evolution provide valuable data for refining our understanding of the underlying processes that have shaped galaxies over cosmic time (Obreschkow et al., 2009a; Crain et al., 2016).

Overall understanding how the HI content of galaxies changes with redshift is crucial for gaining insights into galaxy formation, baryonic physics, the Epoch of Reionization, cosmological models, and the overall evolution of the universe.

1.2 Redshift evolution of the cosmic star-formation rate density and HI mass function

The understanding of HI gas in galaxies has been significantly shaped by nearby galaxy surveys such as the HI Parkes All-Sky Survey (HIPASS, Barnes et al., 2001) and the Arecibo Fast Legacy ALFA Survey (ALFALFA, Giovanelli et al., 2005).

These surveys have played a crucial role in evaluating the total mass density of neutral hydrogen, Ω_{HI} . While these surveys provide substantial knowledge about HI gas in the local universe, there is still a lack of well-constrained measurements of Ω_{HI} over the intermediate redshift range of $0.1 < z < 1.6$, even with the use of damped Lyman α absorbers (DLAs) (Elson et al., 2019b). To overcome this, stacking HI data from existing radio telescopes has been employed to measure Ω_{HI} at the lower end of this redshift range (for example: Lah et al., 2007, 2009; Delhaize et al., 2013). The LADUMA (Looking at the Distant Universe with the MeerKAT Array) survey utilizes HI detections of individual galaxies, as well as spectral line stacking, to probe a larger fraction of this redshift range (Blyth et al., 2016). Figure 1.3 illustrates the redshift-dependent changes in Ω_{HI} based on different studies.

Analysis of Figure 1.3 reveals that all results obtained through HI stacking are contained within the range of $0 < z < 0.4$. These studies show a slight linear relationship between Ω_{HI} and redshift (for example: Zwaan et al., 2005; Rhee et al., 2023; Jones, 2016; Delhaize et al., 2013). At higher redshifts, the estimates of Ω_{HI} rely on DLAs and MgII absorbers, which also exhibit a linear increase in Ω_{HI} with redshift (for example: Chowdhury et al., 2020; Prochaska et al., 2005; Bird et al., 2017).

The predictions made by different models regarding the cosmic gas density of galaxies and its variation with redshift exhibit some divergence. This divergence arises due to uncertainties in understanding galaxy formation, particularly regarding feedback from stars and active galactic nuclei (AGN), as well as the challenge of obtaining a wide spatial range to track the necessary gas properties (Delhaize et al., 2013). Furthermore, the limitations of current technology contribute to uncertainties in understanding the exact conditions required for star formation (Healy et al., 2019). To address these uncertainties, observational constraints become vital. In the absence of such constraints, theory can offer insights into the evolving properties of Ω_{HI} and cold gas in general (Elson et al., 2016). Semi-analytic models of galaxy formation, such as those proposed by Obreschkow et al. (2009a); Lagos et al. (2011); Popping et al. (2014), have focused on predicting the cold gas properties of galaxies (Elson et al., 2016).

Lagos et al. (2011) present self-consistent models for predicting the HI and H₂ content of galaxies. Their predictions, as shown in Figure 1.4, demonstrate that the HI mass function (HiMF) evolves weakly with redshift. The number density of galaxies

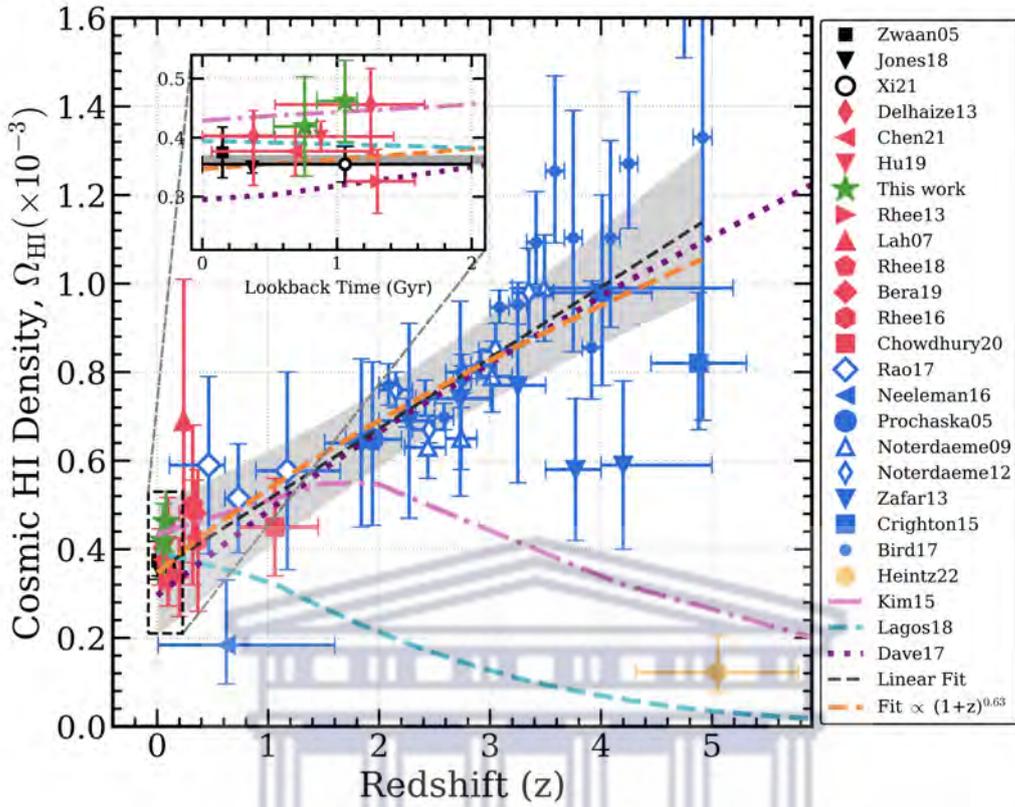


Figure 1.3: The cosmic neutral atomic hydrogen (HI) density Ω_{HI} computations collected by Rhee et al. (2023). There are three different techniques of determining the HI density shown in this plot: measurements from DLAs and MgII absorbers (colour-coded in blue), HI stacking technique (colour-coded in red and green) and directly detecting the HI in galaxies (colour-coded in black). The colours and shapes of each data point correspond to the author made the calculations shown in the legend.

with low HI masses increases with redshift, while the number density of massive galaxies decreases with increasing redshift. These predictions align well with observations from both the nearby universe and high-redshift surveys. Lagos et al. (2014) present updated models building upon Lagos et al. (2011), which predict the evolution of the cosmic HI density for $z < 3$, in agreement with observations in the local universe. According to their models, the HI density is dominated by galaxies with a star formation rate $\text{SFR} < 1 M_{\odot}\text{yr}^{-1}$ and low stellar mass of $M_{\star} < 10^9 M_{\odot}$.

Obreschkow et al. (2009a) present a simulation that explores the cosmic evolution of the atomic and molecular phases of cold hydrogen gas in approximately 3×10^7 galaxies. Their approach utilizes physical prescriptions based on the total cold gas masses of galaxies and other galaxy properties to assign HI and molecular hydrogen (H_2) to these galaxies. Their results encompass the local HI and H_2 mass functions,

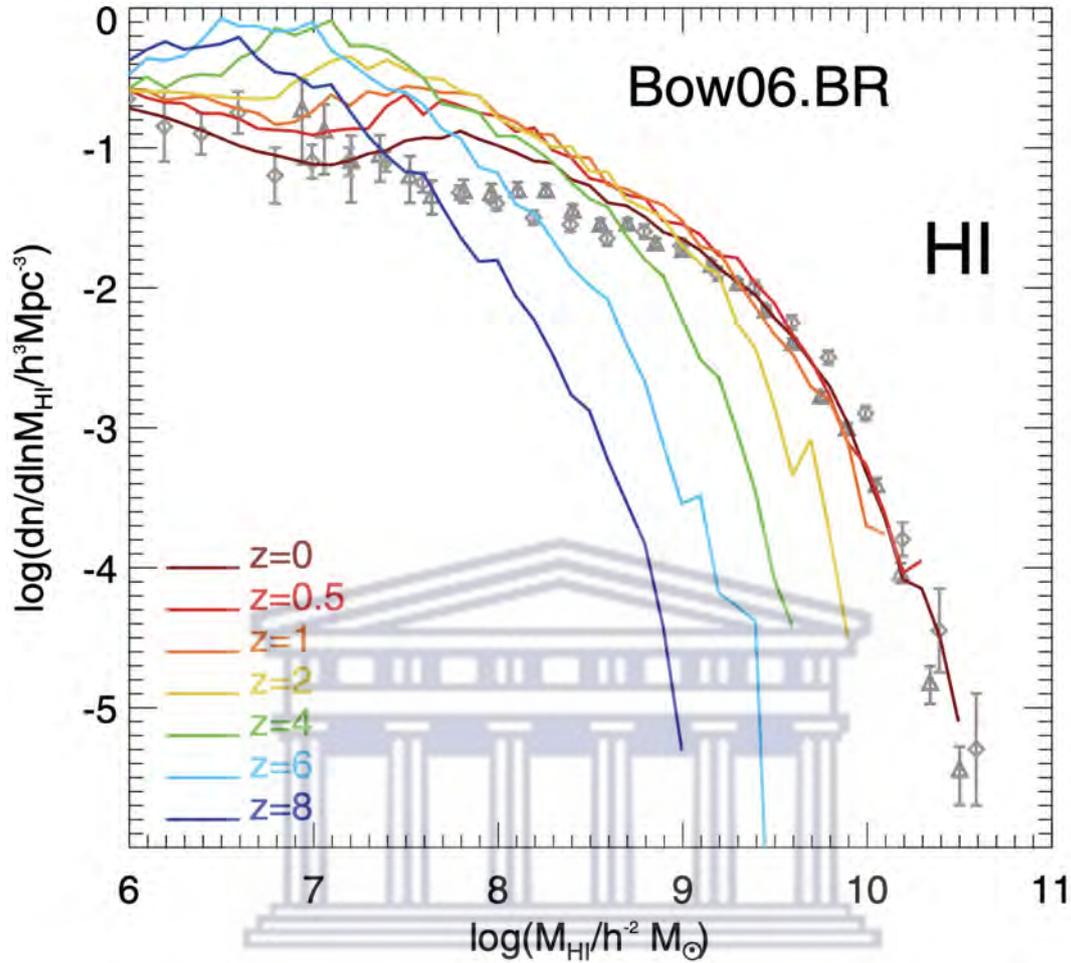


Figure 1.4: The HI Mass Function for the Bower et al. (2006) model at different redshifts from $z = 8$ to $z = 0$, for measurements collected by (Lagos et al., 2011). The legend shows the colours and the corresponding redshifts for each HIMF. The unfilled grey triangles and squares represent the observational results at $z = 0$ from Zwaan et al. (2005) and Martin et al. (2010).

the cold gas mass-diameter relation, the Tully-Fisher relation (TFR), and the CO luminosity function. All of these outcomes match observational data from the local universe. The authors also provide predictions for high-redshift scenarios, including cold gas diameters and the TFR, which display significant changes with lookback time.

Popping et al. (2014) employ semi-analytic models of galaxy formation to study the evolution of atomic and molecular gas in galaxies. They consider two scenarios for molecular formation: one pressure-based and one metallicity-based. The predictions from both scenarios agree with observed HI and H₂ mass functions in the local

universe. Their models predict that the high-mass end of the HI mass function remains relatively constant at redshifts $z < 2$, while galaxy gas fractions remain high from $z \sim 3$ to 6 and decline rapidly at lower redshifts. However, their predictions slightly overestimate the low-mass end of the HI mass function.

All of the aforementioned models emphasize the pivotal role played by HI in galaxy formation and evolution. To maximize their utility, [Elson et al. \(2016\)](#) developed a set of tools that convert these models into formats that can be treated as real data. This toolkit generates synthetic data cubes for large cosmological volumes, employing an input catalog of evaluated galaxy properties. The data cubes provide detailed models of the spectral and spatial distributions of HI in galaxies. These synthetic products can guide the planning of radio and multi-wavelength observations for future HI galaxy surveys, as well as facilitate the development of calibration, imaging, and analysis methods. They can be employed through mock HI stacking experiments for both low and high-redshift galaxy samples, offering valuable insights into their HI gas content.

Large HI galaxy surveys, such as the LADUMA survey ([Blyth et al., 2016](#)), aim to track the cosmic evolution of HI over extensive regions up to $z > 1$. However, direct methods of HI detection are limited to a few hundred megaparsecs, making it challenging to directly image high-redshift galaxies within a reasonable timeframe ([Elson et al., 2016](#)). The toolkit developed by [Elson et al. \(2016\)](#) addresses this issue. The HI spectral line stacking method enables the measurement of the total HI content of populations of galaxies at high redshifts, including galaxies that are individually too faint to detect. The LADUMA survey employs the MeerKAT telescope, one of the most powerful telescopes located in South Africa, to carry out this research.

1.3 HI spectral line stacking method

Investigating the HI content in galaxies is crucial for understanding galaxy evolution. Neutral atomic hydrogen (HI) serves as the primary ingredient for the formation of molecular hydrogen (H_2), which, in turn, plays a pivotal role in star formation. HI can be observed through the 21-cm emission line; however, the line is faint. Direct detection of HI over a statistically significant galaxy sample is possible at low redshifts ($z < 0.1$) and has already been accomplished through surveys of nearby

galaxies, as mentioned earlier. However, at higher redshifts ($z > 0.1$), direct detection of HI in the emission line becomes challenging. Over the past two decades, authors have employed the HI spectral line stacking technique to study the HI content in galaxies at high redshifts (Zwaan et al., 2001; Chengalur et al., 2001; Fabello, 2012; Sinigaglia et al., 2022).

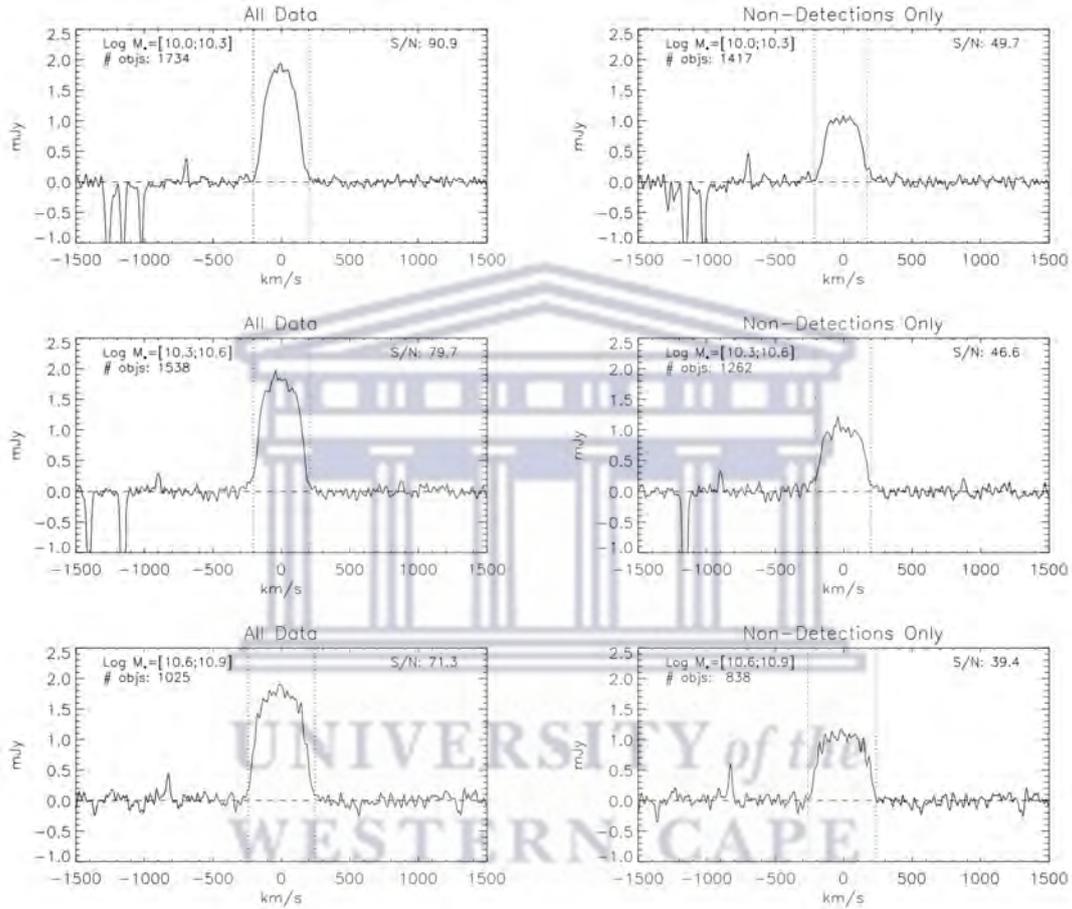


Figure 1.5: The spectra obtained in Fabello et al. (2011) using ALFALFA blind survey. The left column shows stacked spectra acquired from ALFALFA detections and non-detections. The right column shows spectra acquired only from stacking ALFALFA non-detections, to demonstrate that even if there is no galaxy detection in the sample, the stacking methods recover the signal.

The stacking method is a statistical technique that involves taking a sample of spectra predominantly classified as non-detections and creating a high signal-to-noise spectrum representative of the total HI mass in a galaxy sample, thereby providing the average HI mass of the sample. In this context, non-detections refer to sources whose individual HI spectra did not yield a statistically significant (i.e. above a given signal-to-noise ratio) detection. The stacking method follows a simple

principle: Small sub-volumes are extracted from the full HI data cube at the known optical positions of galaxies. These sub-volumes are converted into 1-D spectra and then aligned based on each galaxy's known redshift. Finally, the 1-D spectra are co-added to produce a high signal-to-noise spectrum that can be integrated to measure the average HI mass of the entire galaxy sample.

Initially, the HI stacking method was used solely to enhance the signal of undetected galaxies in the emission line. Due to the availability of a large number of spectra from nearby galaxy surveys (e.g., HIPASS and ALFALFA), the stacking process utilizes both detected and non-detected galaxy spectra, resulting in a more accurate estimation of the average HI content of the sample (Fabello et al., 2011; Brown et al., 2015). Figure 1.5 illustrates the stacking results obtained in Fabello et al. (2011) using the ALFALFA blind survey, with stellar mass bins at a redshift range of $0.025 < z < 0.05$. The left column displays spectra acquired from stacking both ALFALFA detections and non-detections, while the right column demonstrates spectra obtained solely from stacking ALFALFA non-detections, highlighting that the stacking method can recover the signal even in the absence of direct galaxy detection in the sample.

1.4 Sample selection effects in stacking experiments

The spectral stacking technique is a powerful tool commonly used to study the HI content of non-detected galaxies in the emission line. While the HI stacking method is conceptually straightforward, accurately measuring galaxy properties from stacked spectra is challenging due to confusion effects. When generating a spectrum for a target galaxy, the resulting spectrum often contains not only the emission from that galaxy but also additional flux from nearby galaxies. The contamination from neighboring galaxies must be corrected for before measuring any galaxy parameters. Several methods exist in the literature to address contamination in stacked spectra (for example: Fabello, 2012; Elson et al., 2016, 2019b; Jones, 2016).

Jones (2016) investigated the contribution of source confusion in HI stacking experiments within the redshift range of $z = 0.04 - 0.13$ using the ALFALFA correlation function. They found that the amount of confused mass per galaxy is approximately $\sim 1.3 \times 10^{10} M_{\odot}$. Furthermore, Elson et al. (2016) examined source confusion in mock LADUMA data cubes within the redshift ranges of $z = 0.04 - 0.13$ and

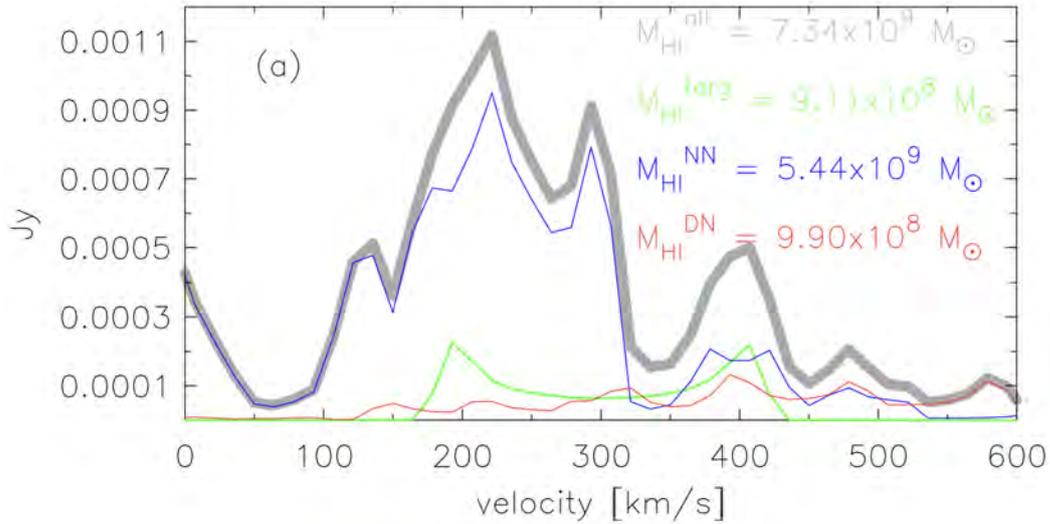


Figure 1.6: The HI spectra for different mass components extracted from the mock LADUMA HI data cube in [Elson et al. \(2016\)](#). The grey curve represents the whole flux in the extracted sub-volume. The green-curve represents the flux of the target galaxy in the sub-volume. The blue curve represents the flux contributed by the nearby neighbours (NN) and the red-curve represents the flux contributed by distant neighbours (DN). The NN and DN contribute a flux that is ~ 7 times greater than the target galaxy flux on the extracted sub-volume.

$z = 0.7 - 0.758$, employing different spatial resolutions. Their method involved decomposing the extracted spectrum into flux contributions from nearby neighbours (NN), distant neighbours (DN), and the target galaxy. In the low-redshift range, they determined an average contamination mass per galaxy of approximately $\sim 1.5 \times 10^{10} M_{\odot}$, similar to the findings in [Jones \(2016\)](#). Figure 1.7 presents the average contamination HI mass contributed by DN+NN (panel (a)), NN only (panel (b)), and DN only (panel (c)) at different resolutions for both low-redshift and high-redshift samples. The figure illustrates that high-redshift stacks exhibit a lower average contaminant mass compared to their low-redshift counterparts.

These aforementioned studies have significantly contributed to our understanding of source confusion and methods for correcting it in stacking experiments. However, when working with high-redshift optical catalogues, the challenge becomes even more complex, as the rates of source confusion depend not only on the 3D clustering of galaxies but also on the selection effects imposed by the optical catalogue used to locate galaxies in the HI data cube. High-redshift optical catalogues are typically magnitude-limited, introducing biases towards certain types of galaxies. For exam-

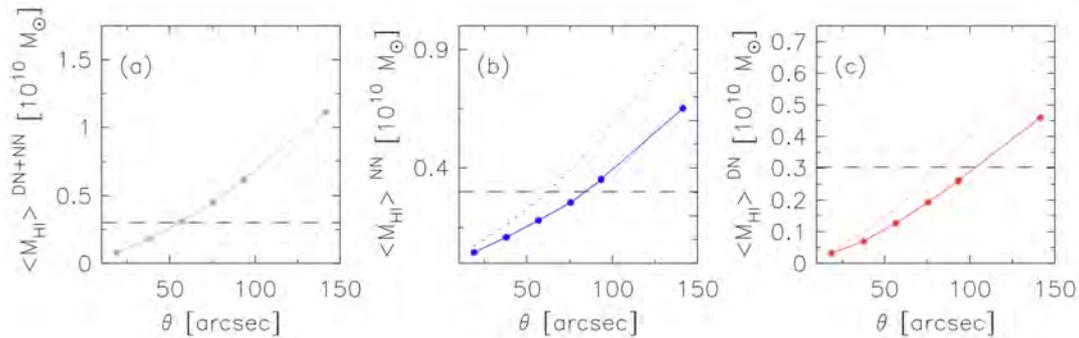


Figure 1.7: The average contamination HI mass contributed to the stacked spectrum at different resolutions. In all the panels the solid-curve represents the contaminant mass for high-redshift stacks, while the dashed curve represents the contaminant mass for low-redshift stacks. The horizontal black dashed line in all the panels represents the true evaluated average HI mass of the galaxies that were stacked in [Elson et al. \(2016\)](#). Image credit to [Elson et al. \(2016\)](#).

ple, stacking galaxies that are highly luminous in the B-band may predominantly include galaxies with high star-formation rates, affecting their HI content. Similarly, stacking earlier-type galaxies may be more susceptible to the confusion effects induced by spatial clustering. This work aims to conduct a detailed investigation into the ways in which the results of high-redshift stacking experiments are influenced by the incompleteness of optical catalogues.

1.5 The LADUMA MeerKAT Large Survey Project

Due to the long observation intervals required, directly detecting HI with modern radio telescopes beyond $z \sim 0.25$ is challenging. To overcome this limitation and observe a larger number of objects in a shorter time, next-generation telescopes such as the Square Kilometre Array (SKA) and SKA Pathfinders (MeerKAT in South Africa, and ASKAP in Australia) have been constructed with larger fields of view compared to existing telescopes.

One of the authorized Large Survey Projects for the ongoing MeerKAT telescope is the Looking At the Distant Universe with the MeerKAT Array (LADUMA) survey ([Holwerda et al., 2011](#)). Covering an area of the sky of approximately 1 deg^2 at $z = 0$, which increases to around 5.5 deg^2 at $z = 1.4$, LADUMA is set to be one of the deepest HI surveys conducted, with thousands of individually-detected galaxies

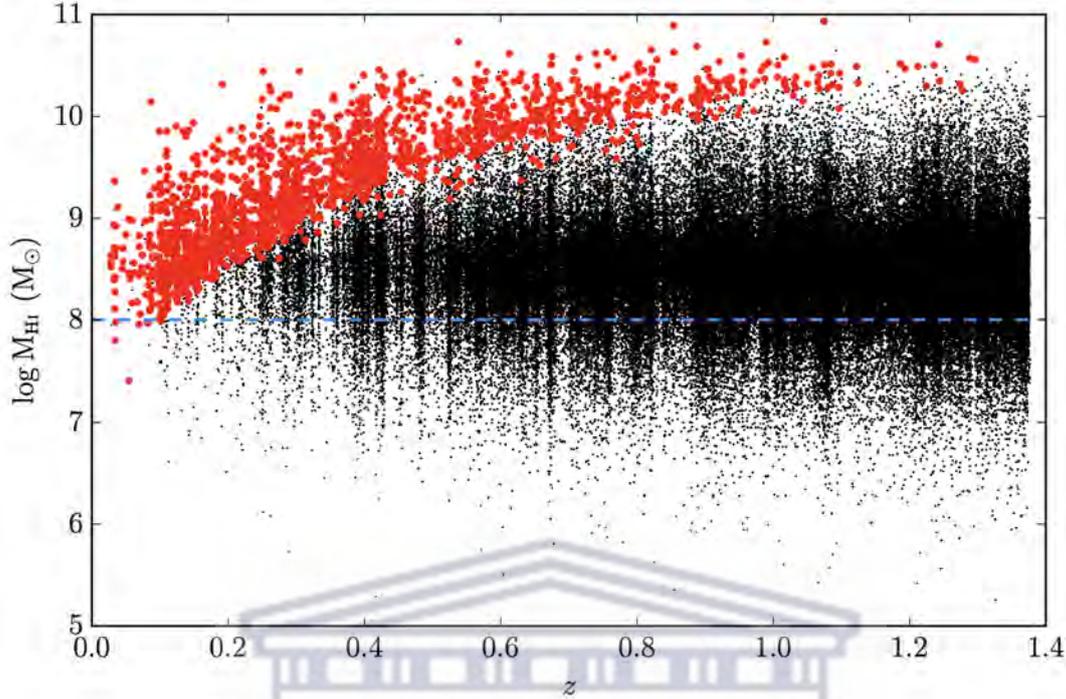


Figure 1.8: The data used for this figure are from the S^3 -SAX catalogue (Obreschkow et al., 2009a,b,c). The dashed blue line shows that the simulations used to generate the catalogue are complete for HI masses larger than $\log_{10}(M_{\text{HI}}/M_{\odot}) = 8$. The black data points represent the galaxies for which there will be no direct detection of HI, whereas the red data points represent the galaxies for which a 5σ detection is expected.

(Blyth, 2014). Despite the excellent sensitivity of MeerKAT (rms noise of 1.6 Jy per channel at a spectral resolution of 7 km.s^{-1}), LADUMA does not expect to directly detect every galaxy in the survey region. It is estimated that the LADUMA survey volume contains over 10,000 optical galaxies, but only around 3500 of those galaxies are likely to have 5σ HI detections (Blyth et al., 2016). A simulation of the HI mass distribution of galaxies in the LADUMA survey volume as a function of redshift is shown in Figure 1.8, indicating that hundreds of galaxies will remain undetected in the HI emission line.

The LADUMA survey aims to deepen our understanding of galaxy evolution by investigating the following key aspects:

- Cosmic Neutral Hydrogen Density (Ω_{HI}): By estimating the overall HI density at different epochs, researchers can gain insights into how HI has shaped the history of galaxies throughout cosmic time.

- The HI Mass Function (HIMF): Examining how the volume densities of galaxies change with their HI mass is essential for understanding the evolution of the cosmic star formation rate density. This quantity, also known as the HI atoms per unit volume or Ω_{HI} , plays a crucial role.
- Galaxy Evolution: Analyzing the HI content of different galaxy populations as a function of redshift will provide valuable insights into how galaxies evolve.
- Tully-Fisher (T-F) Relation: The LADUMA data will offer HI information for the first time at higher redshifts, enabling a comprehensive study of the entire disk (and subsequently the halo) with sufficient statistics to establish the normalization and slope of the T-F relation at various redshifts.

Direct detections of HI emission from galaxies at higher redshifts ($z \gtrsim 0.6$) will be challenging in the LADUMA dataset. Therefore, the LADUMA team intends to analyze high redshifts through HI stacking (Holwerda et al., 2011). Previous HI stacking investigations discussed in sections 1.3 and 1.4 have been limited to low redshifts ($z \lesssim 0.25$) due to sensitivity of the telescopes. However, with LADUMA's wide redshift range, HI stacking analyses that were previously restricted to the local universe can now be expanded to higher redshift regions.

1.6 Aims

The primary goals of this project are as follows:

- Generate a stacking Python algorithm and validate it using synthetic data products from Elson et al. (2016).
- Apply the stacking algorithm to high-redshift LADUMA synthetic spectral cubes to investigate the potential impact of input catalogue incompleteness on stacking experiments.
- Study the HI content of distant galaxies in LADUMA synthetic spectral cubes and examine how the confusion rate affects the results obtained from stacking experiments.
- Test and demonstrate that some of the existing methods used in publications to correct for contamination rates in stacking experiments are inadequate.

This thesis is structured as follows: chapter 2 provides an overview of the synthetic data products utilized in this study. Chapter 3 describes the HI stacking methods implemented in this work. In chapter 4 we present and discuss the results produced in this work. Chapter 5 investigates the HI content of distant galaxies in different environments. Chapter 6 outlines how some of the confusion correcting methods are inadequate. Chapter 7 measures the HI scaling relations for the LADUMA synthetic spectral cubes. Finally, chapter 8 provides a summary of this work together with the future outlooks on HI stacking experiments. Throughout this study, a Λ CDM cosmology is assumed with a Hubble constant of $H_0 = 67.3 \text{ km.s}^{-1}\text{Mpc}^{-1}$, $\Omega_\Lambda = 0.685$ and $\Omega_M = 0.315$.



2 Simulated data products

This chapter describes the simulated data products used in this study. The synthetic data products we use in this work are based on the [Obreschkow & Meyer \(2014\)](#) catalogue. The catalogue is a lite version of the Semi-Analytic Suite of the SKA Simulated Skies ($S^3 - SAX$), delivered as part of the European SKA design studies (SKADS) in 2009 and it relies on the physical models described in [Obreschkow et al. \(2009a,b,c\)](#).

Section 2.1 describes the methods implemented by [Obreschkow et al. \(2009a,b,c\)](#) to produce the physical models that are further used by [Obreschkow & Meyer \(2014\)](#) to evaluate the additional galaxy properties in the catalogue. Section 2.2 follows by explaining the techniques used by [Elson et al. \(2016\)](#) to convert the evaluated galaxy properties into the HI distributions. Finally section 2.3 presents the LADUMA HI data cubes we use in this work, it further presents various intrinsic and observable properties of the galaxies in the cubes.

2.1 Semi-analytic models

In order to create the data products we use in this work [Elson et al. \(2016\)](#) convert the evaluated galaxy properties from [Obreschkow & Meyer \(2014\)](#), which relies on the physical models described in [Obreschkow et al. \(2009a,b,c\)](#). [Obreschkow et al. \(2009a\)](#) present a simulation of evaluated neutral atomic hydrogen (HI) and molecular hydrogen (H_2) properties for approximately 3×10^7 galaxies acquired from the [De Lucia & Blaizot \(2007\)](#) catalogue, which is based on the Millennium Simulation of cosmic structure. Their techniques use a set of physical prescriptions to allocate HI and H_2 to their galaxy sample based on their total cold gas masses and other galaxy properties. In these physical prescriptions each galaxy is represented by a schematic model object with various other global properties.

They use equation 2.1 the parametrization for the azimuthally averaged radial distribution of HI mass to derive equation 2.2 and 2.3 known as the equations of atomic and molecular surface density profiles, respectively:

$$\frac{\Sigma_{H_2}(r)}{\Sigma_{HI}(r)} \equiv R_{mol}(r) = R_{mol}^c \exp(-1.6r/r_{disk}), \quad (2.1)$$

$$\Sigma_{\text{HI}}(r) = \frac{\tilde{\Sigma}_{\text{H}} \exp(-r/r_{\text{disk}})}{1 + R_{\text{mol}}^c \exp(-1.6r/r_{\text{disk}})}, \quad (2.2)$$

$$\Sigma_{\text{H}_2}(r) = \frac{\tilde{\Sigma}_{\text{H}} R_{\text{mol}}^c \exp(-2.6r/r_{\text{disk}})}{1 + R_{\text{mol}}^c \exp(-1.6r/r_{\text{disk}})}. \quad (2.3)$$

Above R_{mol}^c is the H_2/HI ratio at the center of the galaxy, r_{disk} is the exponential scale length of the disc and $\tilde{\Sigma}_{\text{H}}$ is the normalizing factor. They checked the profiles against the observed HI- and H_2 -density profiles of the nearby galaxies studied in Leroy et al. (2008). They tested the limitations implied by their assumption (i.e. $r_{\text{disk}} \approx 2\tilde{r}_{\text{disk}}$, where \tilde{r}_{disk} is the scale length of the stellar disc that is 30-50 percent smaller than r_{disk}) using two nearby galaxies namely NGC3184 with $r_{\text{disk}} \approx 2\tilde{r}_{\text{disk}}$ and NGC5505 with $r_{\text{disk}} \approx \tilde{r}_{\text{disk}}$. To compute the atomic and molecular surface density profiles for the two galaxies they require a disk stellar mass (M_{\star}^{disk}), the cold gas mass (M_{gas}), the exponential scale length of the disc r_{disk} and the dispersion parameter $\langle f_{\sigma} \rangle$. They determined M_{\star}^{disk} , M_{gas} and r_{disk} by fitting an exponential profile to the stellar component and used the value that is determined in Elmegreen (1993) for $\langle f_{\sigma} \rangle$, which is $\langle f_{\sigma} \rangle = 0.4$ for nearby galaxies. Figure 2.1 represents the results of the $\Sigma_{\text{HI}}(r)$ and $\Sigma_{\text{H}_2}(r)$. The results they obtained match the empirical data, therefore this is one of the proofs that their catalogue is realistic.

One of the most important relations is the measure of the mass function, Obreschkow et al. (2009a) managed to accurately measure the HI- and H_2 -mass functions for their simulation. The mass function is defined as the number density of galaxies as a function of their gas content and it is important to accurately measure this quantity because it is a critical parameter in models of galaxy evolution. Figure 2.2 represent the mass function obtained by Obreschkow et al. (2009a) and it is compared to the empirical HI mass function obtained from Zwaan et al. (2005) and H_2 mass function obtained from Obreschkow (2009), whereby they both use nearby galaxies. The mass functions obtained in Obreschkow et al. (2009a) are realistic because they match nearby galaxy mass functions.

Another important work that Obreschkow et al. (2009a) did was to study the relationship between HI mass, M_{HI} , and HI radius, r_{HI} , ($M_{\text{HI}} - r_{\text{HI}}$ relation) at redshift $z = 0$. The $M_{\text{HI}} - r_{\text{HI}}$ relation is important because it can be used to estimate the distribution of the bright, well-resolved galaxies in the existing nearby galaxy

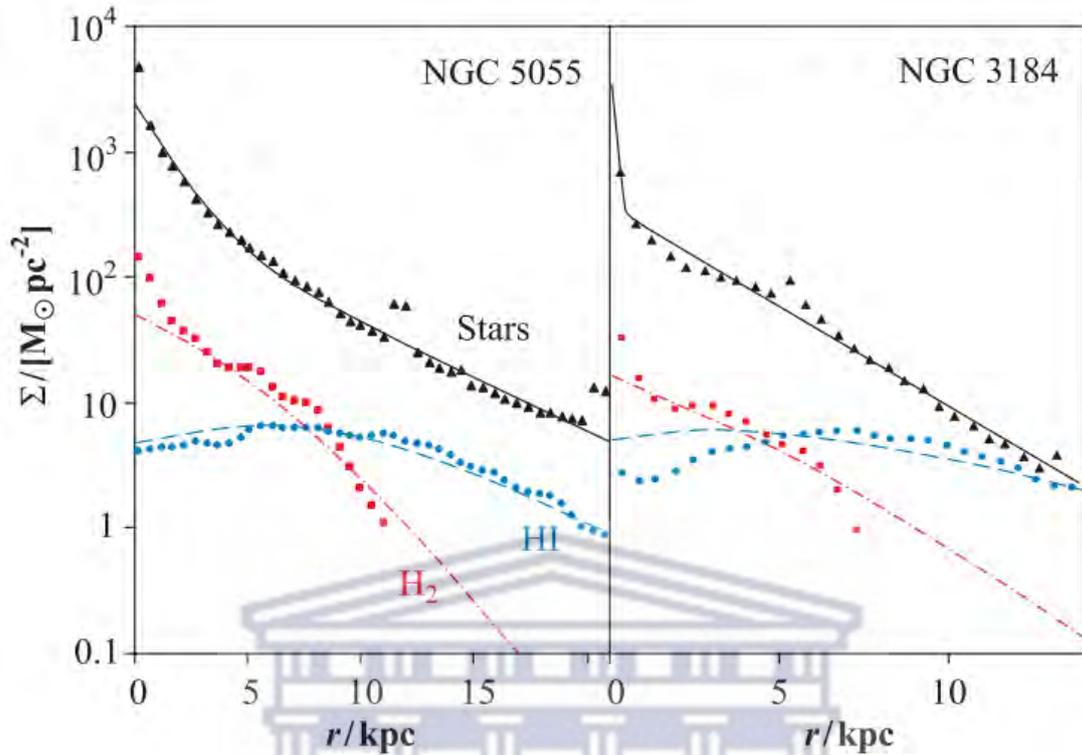


Figure 2.1: The panel on the left and right shows the density profiles for nearby galaxy (NGC5055) and nearby galaxy (NGC3184), respectively. The solid black, dashed blue and red dashed lines in both panels, respectively, represent density profiles for stars, Σ_{HI} and Σ_{H_2} obtained by Obreschkow et al. (2009a) using equation 2.2 and 2.3. The black triangles, red squares and blue dots, respectively, show results obtained by Leroy et al. (2008) for stars, Σ_{HI} and Σ_{H_2} . The results obtained from Obreschkow et al. (2009a) compare well with the results obtained from Leroy et al. (2008).

catalogues. This relation can also be used to predict the HI diameters of unresolved galaxies based on their HI masses, that will be detected by forthcoming experiments. Figure 2.3 shows the $M_{\text{HI}} - r_{\text{HI}}$ relation obtained by Obreschkow et al. (2009a) at redshift $z = 0$. Their results match the results obtained by Verheijen (2001).

The final critical results Obreschkow et al. (2009a) obtained is the Tully-fisher relation (Tully & Fisher, 1977). Tully-fisher relation is defined as the empirical relation between the mass of the spiral galaxy and emission line width. This relation can be used to estimate the distance to the disk galaxies. Obreschkow et al. (2009a) study the Tully-Fisher relation for 1000 randomly selected galaxies from their catalogue at redshift $z = 0$. Figure 2.4 shows the Tully-Fisher relation they obtained based on the galaxy stellar mass M_{\star} and gas mass M_{gas} versus the width of the HI emission

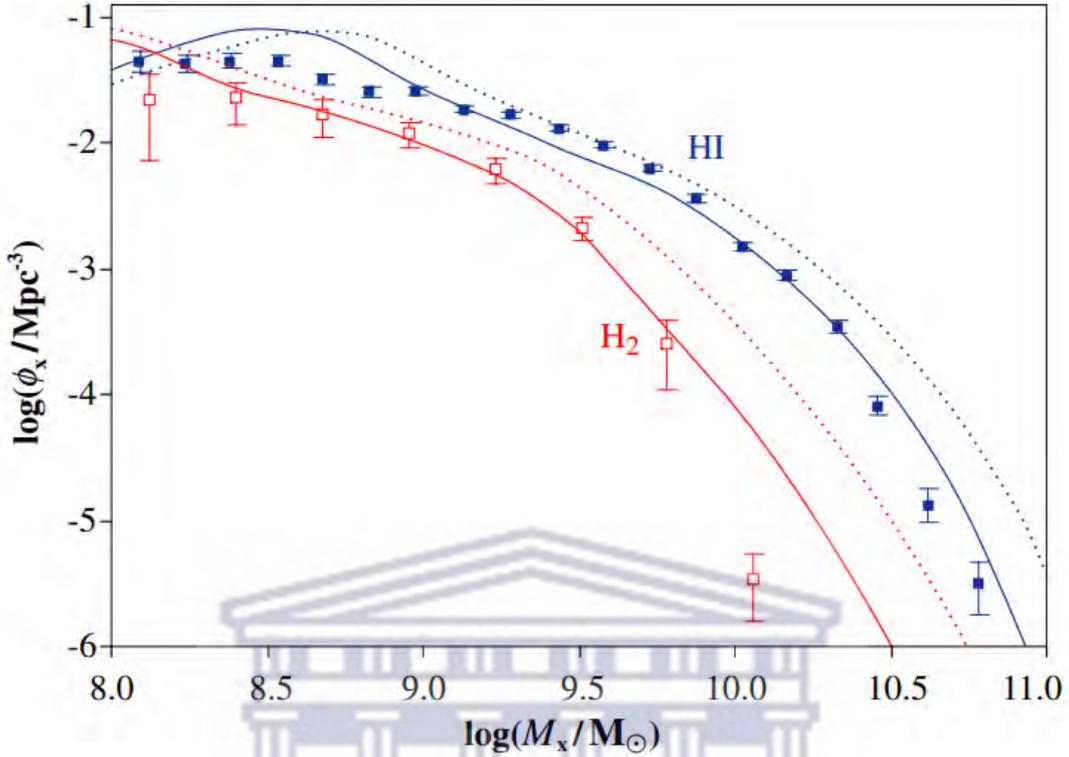


Figure 2.2: The mass function of the simulated galaxies. The solid lines show mass function with HI and H₂, and the dotted lines show the mass function without the HI and H₂ obtained from Obreschkow et al. (2009a). The blue squares with error-bars represent the empirical mass functions from Zwaan et al. (2005), while the open red squares with error bars represent mass functions from Obreschkow (2009). The mass functions obtained by Obreschkow et al. (2009a) match the empirical results.

line measured at 50 and 20 percent of the peak flux density W_{HI}^{50} , W_{HI}^{20} , respectively. To verify that the results they obtained are realistic they compared them to the results obtained by McGaugh et al. (2000).

Because the results obtained by Obreschkow et al. (2009a) are shown to compare well with the observations and their simulations can be trusted to be reliable, further studies can be carried out using their simulations. From the physical models described in Obreschkow et al. (2009a,b,c) that appear to be realistic when they are compared to the local universe studies, Obreschkow & Meyer (2014) evaluate additional galaxy properties of the catalogue. This catalogue has a cone shape with cosmic volume that spans a sky area of 100 deg² and a redshift range $z = 0 - 1.2$. The catalogue is truncated to 21 cm peak flux density above 1 μ Jy in order for its size to be logical. Within this flux limit, the model is complete for HI mass above

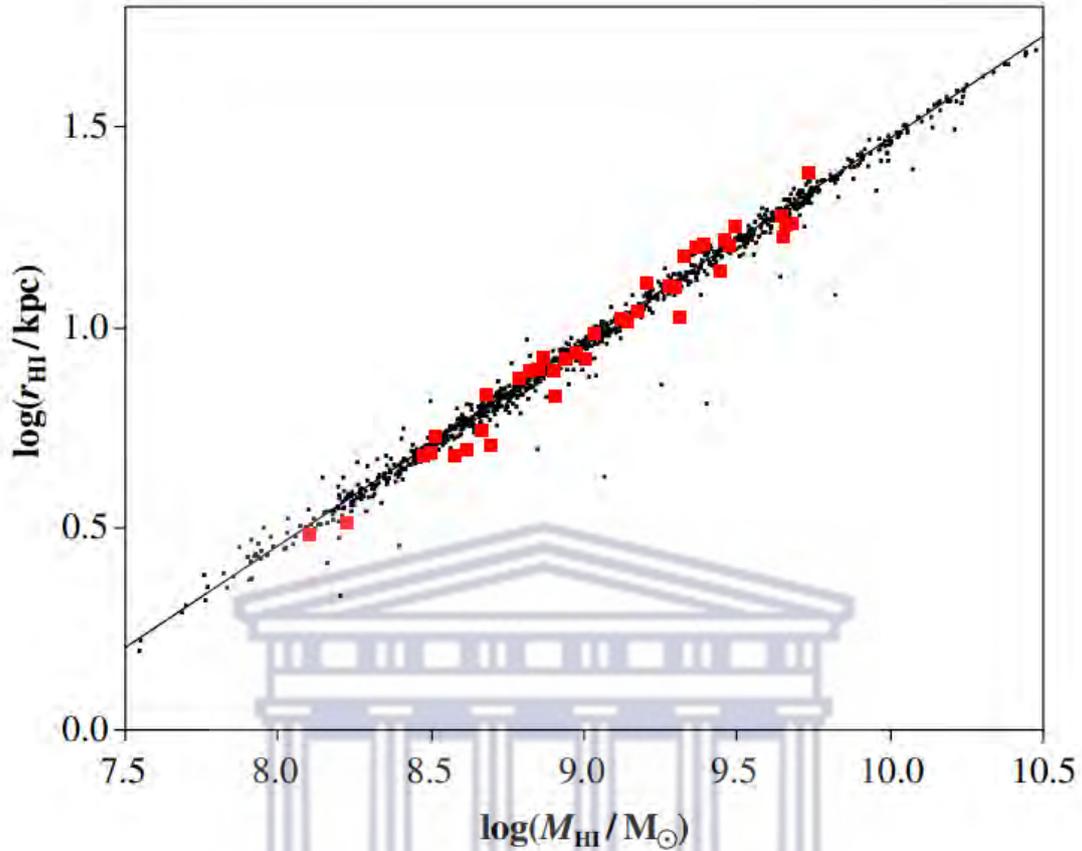


Figure 2.3: The $M_{\text{HI}} - r_{\text{HI}}$ relation for 1000 randomly selected galaxies from Obreschkow et al. (2009a) simulated galaxies at redshift $z = 0$. The black dots represent the 1000 galaxies, while the red squares represent Verheijen (2001) galaxies, and the black line is the fit line.

$10^8 M_{\odot}$ (Obreschkow & Meyer, 2014).

The difference between the Obreschkow & Meyer (2014) catalogue and the $S^3 - \text{SAX}$ is that it includes more optical galaxy properties. The properties are derived from the extinction corrected absolute magnitudes that are simulated using a variable K-correction. Each galaxy in this catalogue contains 20 properties, which are shown in table 2.1.

2.2 HI data cubes of galaxies

Knowing the fact that Obreschkow et al. (2009a,b,c) physical models can accurately reproduce the properties of known nearby galaxies, Obreschkow & Meyer (2014)

Col	Symbol	Unit	Description
1	ID	-	Unique galaxy identifier in the Munich Semi-Analytic Model “DeLucia2006a”
2	RA	deg	Right ascension of galaxy centre
3	Dec	deg	Declination of galaxy centre
4	z	-	Apparent redshift of galaxy centre, including the Doppler component due to peculiar motion relative to the Hubble expansion
5	i	deg	Galaxy inclination defined as the smaller angle ($0^\circ - 90^\circ$) between the line-of-sight and the rotational axis of the galaxy
6	T	-	Numerical Hubble type (99 morphologically unresolved objects)
7	M_\star	M_\odot	Stellar mass
8	M_{HI}	M_\odot	Mass of neutral atomic hydrogen HI, without helium
9	M_{H_2}	M_\odot	Mass of molecular hydrogen H ₂ , without helium
10	$S_{\text{HI}}^{\text{int}}$	Jy km.s^{-1}	Velocity-integrated flux of the redshifted 21 cm HI emission line, with velocity units defined in the galaxy rest-frame
11	$S_{\text{HI}}^{\text{peak}}$	Jy	Peak flux density of the HI emission line; typically the flux density of the ‘horns’
12	$S_{\text{CO}}^{\text{int}}$	Jy km.s^{-1}	Velocity-integrated flux of the redshifted 115.27 GHz ¹² CO(1 – 0) emission line, with velocity units defined in the galaxy rest-frame
13	$S_{\text{CO}}^{\text{peak}}$	Jy	Peak flux density of the ¹² CO(J = 1 – 0) emission line; typically the flux density of the ‘horns’
14	W_{HI}^{50}	km.s^{-1}	Width of the HI emission line, in galaxy rest-frame velocity units, measured at 50 percent of the peak flux density
15	W_{HI}^{20}	km.s^{-1}	Width of the HI emission line, in galaxy rest-frame velocity units, measured at 20 percent of the peak flux density
16	$r_{\text{HI}}^{\text{edge}}$	arcsec	Apparent HI radius along the major axis out to a HI disk surface density of $1 M_\odot \text{ Mpc}^2$, corresponding to a face-on column density of $1.25 \times 10^{20} \text{ cm}^{-2}$
17	$r_{\text{HI}}^{\text{half}}$	arcsec	Apparent HI half-mass radius along the major axis
18	M_R	mag	Absolute Vega R-band magnitude, corrected for intrinsic dust extinction; 99 if stellar mass and star formation history are insufficiently resolved to compute M_R
19	m_R	mag	Apparent Vega R-band magnitude; value 99 if no absolute magnitudes available
20	r_e	arcsec	Effective radius, here approximated as the radius containing half the stellar mass if the galaxy were viewed face-on

Table 2.1: The 20 properties for each galaxy in the [Obreschkow & Meyer \(2014\)](#) catalogue. The columns give the description of the mock catalogue in ASCII format. A subset of this information for each galaxy was used by [Elson et al. \(2016\)](#) to model the HI line emission. <http://etd.uwc.ac.za/>

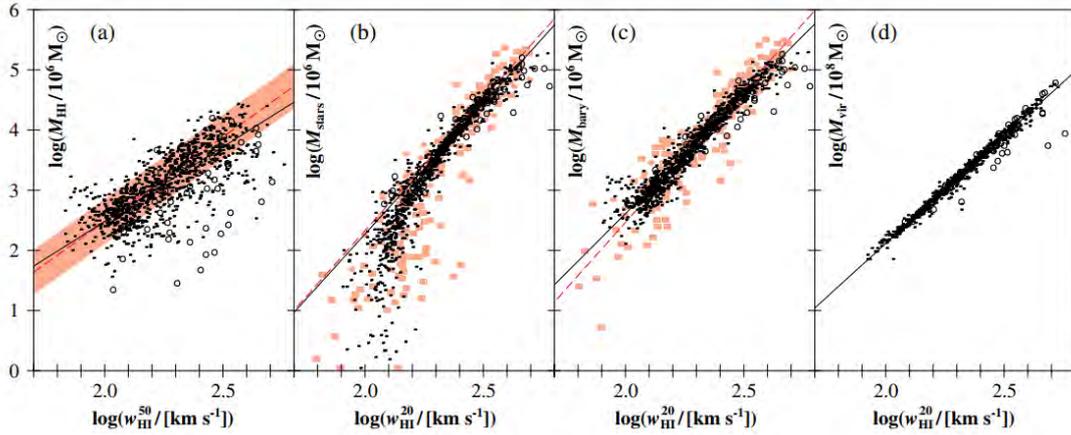


Figure 2.4: The Tully-Fisher relation from Obreschkow et al. (2009a). The black dots and circles represent simulated spiral and elliptical galaxies. Solid lines represent power-law fits to the simulated disk galaxies. Panel a) shows Tully-Fisher relation between M_{HI} and W_{HI}^{50} . The red dashed line and the red shaded zone represent Obreschkow et al. (2009a) observational determination and the 1σ scatter of the HI- Tully-Fisher relation from the HIPASS data. Panel b) shows the simulated stellar-mass Tully-Fisher relation with the observational data from McGaugh et al. (2000). The red dots and dashed line represent observational data and the power-law regression from McGaugh et al. (2000) work. Panel c) shows the simulated baryonic Tully-Fisher relation and observational data from McGaugh et al. (2000). Panel d) shows the simulated total dynamical mass, considered as virial mass M_{vir} and the circular velocity, given by the line width W_{HI}^{20} .

further used these physical models to generate a mock galaxy catalogue with more galaxy properties. The additional properties helped Elson et al. (2016) to successfully convert the evaluated HI properties to the full three dimensional HI models we use in this work. In order for Elson et al. (2016) to properly generate the three dimensional HI models, they use a circular velocity profile and a parameterisation of the radial distribution of the mass for a galaxy. The following subsections present the techniques implemented by Elson et al. (2016) to model the spatial and spectral distributions of the three dimensional HI emission from the galaxies.

2.2.1 HI mass profiles

One of the fundamental requirements to generate a full three dimensional model of the spatial and spectral distributions of a galaxy's HI flux is a parameterisation of the radial distribution for the HI mass of a galaxy. In order to generate the HI mass

profile of a galaxy in our sample [Elson et al. \(2016\)](#) used equation 2.4. The radial profiles generated from equation 2.4 helps to model the central HI concentration and depressions of the galaxy:

$$\Sigma_{\text{HI}}(R) = \frac{A \exp(-R^2/2h^2)}{1 + \beta \exp(-1.6R^2/2h^2)}, \quad (2.4)$$

where A is the normalization parameter that is used to control the total HI mass. The numerator is a one-dimensional Gaussian, where h is the standard deviation. On the denominator the parameter β controls the central concentration of HI. β values that are greater than zero yield the central HI depressions. Equation 2.4 matches the HI distributions observed by [Bigiel & Blitz \(2012\)](#) in 33 nearby spiral galaxies, that are from HERACLES and HI data from THINGS.

[Elson et al. \(2016\)](#) constructed the HI mass profiles by converting the evaluated apparent HI half-mass radius along the major axis, $R_{\text{HI}}^{\text{half}}$, to exponential disc scale length, R_h , and then they set $h = R_h$ from equation 2.4. They further set $\beta = R_{\text{mol}}^c$, where R_{mol}^c is the H₂/HI ratio and then they modelled every galaxy out to a radius $R = 3.5h$. Figure 2.5 shows the HI mass profiles for two galaxies from [Obreschkow & Meyer \(2014\)](#) catalogue produced by [Elson et al. \(2016\)](#).

2.2.2 Rotation curves

[Elson et al. \(2016\)](#) used template rotation curves for disc galaxies presented in [Catinella et al. \(2006\)](#) to model rotation curves. [Catinella et al. \(2006\)](#) used a homogeneous sample of ~ 2200 nearby galaxies with high sensitivity long-slit optical spectroscopy and detailed I-band photometry. Their template rotation curves are separated in luminosity classes and span six magnitudes in I-band luminosity. They used the Polyex model ([Giovanelli & Haynes, 2002](#)) presented by equation 2.5 to fit their template rotation curves:

$$V_{\text{PE}} = V_0 \left(1 - e^{-R/R_{\text{PE}}}\right) \left(1 + \frac{\alpha R}{R_{\text{PE}}}\right), \quad (2.5)$$

where V_0 , R_{PE} , α , respectively, determine the amplitude, the exponential scale of the inner region, and the slope of the outer part of the rotation curve. [Elson et al.](#)

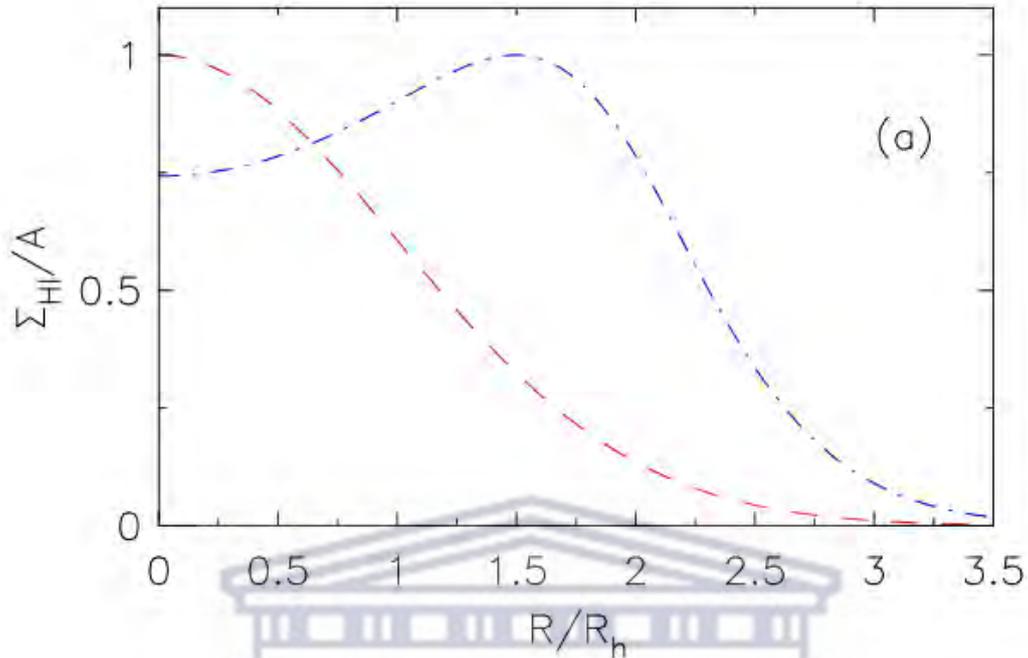


Figure 2.5: The HI mass profiles from Elson et al. (2016) for parameter pairs $h, \beta = (R_h, 0.0), (R_h, 10.0)$, respectively, shown as red dashed and blue dashed, and they have been scaled to have the highest value of 1. Catinella et al. (2006) generated the HI mass profiles for every galaxy in the catalogue and then used them in the process of locating the galaxies in the cube.

(2016) used the Polyex model for several luminosity classes presented in Catinella et al. (2006) to construct rotation curves for galaxies based on their evaluated R-band absolute magnitudes. The evaluated R-band absolute magnitudes were converted to I-band magnitudes. Figure 2.6 represents the Polyex model fits produced from various luminosity classes from Catinella et al. (2006).

2.2.3 Three-dimensional modelling

In order to properly locate a galaxy within the HI data cube based on equation 2.4 the parameterisation for radial distribution and equation 2.5 the Polyex model. Elson et al. (2016) used their path, which assumes axisymmetry and models a galaxy as a collection of many HI clouds. This approach requires that the rotation curve of each galaxy must be provided with the inclination and position angle of the HI disc.

The approach also requires the HI velocity dispersion and the HI mass profile as

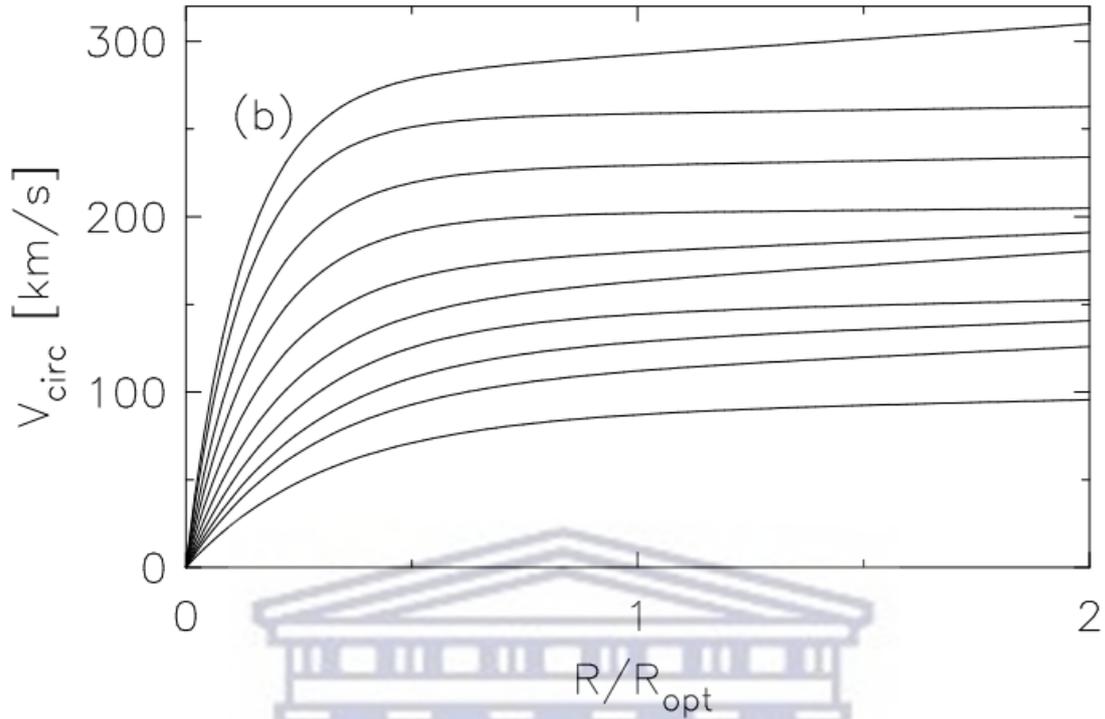


Figure 2.6: The Polyex model fits to rotation curves plotted against the optical radius constructed by Elson et al. (2016). The mean I-band absolute magnitudes of different models presented here can be found in Catinella et al. (2006); Elson et al. (2016). Catinella et al. (2006) generated the circular velocity for every galaxy in the catalogue and then used them to locate the galaxies in the cube.

function of radius. They further assumed an infinitely thin HI disc, and then generated random numbers presenting the spatial and spectral coordinates of individual HI clouds. Then the approach used the position and velocity of the cloud to place them at the appropriate location within the data cube.

Figure 2.7 shows channel maps for a single galaxy model. This galaxy has a unique identifier ID = 375700. The model is not flux-calibrated, colours in the maps represent the relative distribution of HI flux in a particular channel. Galaxy 375700 has an HI mass of $M_{\text{HI}} = 8.18 \times 10^7 M_{\odot}$, velocity-integrated flux of redshift 21 cm HI emission line $S_{\text{HI}}^{\text{int}} = 3.96 \times 10^{-9} \text{ Jy km.s}^{-1}$ and HI disc that is inclined at 63.8 deg.

The three dimensional model for each galaxy in the sample is generated and then all the models are brought together into a single large HI data cube. The models are realistically placed in locations provided in the Obreschkow & Meyer (2014) catalogue, which means all the galaxies in the cube are correctly spatially distributed.

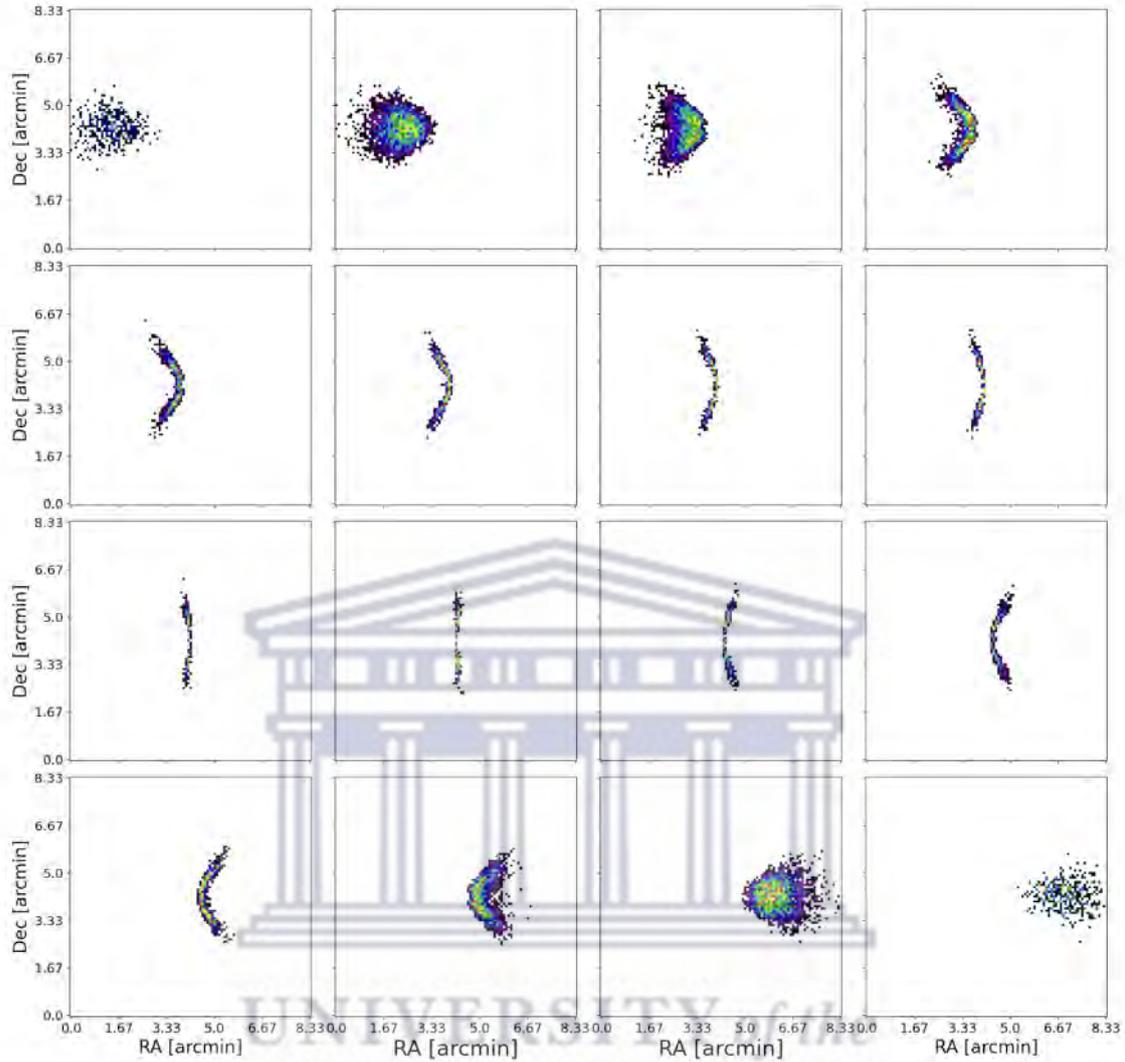


Figure 2.7: The channel maps of a single galaxy with ID = 375700. The channels in the cube are separated by 5 km. s^{-1} . This galaxy has the spatial dimensions of 101-by-101 pixels and it spans 32 channels. For this model we show every second channel of the galaxy.

Figure 2.8 shows HI data products produced from the galaxy model shown in figure 2.7. Panel a) shows the total intensity map (i.e. moment0 map). Panel b) shows intensity-weighted mean velocity field (i.e. moment1 map). This map is used to compute average velocity of the field. Panel c) shows the position-velocity slice taken along the major axis of the galaxy and panel d) shows the global profile of this galaxy. The global profile is generated by taking the sum of the HI emission in each channel of the data cube and plotting it against the velocity of the galaxy in each channel.

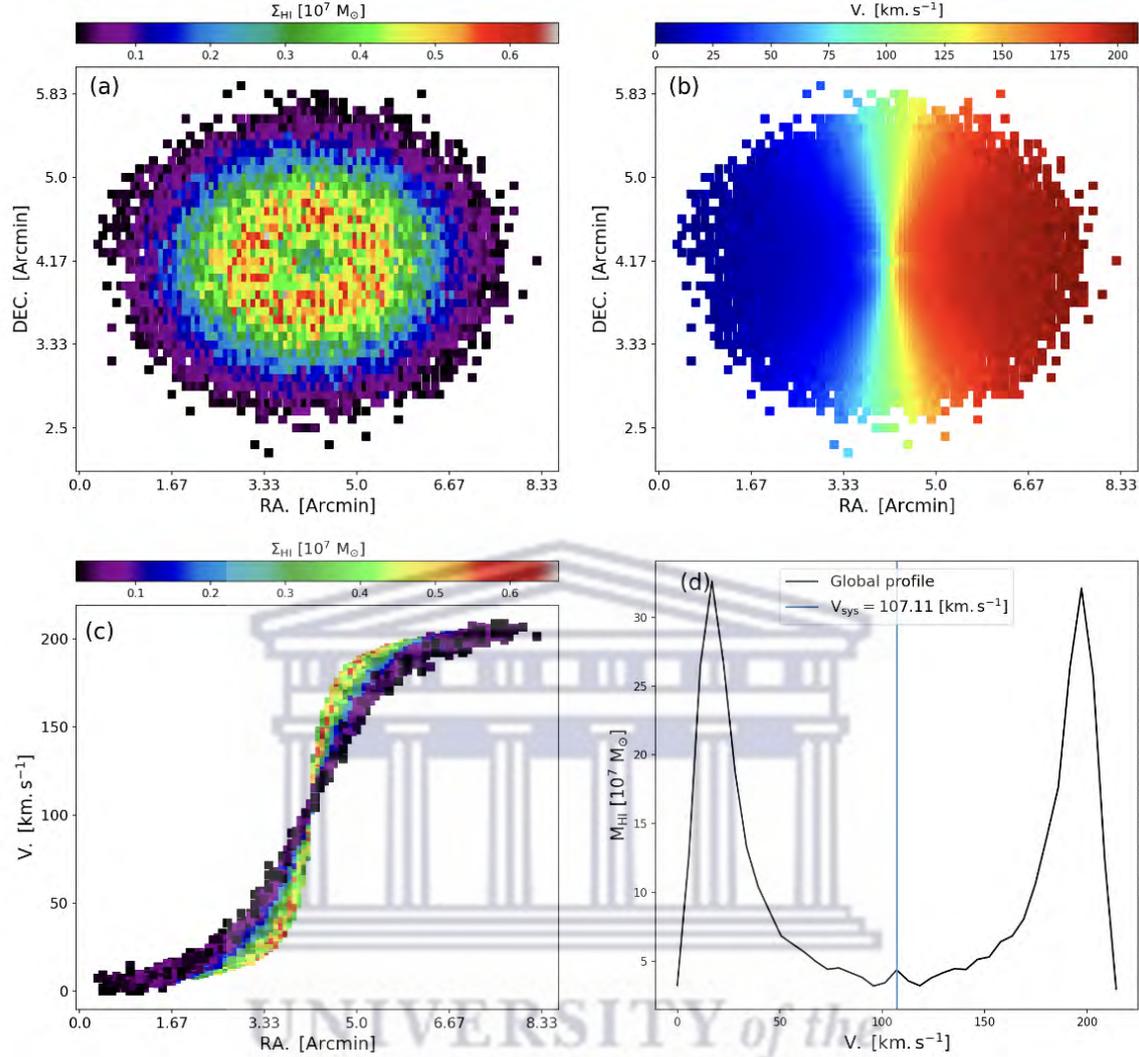


Figure 2.8: The HI data products produced from the galaxy model shown in figure 2.7. Panel (a) shows the HI total surface density. Panel (b) shows moment 1 map (i.e. intensity-weighted mean velocity field). Panel (c) shows position-velocity slice taken along the major axis of the galaxy. Panel (d) shows the global profile showing the total HI mass in each channel. The vertical blue line shows the systemic velocity of this galaxy.

2.3 LADUMA HI data cube

In this work we use the cubes from [Elson et al. \(2016\)](#). We choose to focus on these cubes because they best represent a high redshift z portion of the LADUMA cubes. LADUMA will conduct HI stacking experiments by using data cubes with a frequency-dependent spatial resolution of $\sim 10 - 20$ arcsec ([Elson et al., 2019b](#)). Here we use a synthetic data cube that spans a sky area of $1^\circ \times 1^\circ$ and a redshift range $0.7 < z < 0.758$. This cube is made-up of 26511 galaxies with a total HI mass

of $1.714 \times 10^{13} M_{\odot}$. Table 2.2 present the summary of user specified inputs for our cubes.

Item	Inputs
RA range	1 deg.
Dec range	1 deg.
z range	0.7 - 0.758
Cosmology	$h = 0.7, \Omega_m = 0.315, \Omega_{\Lambda} = 0.685$
Channel width	26 kHz
pixel scale	5 arcsec
BMAJ	15 arcsec
BMIN	15 arcsec
RMS noise scale	2.88×10^{-5} Jy/beam

Table 2.2: Table summarising user specified inputs for LADUMA synthetic data cubes used in this work.

Figure 2.9 shows the channel maps of the LADUMA data cube used in this work. The channel maps are from the noise-free version of our cubes that was smoothed using a Gaussian point-spread-function. Each panel shows the sum of the 53 adjacent channels in the cube. The top right corner shows the frequency of the central channel in the 53 channels in units of Hz. Each galaxy in this cube has had its HI flux modelled in the ways described in section 2.1 and 2.2. The channel maps confirms that our cube is made-up of clusters of galaxies, from small galaxies to large galaxies.

Figure 2.10 shows an HI total intensity map produced from this cube. The total intensity map shows that our cube is made-up of a lot of galaxies that are modelled and then placed into the HI data cube using the positions provided in the Obreschkow & Meyer (2014) catalogue. The parts of the map that have high flux are highly clustered and most of them are concentrated with a single large galaxy and several small galaxies behind them.

2.3.1 Intrinsic and observable properties of the LADUMA cubes.

In this subsection we focus on the different intrinsic and observable properties of the galaxies in our cube. In figure 2.9 we have shown the channel maps of the full-size HI data cube and in figure 2.10 we have shown the HI total intensity map. Figure 2.11 shows some of the galaxy properties of our cube. Panel a) shows the right

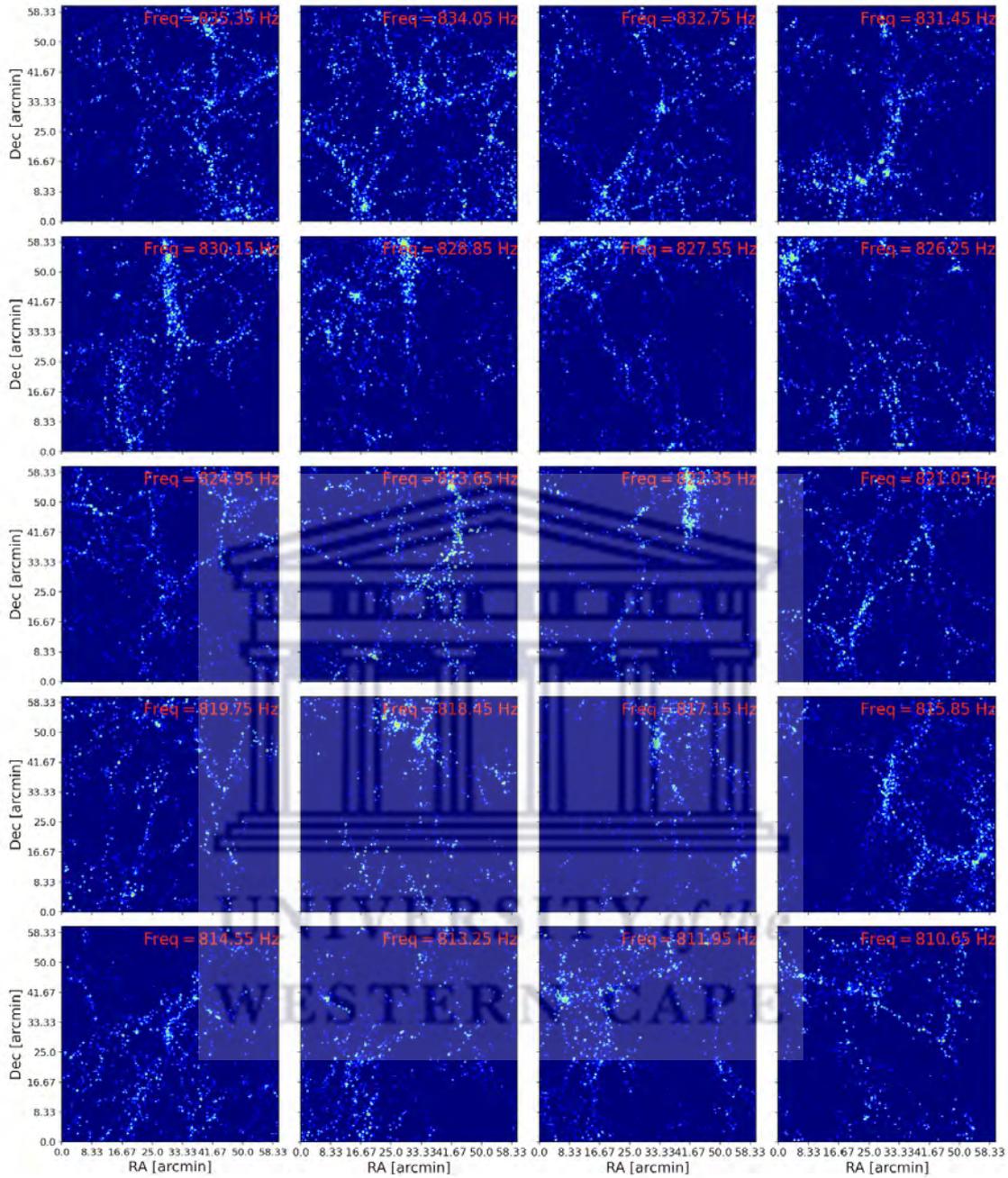


Figure 2.9: The channel maps showing HI line emission (noise-free) of a cube spanning a sky area of $1^\circ \times 1^\circ$, a redshift range $0.7 < z < 0.758$ and it contains 26511 galaxies. Each panel here shows the sum of the 53 adjacent channels in the cube. The top right corner shows the frequency of the central channel in the 53 channels in units of Hz. Each galaxy shows the frequency of the central channel in the 53 channels in units of Hz. Each galaxy in this cube has had its HI flux modeled in the ways described in section 2.1 and 2.2.

ascension and declination position of the 4271 galaxies from the 26511 galaxies of the cube. This clearly shows that our cube is made-up of clusters of galaxies and there are some areas that are highly clustered compared to the others.

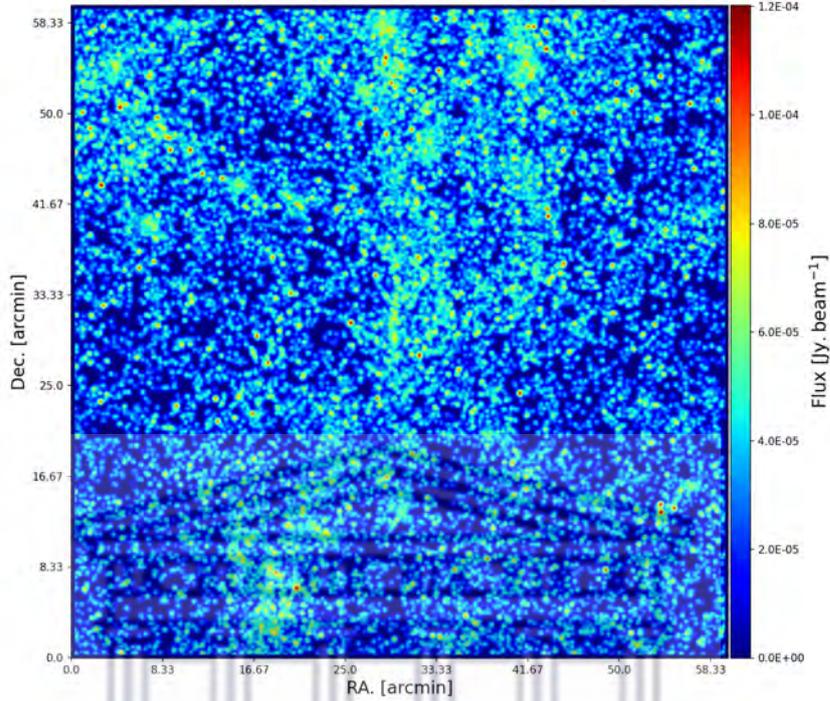


Figure 2.10: An HI total intensity map for the full-size synthetic cube. This map covers a sky area of $1^\circ \times 1^\circ$ and it contains the HI line emission from 26511 galaxies.

Panel b) shows the distribution of the galaxy's stellar masses, M_* . At $M_* \sim 10^{8.5} M_\odot$ the distribution become discontinuous, this behaviour is due to the fact that the $S^3 - SAX$ catalog is incomplete at $M_* \lesssim 10^{8.5} M_\odot$. Panel c) shows the distribution of the HI mass. A lot of galaxies have an HI mass range $10^{7.5} \lesssim M_{\text{HI}} [M_\odot] \lesssim 10^{9.5}$. Panel d) shows the HI gas fraction distribution. Large galaxies have small HI gas fraction because they are old and their HI mass is almost depleted. Panel e) shows luminosity distance distribution, and this distribution is almost uniform most likely because at our redshift range the volume of the universe probed increases with redshift. As redshift increases, we are sampling a larger volume of the universe, leading to a more uniform distribution of luminosity distances. This is because the volume element increases with distance, causing an equal probability of galaxies being found at different distances within the redshift range.

Panel f) shows distribution of the inclination angle of the galaxies in the sample. The distribution is non-uniform due to galactic morphology. Different types of galaxies exhibit different inclination angle distributions. Spiral galaxies, for instance, often

have a wider range of inclination angles compared to elliptical galaxies. This is because spiral galaxies have a flattened disk structure, and their inclination angles can span from face-on (low inclination) to edge-on (high inclination). On the other hand, elliptical galaxies, which are more spheroidal in shape, tend to have a narrower distribution of inclination angles. Panel g) shows the distribution of the HI diameter. Panel h) shows the relation between HI gas fraction and the galaxy's stellar mass. The galaxies with small stellar mass have large HI gas fraction because they are young and active. Large galaxies have small HI gas fraction because they are old and passive. Panel i) shows that the HI gas fraction increases with the HI mass. And panel j) shows the relation between the HI mass and the galaxy's stellar mass. Overall figure 2.11 proves that the synthetic data cubes used in this work are reliable and they contain diversity of galaxies with various reliable properties.

Now that we have shown some of the properties of the galaxies in our cube, we also look at the scaling relation generated from the galaxy properties. We found it important to look at several scaling relations generated between galaxy properties in the cubes, because they verify that the cube we are using in this work is realistic as Obreschkow et al. (2009a) did to check if their models are realistic. Figure 2.12 shows some of the important scaling relations we generate in this work using LADUMA synthetic data cube.

Panel (a) shows a relationship between main-sequence galaxy stellar mass and absolute magnitude of the galaxies. This shows that massive galaxies are much brighter compared to the small galaxies. Panel (b) shows the HI size-mass ($D_{\text{HI}} - M_{\text{HI}}$) relation of the galaxies in our sample. This relation is extraordinarily correlated due to the fact that late type galaxies have a homogeneous radial profile of HI surface density in the outer regions when the radius is normalized by D_{HI} (Wang et al., 2016). We compared our results to Wang et al. (2016) and Broeils & Rhee (1997) where they, respectively, used 500 nearby galaxies and 108 nearby galaxies. The green line represents the results obtained by Wang et al. (2016), the yellow line represents results obtained by Broeils & Rhee (1997) and the red line are the results we obtained in this work.

Panel (c) shows the Tully-Fisher relation (Tully & Fisher, 1977), that is computed by scaling W_{HI}^{50} by $\sin(i)$, where i is the inclination angle of the galaxy. Panel (d) and (e) show the baryonic Tully-Fisher relation for galaxies of our sample. In order to calculate the baryonic mass M_{bar} we use equation 2.6 adopted from Trachternach

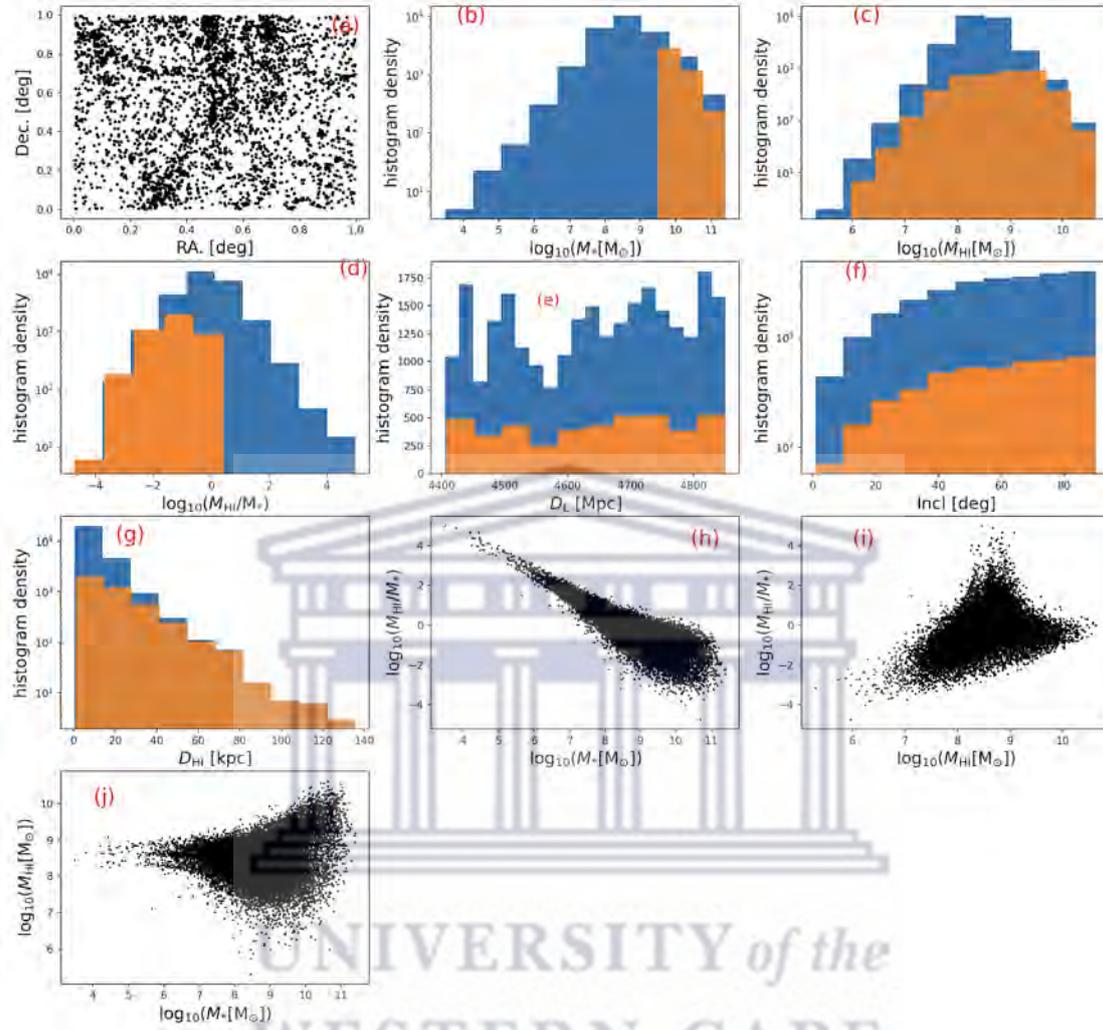


Figure 2.11: The evaluated galaxy properties for LADUMA HI data cube. The blue histogram correspond to the full sample of the galaxies in the synthetic data cube, while the orange histogram correspond to the results for a sample of galaxies having stellar masses $M_{\star} > 10^{9.5} M_{\odot}$. Panel (a) shows the right ascension and declination position of 4271 galaxies from the 26511 galaxies of the cube. Panel (b) shows distribution of the galaxy's stellar mass. Panel (c) shows distribution of an HI mass. The total HI mass of the full sample is $1.714 \times 10^{13} M_{\odot}$. Panel (d) shows HI gas fraction distribution. Panel (e) shows luminosity distance distribution. Panel (f) shows distribution of galaxy inclination angle. Panel (g) shows HI diameter distribution. Panel (h) shows the relation between HI-stellar mass ratio and galaxy stellar mass, it indicates that small galaxies are gas rich while large galaxies are gas poor. Panel (i) shows the relation between HI-stellar mass ratio and HI mass. Panel (j) shows the relation between galaxy HI mass and galaxy stellar mass.

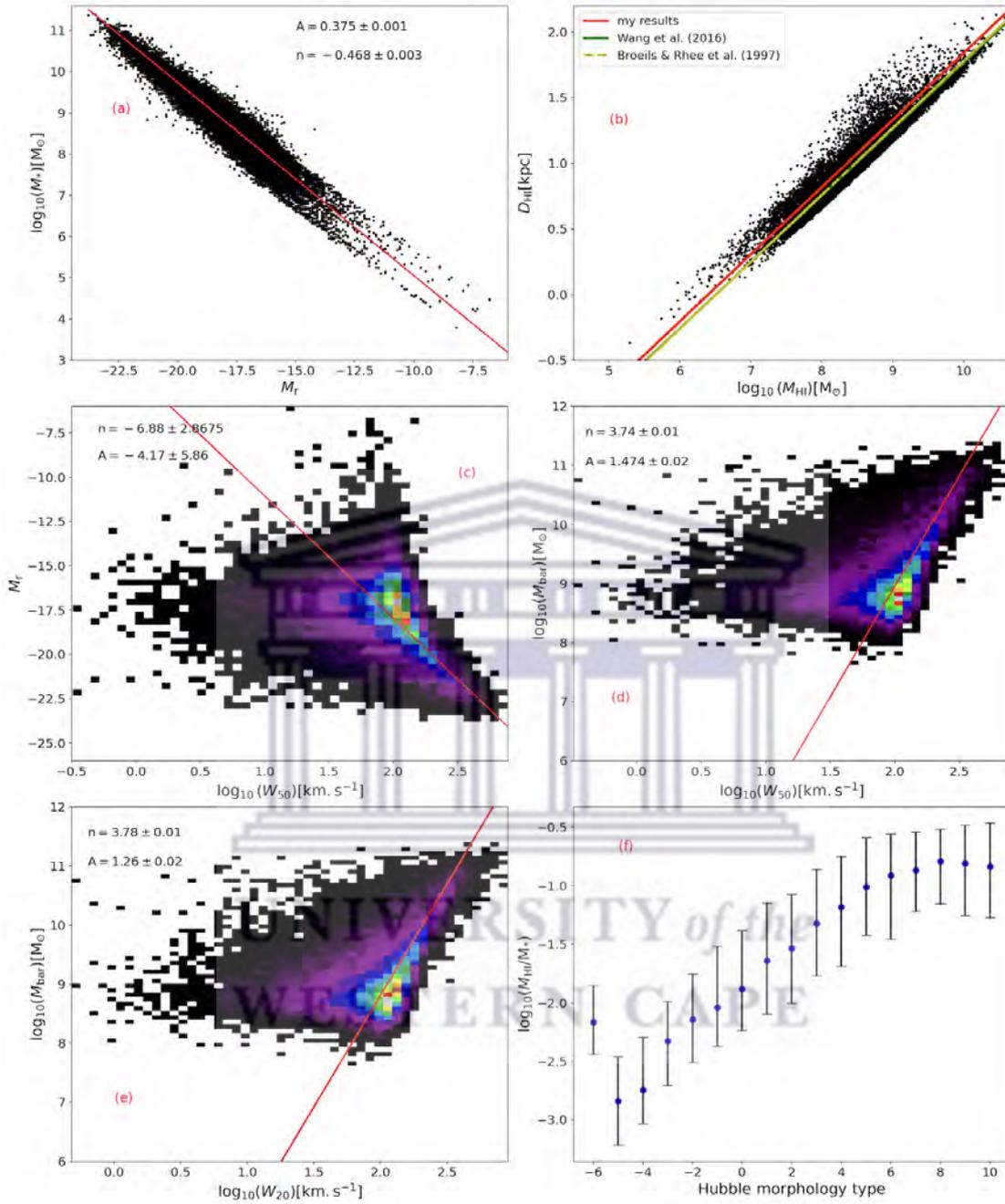


Figure 2.12: Shows the scaling relations of our galaxy sample. Panel a) shows the relation between absolute magnitude and galaxy’s stellar mass. Panel b) shows HI size-mass ($D_{\text{HI}} - M_{\text{HI}}$) relation. The red line represents our best fit line, the green line represents Wang et al. (2016) best fit line and the yellow line represents Broeils & Rhee (1997) best fit line. Panel c) shows the Tully-Fisher relation. Panel d) and e) shows the baryonic Tully-Fisher relation for galaxies of our sample. Panel (f) shows the relation between Hubble morphology type and an HI gas fraction. The error-bars represent the 25th, and 75th percentile and the blue dots represent the HI gas fraction median for each morphology type sub-sample.

et al. (2009):

$$M_{\text{bar}} = M_{\star} + M_{\text{gas}}, \quad (2.6)$$

where M_{\star} is the galaxy's stellar mass, M_{gas} is the mass of the galaxy gasses, which includes HI and H₂ mass. Again we scale the W_{HI}^{50} and W_{HI}^{20} by $\sin(i)$, where i is the inclination angle of the galaxy in order to obtain the line-of-sight of the HI line width of a galaxy. The baryonic Tully-Fisher relation have a lot of horizontal spread at $M_{\text{bar}} \approx 10^9 M_{\odot}$ because galaxies of this size have a large difference in the HI emission line width due the Black hole mass (Di Cintio et al., 2017). Panel (f) shows the relation between Hubble morphology type and the HI gas fraction. The HI gas fraction increases from early to late type galaxies, this is because early type old galaxies are passive and HI poor, while the young late type galaxies are active and HI rich. The error-bars represent the 25th, and 75th percentile and the blue dots represent the HI gas fraction median for each morphology type sub-sample. The first point in this panel represent Compact elliptical galaxies (cEs). These types of galaxies forms when early-type disk galaxy with a compact bulge loses a large fraction of its initial disk mass through dissipative tidal interactions with a more massive host galaxy and only the central bulge component survives (Kim et al., 2020; Bekki et al., 2001; Chilingarian et al., 2009). The compact elliptical galaxies are metal-rich and have low-mass, thus they have higher HI gas fraction compared to the other elliptical galaxies. In particular, all the relations produced in this figure are in agreement with the relations produced at redshift $z \sim 0$ in publications (for example: Trachternach et al., 2009; Broeils & Rhee, 1997; Wang et al., 2016). This further proves that the catalog we use in this work is realistic.

2.3.2 The angular 2 point correlation function

We investigated the clustering of the galaxies in our cube using 2 point correlation function. Figure 2.13 shows the angular 2 point correlation function. The angular 2 point correlation function is computed using equation 2.7. This equation is adopted from Chen & Schwarz (2016), it is the so called Landy-Szalay estimator (Landy & Szalay, 1993):

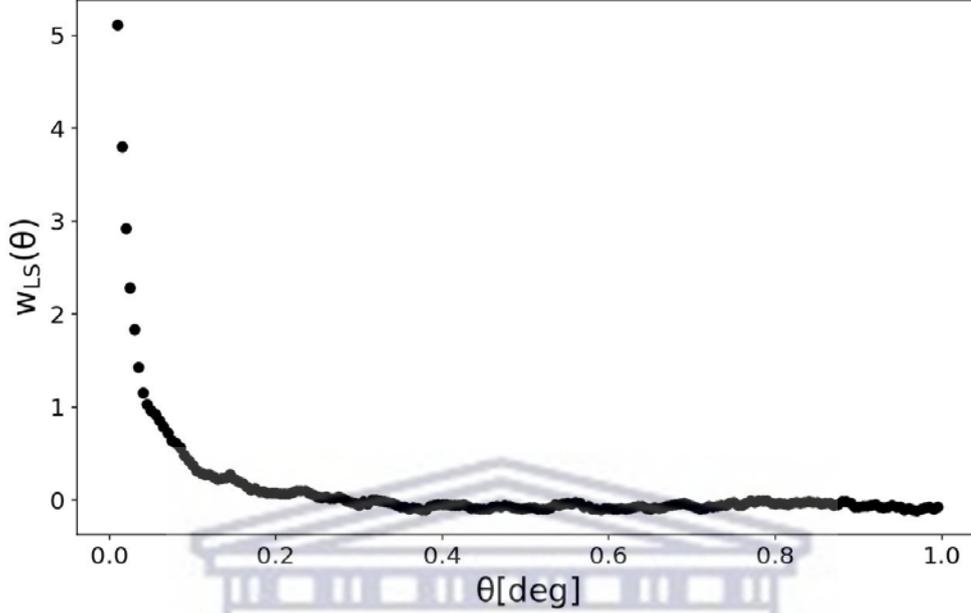


Figure 2.13: The angular 2 point correction function computed in terms of redshift z . It shows a clear excess of clustering at angular separation $\theta \lesssim 0.2$ deg. At higher values of angular separations $\theta \gtrsim 0.2$ deg the estimator approaches zero indicating that there is less clustering at those separations.

$$w_{\text{LS}}(\theta) = \frac{N_r(N_r + 1)}{N(N + 1)} \frac{DD}{RR} - N(N_r + 1) \frac{\overline{DR}}{RR} + 1, \quad (2.7)$$

where N_r is the number of galaxies in the random field, DD is the number of pairs at separation θ , N is the total number of sources considered in the study, \overline{RR} is the averaged pair count over a number of large random simulations, and \overline{DR} is the averaged data-random cross pair count for a number of large random simulations. In order to calculate the angular separation (i.e. θ_{ij}) of the galaxies we use the distance equation 2.8:

$$\theta_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}. \quad (2.8)$$

The estimator clearly shows that there is an excess of clustering (i.e. excess of power) at angular separations $\theta \lesssim 0.2$ deg. At higher values of angular separations $\theta \gtrsim 0.2$ deg the estimator approaches zero, which indicates that there is less clustering at those separations. The study of the angular 2 point correlation function is relevant

to this study because it helps us to understand at what separations can we expect high contamination rate when we extract the sub-volumes from our cubes during the HI stacking process in chapter 3. For our cubes we are expecting to see high contamination rate at the angular separations $\theta \lesssim 0.2$ deg because of the excess of power we observe from figure 2.13.



3 HI stacking techniques

This chapter presents the methodology employed to measure the average HI (neutral hydrogen) content within and around galaxy clusters using the HI stacking technique. The stacking of HI line spectra has been widely implemented by various researchers. For instance, [Zwaan et al. \(2001\)](#) and [Chengalur et al. \(2001\)](#) utilized this technique to study the average HI content of the rich galaxy clusters Abell 2218 and Abell 3128, respectively. The study of Abell 3128 was based on the Australia Telescope Compact Array, while Abell 2218 utilized the Westerbork Synthesis Radio Telescope. In both cases, only a limited number of galaxy detections were found within the HI data cubes. Therefore, the authors employed the HI stacking method to investigate the average HI content of their respective galaxy sub-samples. Following these studies, subsequent authors have also utilized the HI stacking procedure to examine the average HI content of galaxy clusters, as demonstrated by [Fabello et al. \(2011\)](#); [Healy et al. \(2019\)](#); [Elson et al. \(2016, 2019b\)](#), and others.

The HI stacking technique operates on a simple principle: small sub-volumes are extracted from the HI data cube at the known optical positions of galaxies (RA, Dec, channels). These sub-volumes are then converted into 1-D spectra and spectrally aligned based on each galaxy's known redshift. Ultimately, the 1-D spectra are co-added to generate a high signal-to-noise spectrum, enabling the measurement of the total HI mass of the galaxy sample.

Figure 3.1 displays the stacked spectra presented in [Chengalur et al. \(2001\)](#). Panels (A) and (B) correspond to a sample of 148 galaxies from the ESO Nearby Abell Cluster Survey, while panels (C) and (D) depict 123 galaxies located outside the cluster core. All four spectra have a velocity resolution of 140 km.s^{-1} . Panels (A) and (C) show the stacked spectra for the respective samples, while panels (B) and (D) display reference spectra generated by assigning random redshift values to each sample's spectra. In the following sections, a detailed explanation of each step in the HI stacking process will be provided.

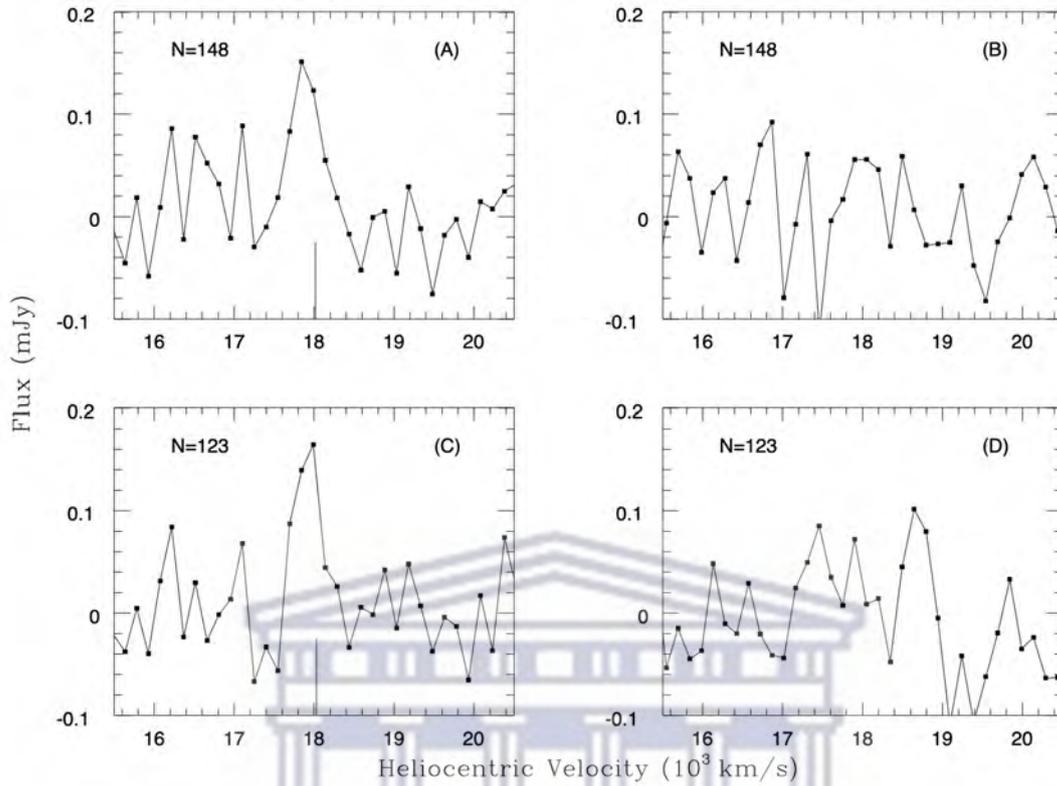


Figure 3.1: The sample of 148 galaxies in panels (A) and (B) corresponds to data from the ESO Nearby Abell Cluster Survey. The 123 galaxies in panels (C) and (D) are situated outside the cluster core. The velocity resolution for all four spectra is 140 km.s^{-1} . Panels (A) and (C) display the stacked spectra for the two samples, whereas panels (B) and (D) display the reference spectra that were produced by generating random redshift values for each sample's spectra. (image credit to: [Chengalur et al., 2001](#)).

3.1 Spectral extraction

The HI data cube refers to the three dimensional cube that contains the HI spectral line emission for thousands of galaxies. In order to get an HI spectrum of each galaxy in the full HI data cube, a sub-volume centered around a galaxy should be extracted from the full HI data cube. The sub-volumes are extracted from the full HI data cube using Python code. The [Obreschkow & Meyer \(2014\)](#) catalogue provides the RA, Dec and redshift z positions of the galaxies in the sample. The code uses the galaxy positions to extract the HI sub-volume corresponding to each galaxy. The sub-volumes enable us to focus our analysis around individual undetected galaxies and get an HI line spectrum from which we can get the approximate HI mass of the galaxy sample and consequently the average HI mass of the HI data cube.

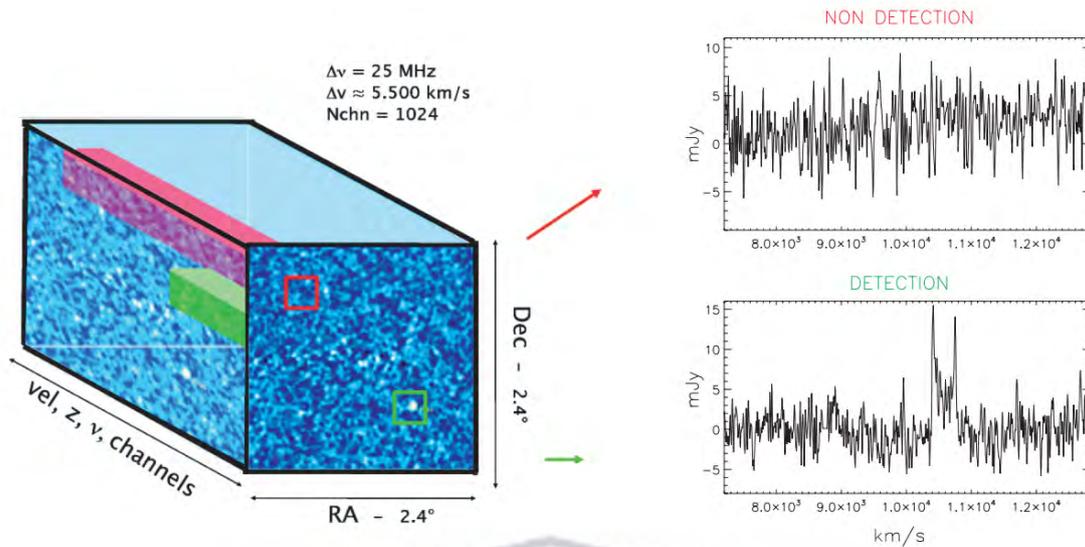


Figure 3.2: The image that illustrates how the sub-cubes are extracted from the three dimensional HI data cube using known optical positions of the galaxies and convert them to the HI line spectra. All the galaxies are buried in noise. The red sub-cube centered around a galaxy illustrates HI non-detection, while the green sub-cube illustrates HI detection. Image credit to [Fabello et al. \(2011\)](#).

Figure 3.2 illustrates the process of extracting a sub-volume centered around an individual galaxy and converting it into a one-dimensional spectrum, as demonstrated by [Fabello et al. \(2011\)](#). The LADUMA HI data cubes utilized in this study have a spectral range of 1061 channels with a frequency channel width of 26 kHz. The galaxy sub-volume is extracted from all channels and subsequently truncated to focus the analysis on the target emission. A spatial aperture size of 8-by-8 pixels is employed, corresponding to a size of 40 arcsec based on a pixel scale of 5 arcsec/pixel (as shown in Table 2.2). This aperture size ensures that the entire flux emitted by the galaxy is contained within the extracted sub-volume, avoiding the overestimation or underestimation of galaxy flux depending on the sub-volume. However, contamination by nearby galaxies remains a concern and it should be corrected.

To ensure sample completeness, 12,007 galaxies with $M_{\star}/M_{\odot} \leq 10^{8.5}$ are flagged out, as the $S^3 - SAX$ catalog has been shown to be complete in M_{\star} down to $M_{\star}/M_{\odot} \sim 10^{8.5}$. Cubelets are extracted for the remaining 14,504 galaxies with $M_{\star}/M_{\odot} > 10^{8.5}$ in the full HI data cube, and later stacked into different sub-samples based on properties such as galaxy's stellar mass, color, specific star formation rate (sSFR), and others, which will be discussed in Chapter 4. Two HI data cubes are utilized in this study, both of which are spatially smoothed using a Gaussian point-spread-

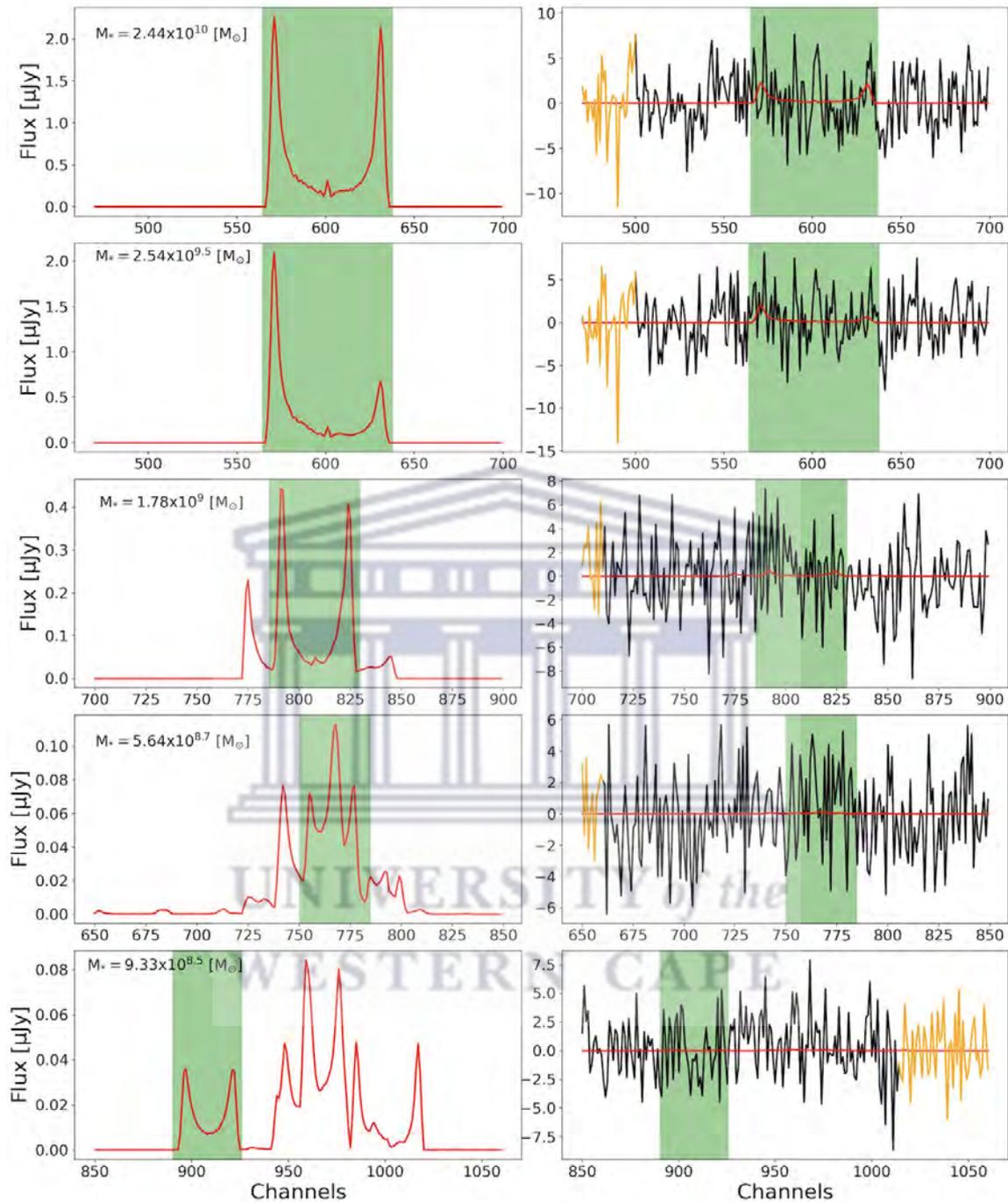


Figure 3.3: The 5 of the 14504 HI profiles selected from the sample. The left column shows the HI line spectrum extracted from the spatially smooth and noise-free HI data cube. The green region in all the panels shows the section of the spectra where the galaxy emission is expected to be located. The region is specified by the user and it is part of the galaxy mask. On the right column the red underlying profiles presents an HI line spectrum extracted from noise-free cube, while the black profile represents the same HI spectrum extracted from the noise-filled HI data cube. The orange part of the HI line spectrum will be wrapped around and appended to the other end of the spectrum when central aligning the target emission of the spectrum.

function. The difference between the cubes lies in the noise content, with the first cube being noise-free and the second cube being noise-filled. [Elson et al. \(2016\)](#) filled the second cube with noise using a Gaussian noise distribution with a standard deviation of 2.88×10^{-5} Jy. This is the standard deviation of the expected noise level on LADUMA observations at our redshift range. The noise in the HI data arises from several sources, including instrumental noise, sky background radiation, and system temperature. The noise properties are characterized by the sensitivity of the MeerKAT telescope and are considered in the data analysis and interpretation.

Figure 3.3 shows 5 of the 14504 galaxy spectra generated from the sub-cubes extracted from the HI data cubes. The profiles on the left column are extracted from the noise-free and spatially smoothed cube, hence we see that the profiles have extra flux contributed by the nearby galaxies. The extra flux to the target galaxy sub-volume is contributed by the galaxies that we refer to as "nearby neighbours" and "distant neighbours". The "nearby neighbours" are the galaxies that lies within the target galaxy sub-volume and they contribute most of the contaminant flux. "Distant neighbours" contamination emission originates from galaxies that are outside of the sub-volume but have flux bleed into it when their point-source-like flux distributions are convolved with the point spread function. The spectra on the left column reports the stellar mass of the selected galaxy. The right column shows the same HI profiles, but in this case they are extracted from the smoothed and noise filled cube. The profiles show no flux detection because they are buried in noise in the cube. The orange part of the HI line spectrum will be wrapped around and appended to the other end of the spectrum when we move each galaxy target emission to the center of the spectrum (the green region) in section 3.4. The spectra are selected in different galaxy's stellar mass bins. Our galaxies are separated in bins of $\log_{10}(M_{\star})$ with 0.5 dex. We choose galaxies in different stellar mass bins in order to demonstrate that as we extract HI sub-volumes from large galaxies to small galaxies the confusion rate increases with the decreasing galaxy's stellar mass.

3.2 Spectrum de-redshifting and wrapping

Now the sub-volume for each galaxy is extracted from the full HI data cube and converted to the HI line spectrum as described in section 3.1. The next step shown in figure 3.4 is to de-redshift the spectrum from the observed frequency, ν_{obs} , to the

rest-frame frequency, v_{rf} , by means of equation 3.1 (for example: Healy et al., 2019; Sinigaglia et al., 2022):

$$v_{rf} = v_{obs}(1 + z), \quad (3.1)$$

we then convert the spectrum to the units of velocity as $v = cz$, with the channel width given by equation 3.2:

$$dv = \frac{df \cdot c}{f_{emit}}(1 + z), \quad (3.2)$$

where, df , is the frequency channel width of the cube in kHz, c , is the speed of light in vacuum in km/s, and f_{emit} , is the emitted rest-frame frequency in kHz. The de-redshifted spectra from the sub-volumes extracted from the noise-free version of the cube is shown on the left column of figure 3.4, while the second column from the left shows the same spectra (the red curve) in this case it includes the one extracted from the noise-filled version of the cube (the black curve).

After de-redshifting the spectra and converting them to units of velocity, a wrapping process is performed to center the galaxy's target emission at the 0 km/s velocity position (i.e. $f_{emit} = 1420$ MHz). This process is described in the column 3 of figure 3.4. During the wrapping, the yellow part of the spectra is cut and appended at the other end to maintain the length and noise properties of the spectra.

To ensure consistent analysis, the spectra are then resampled to a reference spectral (or velocity) resolution. A common velocity width of $dv = 9.34$ km/s is chosen, which corresponds to the minimum redshift of $z = 0.7$. All spectra are adjusted to have a velocity channel width of 9.34 km/s.

It is important to ensure that the integrated flux density is conserved during the resampling process. This ensures that the total flux of the spectra is preserved despite the adjustments made.

In the final step, the spectra are truncated to a velocity range of ± 800 km/s for completeness. However, the analysis primarily focuses on a narrower velocity range of ± 300 km/s. This narrower range is chosen because all the galaxies in the data cubes were modeled to have a maximum rotation speed of ± 300 km/s.

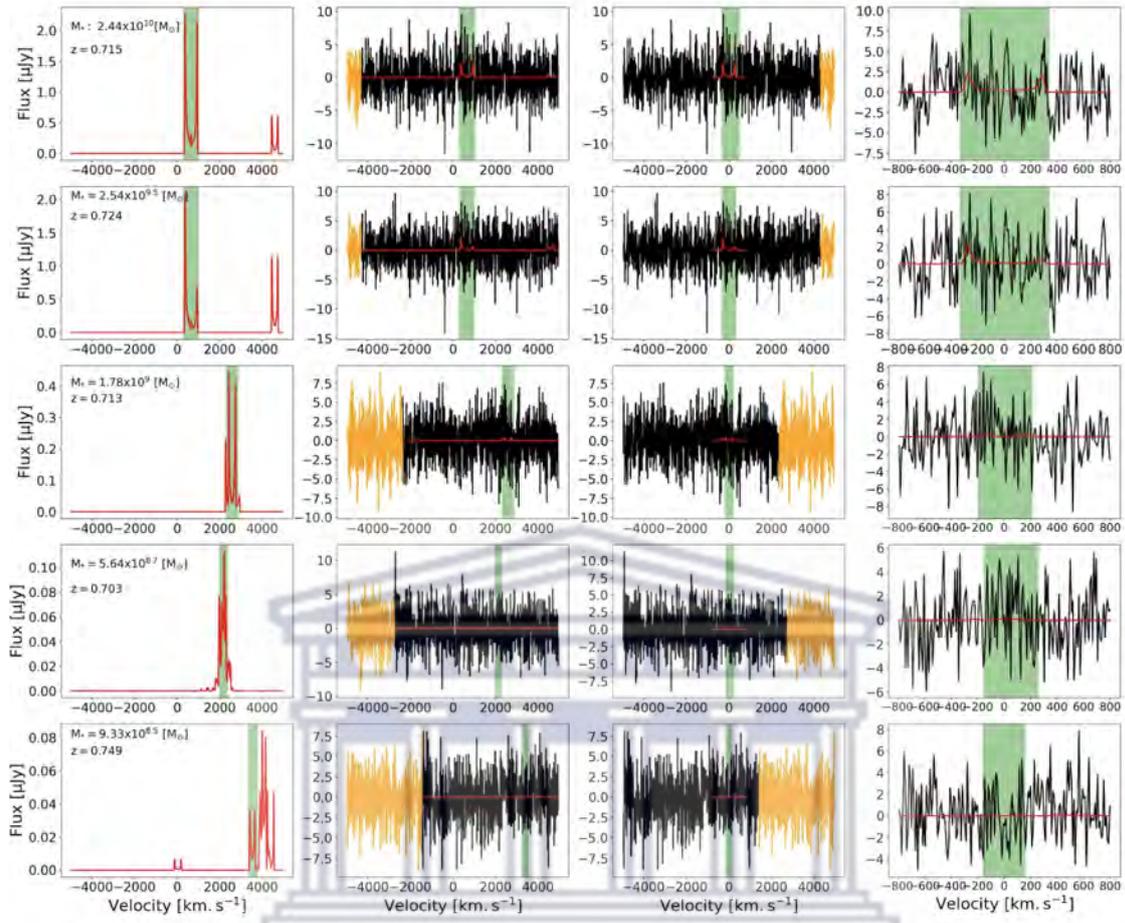


Figure 3.4: The wrapping process of the HI spectrum. The left column shows the profiles of the different galaxies extracted from the noise-free and spatially smoothed cube. The column also shows stellar mass of the galaxies, M_* , and the redshift z . The target emission of the galaxies are centered at different channels from top to bottom in the column. The second column from the left shows the profiles extracted from the noise-free cube (i.e. represented by the red curve) and the black profiles represent the spectrum extracted from the noise-filled and spatially smoothed HI data cube. The orange part of the spectrum will be wrapped around and appended to the other end of the spectrum when the target emission is shifted to the centre (i.e. 0 km.s^{-1}). The third column from the left shows the wrapped spectrum. The orange part is the one that was wrapped around and appended to the other end of the spectrum. The green region in all the panels represents the part of the spectrum where the target emission is expected to be located. This region width is specified by the user and it is the part of the galaxy mask. The right column shows the spectrum that is truncated. We truncate the spectrum so that we are able to focus our attention on the target emission only.

The method described for de-redshifting, wrapping, and resampling the spectra is similar to the HI Stacking Software (HISS) developed by (Healy et al., 2019).

3.3 Co-adding the spectra

In this study, the primary objective is to investigate the HI (neutral atomic hydrogen) content of the galaxies in the data cube. Before co-adding the spectra, the flux density profiles need to be converted to HI mass (M_{HI}). This conversion is based on an equation that assumes a spherical HI cloud that is optically thin and has a uniform internal velocity distribution. The equation used for this conversion is represented by Equation 3.3, as described in Meyer et al. (2017):

$$\frac{M_{\text{HI}}}{[M_{\odot}]} = \frac{2.36 \times 10^5}{1+z} \frac{D_{\text{L}}^2}{[\text{Mpc}]} \frac{\Sigma_i S_i \Delta v}{[\text{Jy km.s}^{-1}]}, \quad (3.3)$$

where M_{HI} is the HI mass in units of M_{\odot} , D_{L} is the luminosity distance in units of Mpc, z is the redshift and $\Sigma_i S_i \Delta v$ is the flux density in Jy km.s^{-1} , where S_i and Δv , respectively, represent HI flux and velocity width. To calculate the luminosity distance we assumed the Λ CDM cosmology with the parameters present in table 2.2 (i.e. Hubble constant $H_0 = 67.3 \text{ km.s}^{-1} \text{ Mpc}^{-1}$, $\Omega_{\text{m}} = 0.315$ and $\Omega_{\Lambda} = 0.685$). We applied the parameters in equation 3.4:

$$D_{\text{L}}(z) = (1+z) \Sigma_i \frac{\Delta z}{H_0 [\Omega_{\text{r}0}(1+z_i)^4 + \Omega_{\text{m}0}(1+z_i)^3 + \Omega_{\Lambda}]^{0.5}}. \quad (3.4)$$

After converting the flux density to HI mass. The last step to generate a stacked spectrum is to co-add the spectra. The one dimensional spectra are co-added in order to produce a high signal-to-noise spectrum that can be integrated in order to measure the average HI mass of the galaxies. From the standard approach in stacking analysis we expect that the signal-to-noise of the co-adds increases, while the Gaussian noise decreases with the factor of $1/\sqrt{N_{\text{gal}}}$, where N_{gal} is the number of galaxies in the stacked spectrum. In this work we co-add the galaxy's spectra in different sub-samples to investigate the ways in which high-redshift stacking experiments will have their results affected by the incompleteness of the optical catalogues. The co-adds results for different sub-samples are presented and discussed in chapter 4. The stacked spectrum can be expressed as (Healy et al., 2019):

$$S_{\text{stack}} = \frac{\Sigma_{i=1}^{N_{\text{gal}}} w_i S_i}{\Sigma_{i=1}^{N_{\text{gal}}} w_i} = \frac{\Sigma_{i=1}^{N_{\text{gal}}} S_i}{N_{\text{gal}}}, \quad (3.5)$$

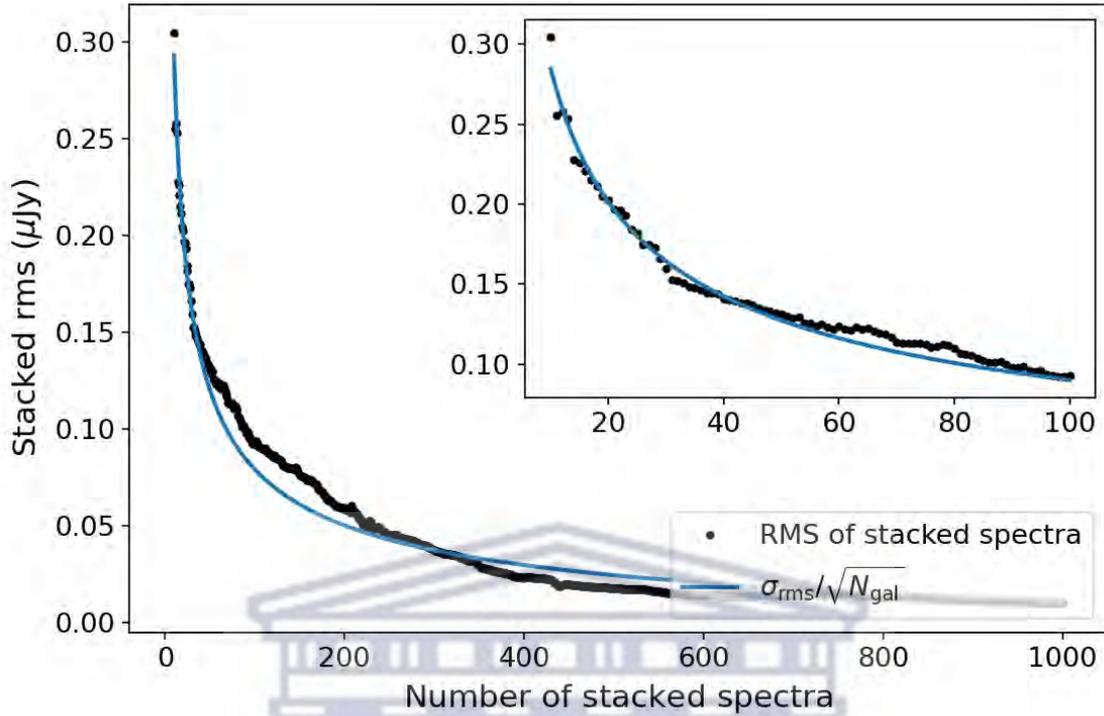


Figure 3.5: The stacked noise as a function of the number of input HI spectra. The black stars represent the measured noise in the stacked spectrum with N_{gal} input spectra and the red curve represents $\sigma_{\text{RMS}}/\sqrt{N_{\text{gal}}}$. The Gaussian noise decreases with the factor of $1/\sqrt{N_{\text{gal}}}$ as expected (for example: Healy, 2021).

where i is the number of the profiles in the stack, S_i is the co-added spectra and w_i is the weighting factor. Each spectrum has its own weighting factor. In this work we follow the standard unweighted instance, where the weighting factor, $w_i = 1$, and $\sum_{i=1}^N w_i = N_{\text{gal}}$. However, the user can choose the weighting factor to be $\sigma_{\text{RMS},i-2}$ (for example: Fabello et al., 2011) or $\sigma_{\text{RMS},i-1}$ (for example: Lah et al., 2007) or $\sigma_{\text{RMS},i} D_L^2(z)$ (for example: Delhaize et al., 2013), where $D_L(z)$ is the luminosity distance and σ_{RMS} is the Gaussian noise standard deviation. From the stacks presented in chapter 4 we have computed the RMS in the central velocity ± 300 km/s in order to check if the RMS calculated will decrease as $1/\sqrt{N_{\text{gal}}}$. Figure 3.5 shows that, indeed as the signal-to-noise ratio of our co-adds increases the Gaussian noise decreases as a function of $\sigma_{\text{RMS}}/\sqrt{N_{\text{gal}}}$ (for example: Healy, 2021) (i.e. decrease with the factor of $1/\sqrt{N_{\text{gal}}}$), where σ_{RMS} is the standard deviation of the Gaussian noise of the stacked spectra.

4 HI stacking results

In this chapter we present and discuss the stacking results for different sub-samples that we stacked. As aforementioned, the stacking method is simple in principle: Small sub-volumes are extracted from the full HI data cube at the known optical positions of galaxies. The sub-volumes are converted into 1-D spectra and then spectrally aligned according to each galaxy's known redshift. Finally, the 1-D spectra are co-added in order to produce a high signal-to-noise spectrum that can be integrated in order to measure the average HI mass of the entire galaxy sample (Zwaan et al., 2001; Chengalur et al., 2001; Fabello et al., 2011).

While the HI stacking method is simple in principle, it is difficult in practice to derive from it accurate measures of galaxy properties. This is largely due to confusion (Fabello et al., 2011; Jones, 2016; Elson et al., 2016; Rhee et al., 2023). Confusion rate refers to the level of contamination or interference from neighboring galaxies that contribute to the observed signal. When stacking multiple spectra together, the combined signal can be affected by the presence of nearby galaxies, which can introduce additional flux and cause confusion.

There are several causes for confusion rate in stacked spectra:

- Nearby galaxies: Stacked spectra may include the emission from nearby galaxies that are spatially close to the target galaxies. These nearby galaxies can contaminate the stacked spectra by contributing their own flux, causing confusion in the interpretation of the combined signal.
- Overlapping spectral features: In stacked spectra, there can be overlapping spectral features from different galaxies. These overlapping features can blend together, making it challenging to distinguish and separate the individual contributions, leading to confusion in identifying the true properties of the stacked sample.
- Redshift uncertainties: Can contribute to the confusion rate in stacked spectra by introducing spectral smearing and blending effects. When the redshifts of individual sources being stacked are imprecise or have significant uncertainties, their spectral features may not align accurately. As a result, distinct lines can blend, appear smeared, or shift in the stacked spectrum, making it challenging

to identify and separate individual source contributions. This uncertainty in redshifts can lead to misinterpretations, difficulties in line identification, and a higher confusion rate in the final stacked spectra.

- **Velocity dispersion:** Differences in the velocity dispersion of galaxies within the stacked sample can contribute to confusion in the stacked spectra. Variations in the velocity profiles of individual galaxies can broaden the combined spectral lines, making it harder to disentangle the contributions from different galaxies.

To mitigate the confusion rate in stacked spectra, careful analysis techniques are employed, including background subtraction, modeling and subtracting the contamination from nearby galaxies, and statistical methods to separate the individual contributions. These approaches help to disentangle the combined signals and reduce the impact of confusion caused by neighboring galaxies.

However, the problem becomes even more difficult in the sense that rates of source confusion depend not only on the 3D clustering of galaxies, but also on the selection effects imposed by the optical catalogue used to locate the galaxies in the HI data cube. Any optical catalogue of high-redshift galaxies will be magnitude limited. This means that the catalogue is biased towards certain types of galaxies. For example: stacking galaxies that are highly luminous in the B-band likely makes them galaxies that have high star-formation rates, which affects their HI content. Or, stacking earlier type galaxies yields a stack more susceptible to the confusion effects induced by spatial clustering.

To address this issue, we stack galaxies in different sub-samples to understand the varying levels of contamination. Sub-samples composed of galaxies that are biased in the catalog will exhibit a higher confusion rate. This chapter is structured as follows: In Section 4.1, we present the stacking results for all galaxies in the sample. Section 4.2 demonstrates the stacking results for galaxies separated into stellar mass bins, while Section 4.3 showcases the stacking results for galaxies divided into star-formation rate (SFR) bins. Additionally, Section 4.4 and 4.5 respectively present the stacking results for galaxies grouped by specific star-formation rate (sSFR) and galaxy color. Finally, in Section 4.6, we compare the confusion rates across all sub-samples. Although we also stacked and quantified the confusion rates for group and non-group galaxies, the results are presented in Chapter 5, specifically in Section 5.2.

4.1 Stacking in all the galaxies

In this section we present the stacking results for all the galaxies in our sample after we have flagged out all the galaxies with $M_*/M_\odot \leq 10^{8.5}$, because our catalogue is incomplete below this stellar mass limit. Figure 4.1 shows the successive stacked spectra after co-adding n-galaxies. We present the stacks after a number of stacked galaxies in order to show that as we stack more galaxies together the signal of the stacked sample starts to show at the velocity range ± 300 km/s (the green region).

On each panel of figure 4.1 we show the spectrum produced from the noise-free version of our cubes (represented by the red curve) and the spectrum produced from the noise-filled version of our cubes (represented by the black curve). Apparently, the stacked spectra show a characteristic shape. Almost all of the stacks have extended wings going beyond maximal model rotation speed (± 300 km/s, the green region) of the stacked galaxies. These wings are made up entirely of the contaminant mass by nearby galaxies, since target galaxy mass is within the velocity range ± 300 km/s. Nonetheless, within velocity range ± 300 km/s there is contaminant mass, too. The stacked spectra reveals a noticeable double horn-shape especially when many galaxies are being stacked. The reasons for this shape are: because the galaxies in our sample were model using template rotation curves that are typically flat at out radii and each target galaxy has no redshifts uncertainties.

The panels report the evaluated average HI mass from the stacked spectra, $\langle M_{\text{HI}}^{\text{Stack}}/M_\odot \rangle$, the true average HI mass based on the evaluated HI masses from the catalogue, $\langle M_{\text{HI}}^{\text{True}}/M_\odot \rangle$, the number of stacked galaxies, N_{gal} , and the confusion rate on the stacks (conf rate), $\langle M_{\text{HI}}^{\text{Stack}} \rangle / \langle M_{\text{HI}}^{\text{True}} \rangle$. The $\langle M_{\text{HI}}^{\text{Stack}}/M_\odot \rangle$ is computed by integrating over the velocity range ± 300 km/s.

Figure 4.2 shows the change in confusion rate as a function of the number of stacked galaxies. From this figure we recognised that for the first few stacked galaxies the confusion rate ranges between ~ 1.27 to ~ 5 . The large deviation in the contamination rate after few stacked galaxies, implies that when stacking all the galaxies in the catalogue it happens that we also stack galaxies with high confusion rate. That is why it is important to specify the types of galaxies you are interested in stacking, instead of stacking the full sample.

On the upcoming sections we stack galaxies according to their properties and that

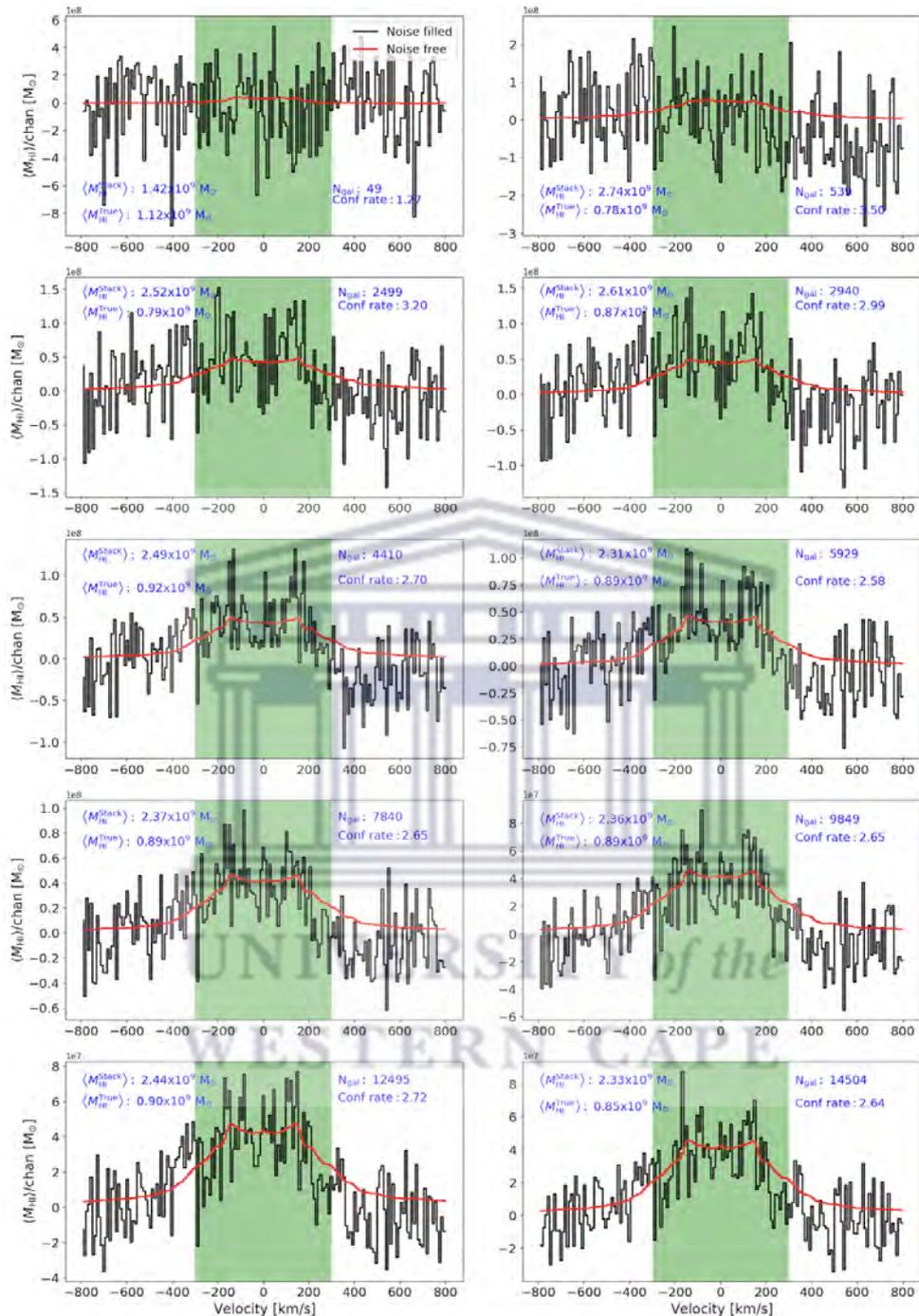


Figure 4.1: Shows the stacking results for the galaxies in the sample. Each panel shows the successive stacked spectrum after co-adding n -number of galaxies spectra (N_{gal}). The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}} / M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}} / M_{\odot} \rangle$, the number of stacked galaxies, N_{gal} , the confusion rate and the signal-to-noise ratio.

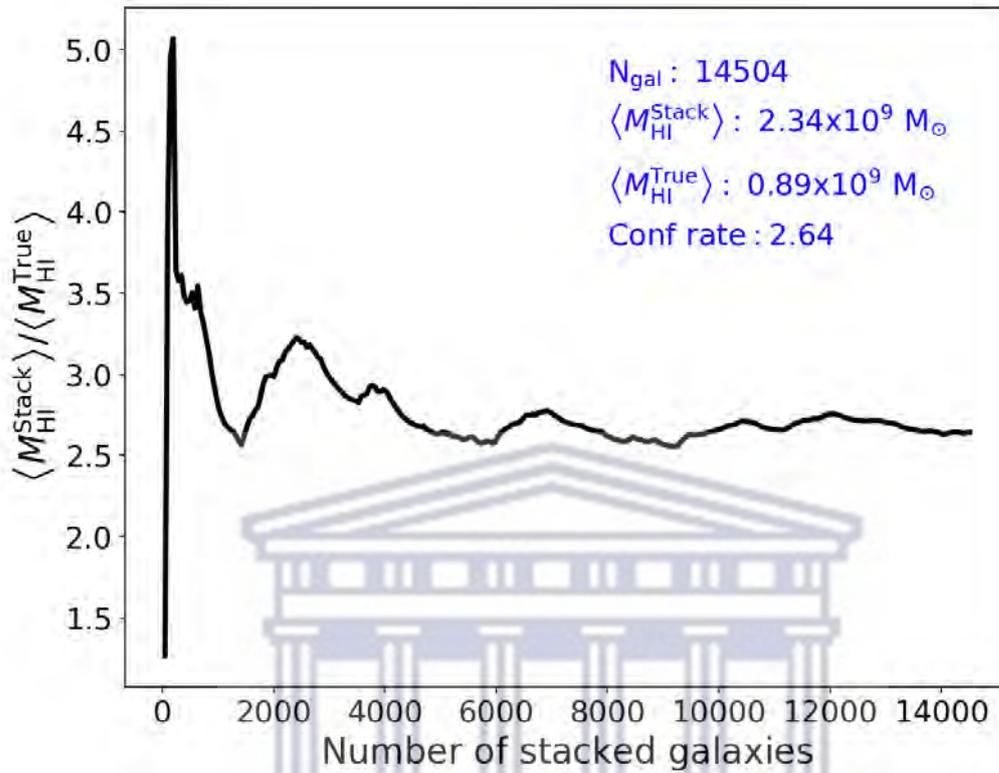


Figure 4.2: Shows how the confusion rate changes as we stack more and more galaxies in the sample. It also reports the total number of stacked galaxies in the sample, N_{gal} , the average HI mass evaluated from the stacks after stacking all the galaxies in the sample, the true average HI mass and the confusion rate after stacking all the galaxies.

Sample	N_{gal}	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	Confusion rate
Full sample	14504	0.89	2.34	2.64

Table 4.1: The resulting parameters for the full galaxy sample.

helps in determining which galaxies mostly yields a stack more susceptible to the confusion effects. Once determined the types of galaxies that yields a spectrum with high confusion rate, one could be able to flag them when performing the stacking experiments, to avoid stacked spectrum with high confusion rate. Table 4.1 presents all the parameters obtained after stacking all the galaxies in the sample.

4.2 Stacking in stellar mass bins

This section presents the stacking results for stellar mass bins. The stellar mass bins are separated in $\log_{10}(M_{\star}/M_{\odot}) = 0.5$ dex, with $8.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 9$, $9 < \log_{10}(M_{\star}/M_{\odot}) \leq 9.5$, $9.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 10$ and $\log_{10}(M_{\star}/M_{\odot}) > 10$. Figure 4.3 shows the stacking results for the first stellar mass bin, $8.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 9$. This figure shows the stacked spectrum after every successive stacked number of galaxies in this bin.

Figure 4.4 showcases the stacking results for different stellar mass bins. The black and red curves represent the stacked spectra obtained from the noise-filled and noise-free versions of our cubes, respectively. Analyzing the results from these stellar mass bins, we observe a consistent trend where the confusion rate decreases as we move from bins containing galaxies with smaller stellar masses to those with larger stellar masses. This pattern was also evident when extracting sub-volumes for target galaxies in different mass bins, as depicted in Chapter 3 figure 3.3.

The higher confusion rate observed in small galaxies can be attributed to their susceptibility to contamination from nearby neighboring galaxies. These smaller galaxies tend to have a larger number of nearby galaxies, which contributes to a higher confusion rate during the stacking process. Furthermore, small galaxies are often located in close proximity to larger galaxies, and since larger galaxies have higher flux densities, even a small fraction of their flux can significantly contaminate the extracted sub-volume of the smaller galaxy. Consequently, this leads to a higher confusion rate in the stacked spectrum of small galaxies.

Stellar mass is often correlated with various galaxy properties, such as age, metallicity, and star formation activity. Lower mass galaxies tend to exhibit a wider range of populations, including younger and older stellar populations, resulting in more diverse spectral features. We speculate that the increased complexity of spectral profiles in lower mass galaxies contributes to a higher confusion rate during the stacking process.

Interestingly, galaxies with the highest stellar masses (bin4: $\log_{10}(M_{\star}/M_{\odot}) > 10$) display a significantly lower confusion rate compared to the other stellar mass bins. This can be attributed to the fact that larger galaxies in the full HI data cubes are often surrounded by or located close to smaller galaxies. As larger galaxies have

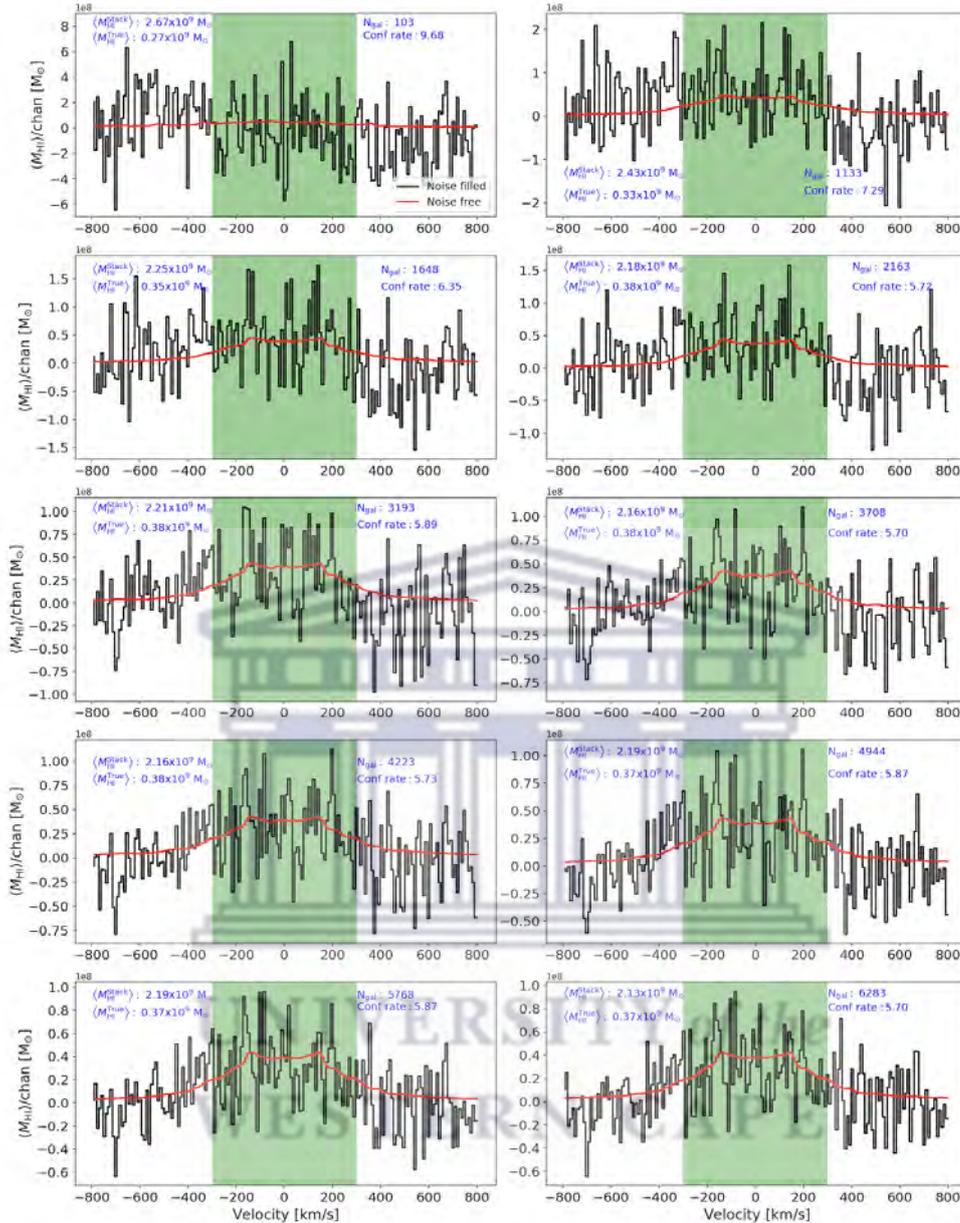


Figure 4.3: The stacking results for the 1st stellar mass bin, with $8.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 9$. Each panel shows the successive stacked spectrum after co-adding n -number of galaxies spectra (N_{galx}). The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{galx} , the confusion rate and the signal-to-noise ratio.

higher flux densities compared to smaller galaxies, they dominate the flux in their extracted sub-volumes, while the contamination contribution from smaller galaxies to the sub-volumes of larger galaxies is relatively minimal. Consequently, larger

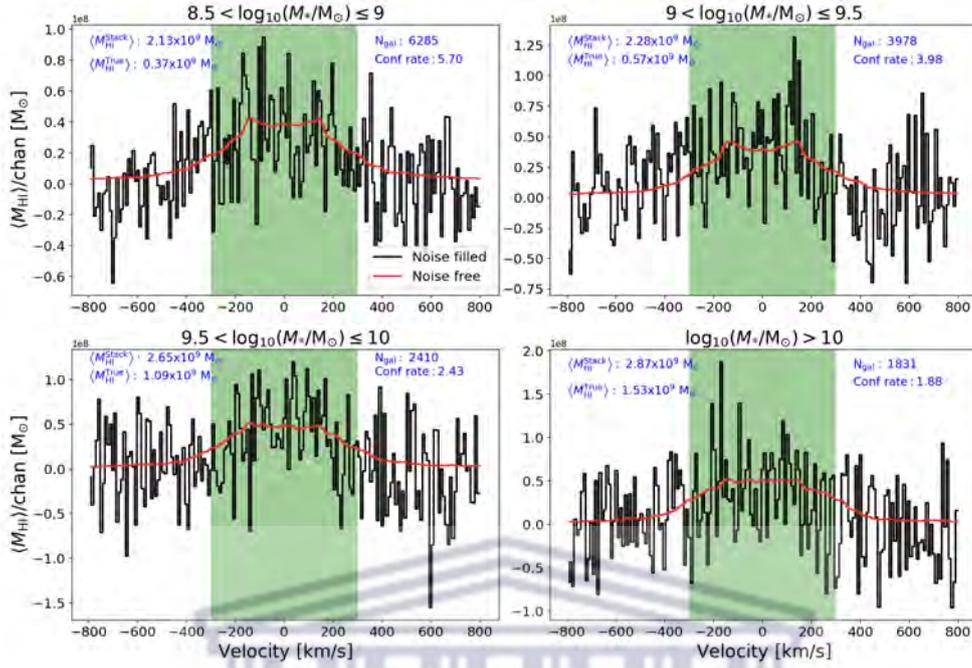


Figure 4.4: Shows the stacking results for all the stellar mass bins. The title of each panel indicates the corresponding stellar mass bin of the galaxies. The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{gal} , the confusion rate and the signal-to-noise ratio.

galaxies exhibit a lower confusion rate in the stacking process.

Understanding the variations in the confusion rate across different stellar mass bins sheds light on the impact of galaxy size and proximity on stacking experiments. These findings underscore the importance of considering galaxy properties and spatial distributions when conducting stacking analyses.

Stellar mass, M_{\star} , bins	N_{gal}	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	Confusion rate
$8.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 9$	6283	0.37	2.65	5.70
$9 < \log_{10}(M_{\star}/M_{\odot}) \leq 9.5$	3978	0.57	2.28	3.98
$9.5 < \log_{10}(M_{\star}/M_{\odot}) \leq 10$	2410	1.09	2.65	2.43
$\log_{10}(M_{\star}/M_{\odot}) > 10$	1830	1.53	2.87	1.88

Table 4.2: The resulting parameters for all the stellar mass bins.

Figure 4.5 provides further insight into how the confusion rate changes as we co-add more and more galaxy spectra in each bin. In the first bin (top left panel), when we

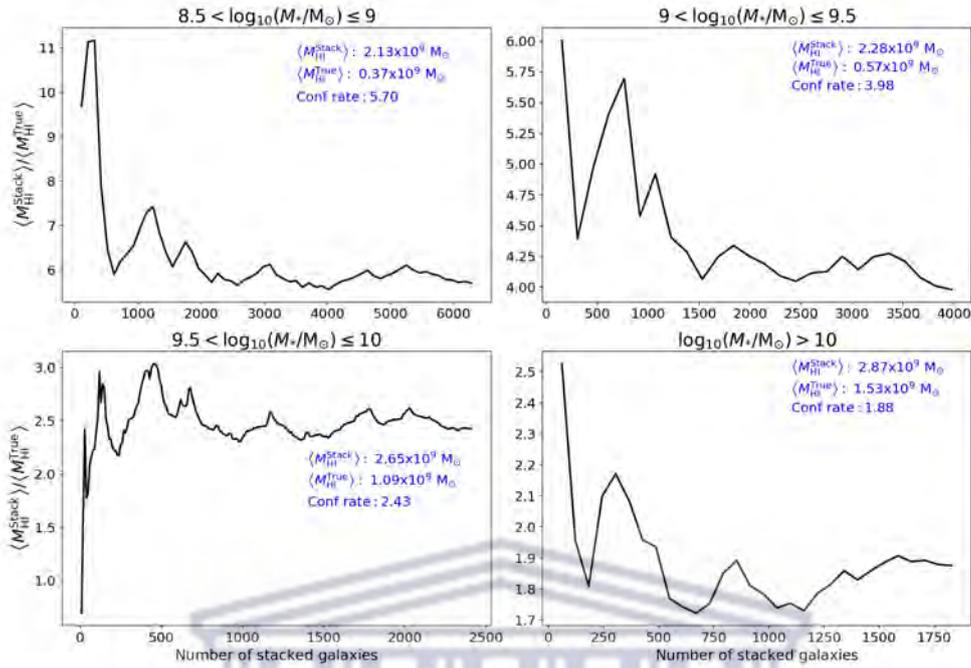


Figure 4.5: The confusion rate for each stellar mass bin as a function of the number of stacked galaxies. The title of each panel indicates the corresponding stellar mass bin of the galaxies. All the panel shows how the confusion rate changes as we stack more and more galaxies in the sample. The panels also reports the number of the stacked galaxies in the sample, N_{galx} , the average HI mass evaluated from the stacks after stacking all the galaxies in each bin, the true average HI mass and the confusion rate.

co-add the first few galaxies (less than or approximately 1000), the confusion rate ranges between approximately 6 to 11. However, as we co-add more galaxies, the confusion rate becomes consistent within the range of approximately 5 to 6. After stacking all the galaxies' spectra in this bin, the resulting confusion rate is 5.70. Similar patterns can be observed in the other stellar mass bins, where the confusion rate gradually decreases as more galaxies are co-added. The specific confusion rate values for each bin after stacking all the galaxies' spectra are provided in Table 4.2.

The inverse relationship between the confusion rate and the number of stacked galaxy spectra (shown in figure 4.5) has crucial insights and implications for scientific research and data analysis. This phenomenon arises due to several key factors, each contributing to the observed trend.

Firstly, when a smaller number of galaxy spectra are stacked, the individual spectral features may not be well-resolved, and their contributions can become convoluted. This overlapping of spectral lines from different sources results in confusion and

ambiguity in identifying the true characteristics of individual galaxies. However, as the number of stacked spectra increases, the combined signal becomes stronger, and the random noise is diluted, leading to a clearer and more distinct stacked spectrum. This improvement in signal-to-noise ratio enhances the ability to identify and separate spectral features accurately, thereby reducing the confusion rate.

Moreover, with a larger dataset, the stacked spectrum represents a more statistically robust average of the underlying galaxy population. The stacking process effectively averages out random variations and instrumental noise, providing a more reliable representation of the typical spectral characteristics of galaxies. This statistical enhancement minimizes the chances of false identifications and reduces the confusion rate in the final stacked spectra.

Furthermore, the inverse relationship between the confusion rate and the number of stacked spectra is indicative of the increasing confidence in the interpretations as the sample size grows. With a limited number of stacked spectra, any peculiarities or variations in individual galaxies can disproportionately impact the combined spectrum, leading to a higher confusion rate. However, as more spectra are combined, the influence of these individual variations diminishes, and the stacked spectrum becomes more representative of the overall galaxy population. This increased representativeness results in clearer spectral signatures and a lower confusion rate.

The implications of this inverse relationship are significant for various aspects of astronomical research. Firstly, the reduction in confusion rate with increasing stacked spectra enhances the reliability of the analysis, allowing researchers to draw more confident conclusions from their data. This improved accuracy is essential for understanding complex astrophysical phenomena and making precise measurements of galaxy properties.

Secondly, the ability to reduce the confusion rate by stacking more spectra can unlock valuable information from faint or low signal-to-noise data. Astronomical observations often involve data with inherent noise and limitations, and stacking offers a powerful tool to overcome these challenges and extract meaningful insights from the data.

Finally, the inverse relationship highlights the importance of sample size and statistical robustness in astronomical studies. A larger dataset not only improves the quality of the stacked spectra but also facilitates subpopulation analyses, enabling

researchers to explore and compare various subsets of galaxies, which may reveal new and interesting astrophysical trends.

Overall, the inverse relationship between the confusion rate and the number of stacked galaxy spectra demonstrates the significance of data quality, statistical significance, and sample size in astronomical research. Stacking spectra is a valuable technique that allows researchers to enhance signal detection, reduce noise, and gain deeper insights into the underlying physical processes shaping galaxies in our universe. By carefully considering the implications of this relationship, astronomers can maximize the scientific value of their data and advance our understanding of the cosmos.

4.3 Stacking in SFR bins

This section presents the stacking results for star-formation rate bins. We pick star-formation rate as one of our sub-samples, because it is one of the most important properties in understating how HI gas content regulates in galaxies. The [Obreschkow & Meyer \(2014\)](#) catalogue does not provide the star-formation rate properties, but from various properties that the catalogue provides shown in table 2.1 we were able to follow the method explained in [Bigiel et al. \(2008\)](#) to compute the star-formation rate for a sample of galaxies. [Bigiel et al. \(2008\)](#) method uses the effective radius of the galaxy, r_e/pc , to compute the area of the galaxy at that radius, assuming that the galaxy's shape is perfectly spherical. This method follows by computing the average molecular hydrogen surface density of the galaxy, $\Sigma_{\text{H}_2}/10\text{M}_\odot\text{pc}^{-2}$. Furthermore, the method applies the $\Sigma_{\text{H}_2}/10\text{M}_\odot\text{pc}^{-2}$ to the Bigiel relation ([Bigiel et al., 2008](#)) given by equation 4.1 to get the average SFR surface density, $\Sigma_{\text{SFR}}/\text{M}_\odot\text{yr}^{-1}\text{pc}^{-2}$:

$$\frac{\Sigma_{\text{SFR}}}{[\text{M}_\odot\text{yr}^{-1}\text{pc}^{-2}]} = a \left(\frac{\Sigma_{\text{H}_1, \text{H}_2, \text{gas}}}{[\text{M}_\odot\text{yr}^{-1}\text{pc}^{-2}]} \right)^N = a \left(\frac{\Sigma_{\text{H}_2}}{[\text{M}_\odot\text{yr}^{-1}\text{pc}^{-2}]} \right)^N, \quad (4.1)$$

where a and N are the power-law index and they are free parameters. To get the total SFR we multiply the average SFR surface density, $\Sigma_{\text{SFR}}/\text{M}_\odot\text{yr}^{-1}\text{pc}^{-2}$, with the area of the galaxy at effective radius. We followed the same method for all the galaxies. Figure 4.6 panel (a) shows the SFR distribution for the galaxies in our sample. This

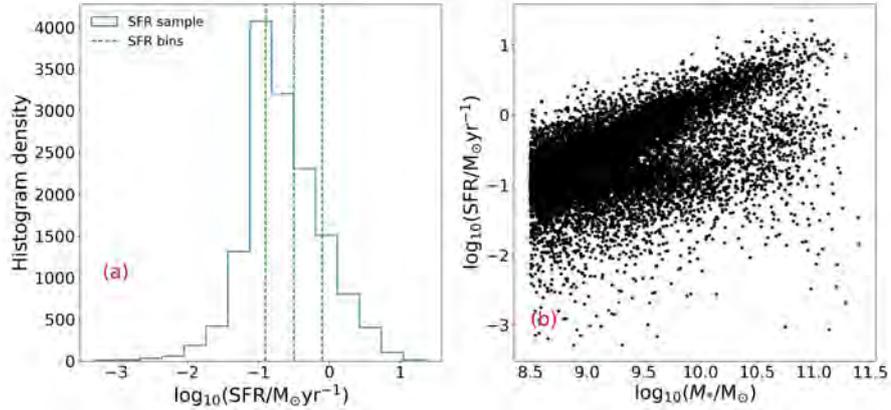


Figure 4.6: Panel (a) shows the star-formation rate distribution for our sample. The green vertical dashed lines indicate the star-formation rate binning used in this study. Bin 1 consists of 4931 galaxies, bin 2 consists of 4426 galaxies, bin 3 consists of 2837 galaxies and bin 4 consists of 2310 galaxies. Panel (b) shows the relation between M_*/M_\odot and $\text{SFR}/M_\odot\text{yr}^{-1}$ for our sample.

panel shows that our galaxy sample has fewer galaxies with $\log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) \geq 0$. The green vertical lines separate our galaxies in SFR bins. The star-formation rate is separated in bins, with $\log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) \leq -0.9$, $-0.9 < \log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) \leq -0.5$, $-0.5 < \log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) \leq -0.1$ and $\log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) > -0.1$. The bin limits were picked such that we have enough galaxies per bin to produce stacked spectra with clear signal. Panel (b) shows the relation between SFR and stellar mass. As expected the SFRs are correlated with the stellar masses of the galaxies. The correlation between SFR and stellar mass, demonstrate that galaxies with higher stellar masses tend to have higher SFRs, and galaxies with lower stellar masses tend to have lower SFRs. The relationship between these two parameters sheds light on the fundamental processes governing galaxy evolution and the interplay between star formation and the growth of galaxies over cosmic time.

Figure 4.7 illustrates the stacking results for the first SFR bin, $\log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) \leq -0.9$. Additionally, Figure 4.8 displays the stacking results for all the SFR bins. The observed confusion rate patterns in the SFR bins closely align with the findings in the stellar mass bins. This correlation is not surprising since galaxies with smaller stellar masses tend to have lower SFRs compared to larger galaxies. Consequently, galaxies with small SFRs are more susceptible to contamination from nearby galaxies, including the larger galaxies in their vicinity, as previously discussed in the context of the confusion rate in the stellar mass bins.

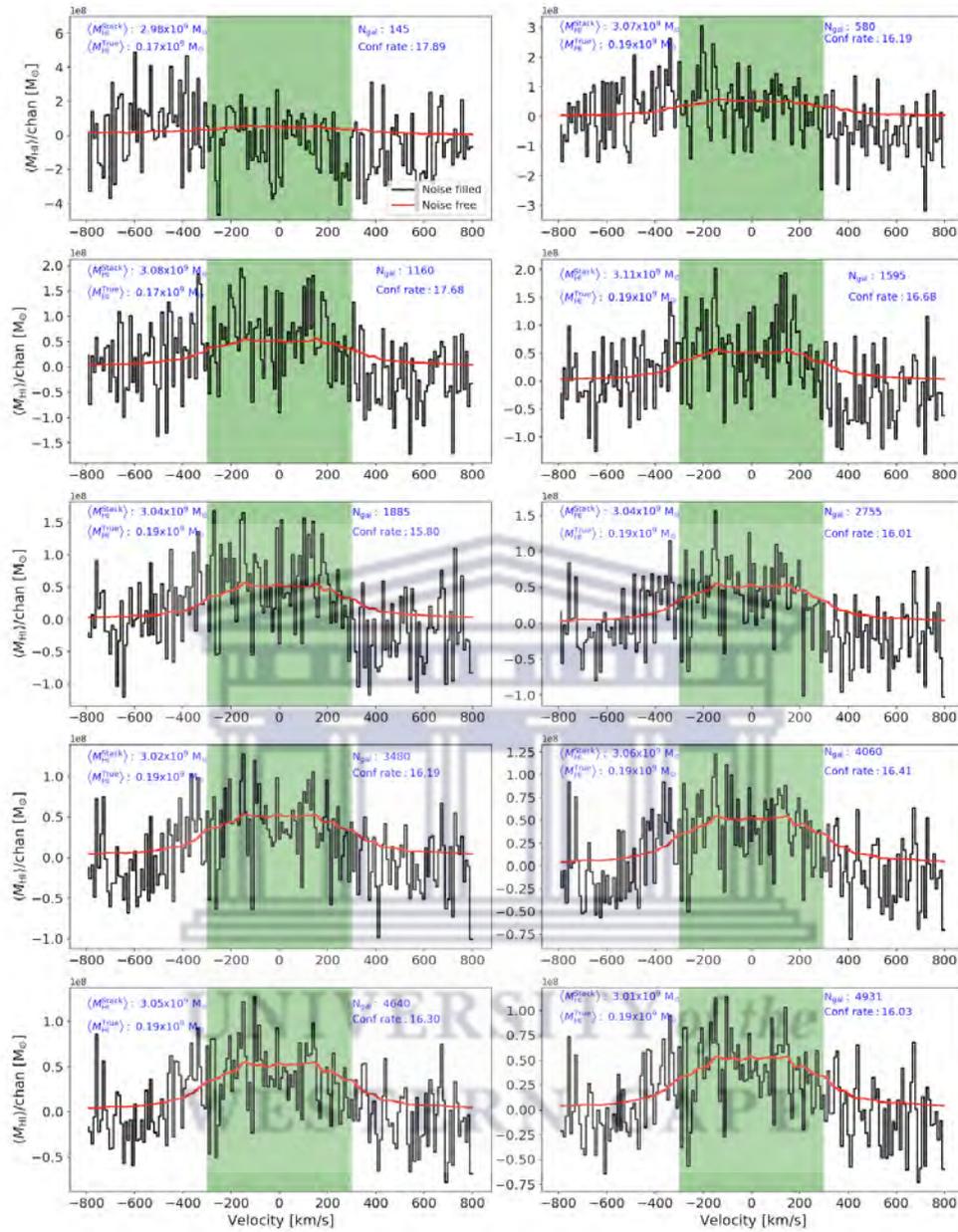


Figure 4.7: The stacking results for the 1st SFR bin, with $\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) \leq -0.9$. Each panel shows the successive stacked spectrum after co-adding n -number of galaxies spectra (N_{galx}). The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{galx} , the confusion rate and the signal-to-noise ratio.

The reason for lower confusion rates in the SFR bin with higher SFRs ($\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) > -0.1$) is that this particular bin predominantly consists of larger galaxies, which experience less contamination from neighboring smaller galaxies.

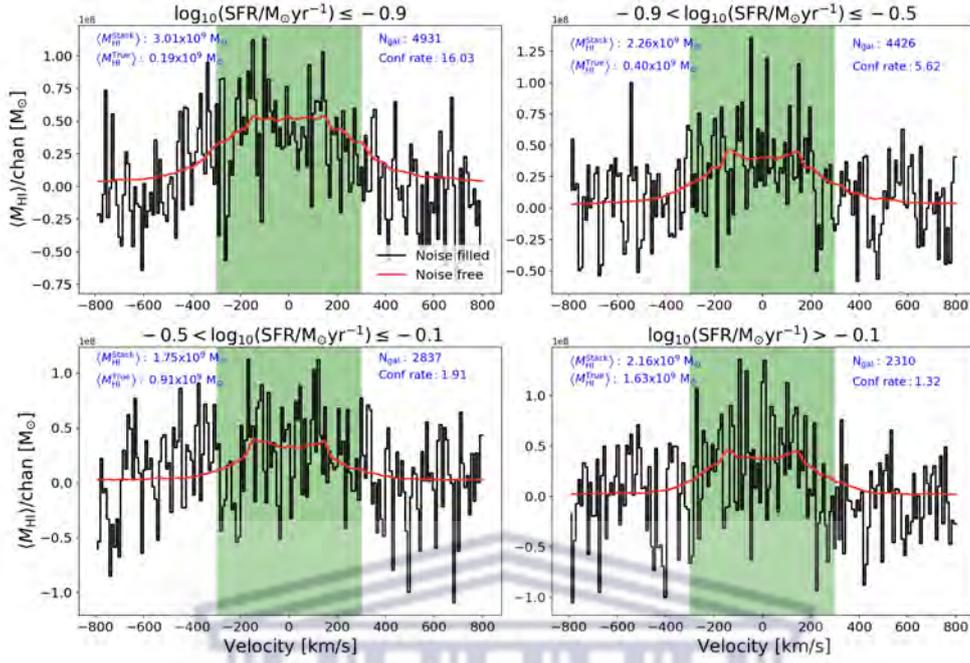


Figure 4.8: Shows the stacking results for all the SFR bins. The title of each panel indicates the corresponding SFR bin of the galaxies. The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{gal} , the confusion rate and the signal-to-noise ratio.

Figure 4.9 provides further insight into the changes in the confusion rate as more galaxy spectra are co-added within each SFR bin. The trends observed in these plots reflect the gradual decrease in confusion rates as more galaxies are included in the stacking process. Table 4.3 summarizes the evaluated parameters obtained from the simulated data (true parameters) as well as the parameters derived from the stacks for each SFR bin.

SFR bins	N_{gal}	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	Confusion rate
$\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) \leq -0.9$	4931	0.19	3.01	16.03
$-0.9 < \log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) \leq -0.5$	4426	0.40	2.26	5.62
$-0.5 < \log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) \leq -0.1$	2837	0.91	1.75	1.91
$\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) > -0.1$	2310	1.63	2.16	1.32

Table 4.3: The resulting parameters for all the SFR bins.

The results obtained from the SFR bin analysis reinforce the role of both stellar mass and SFR in influencing the confusion rate during stacking experiments. Galaxies

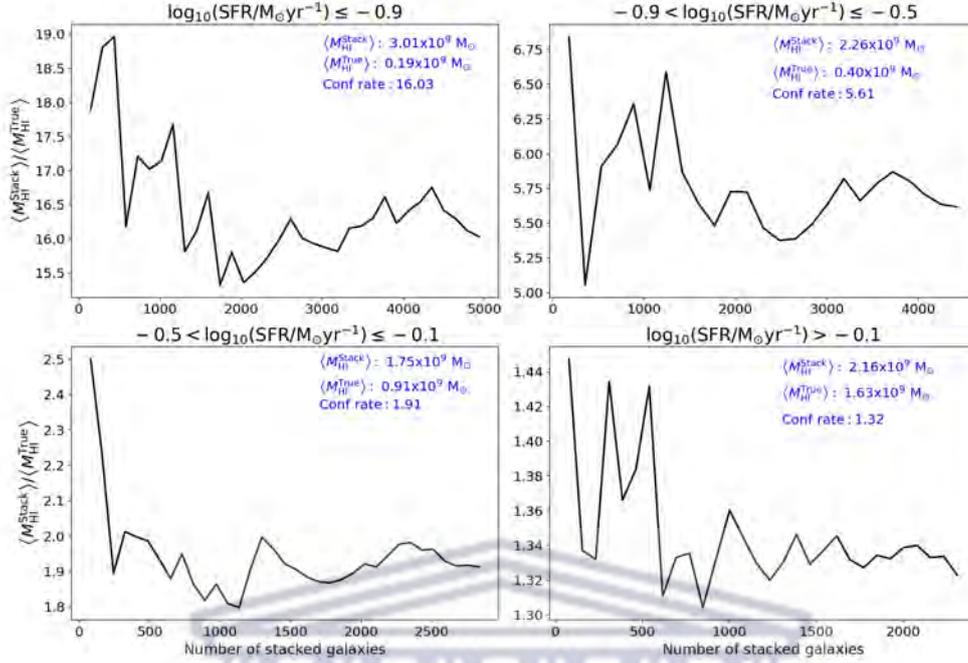


Figure 4.9: The confusion rate for each SFR bin as a function of the number of stacked galaxies. The title of each panel indicates the corresponding SFR bin of the galaxies. All the panel shows how the confusion rate changes as we stack more and more galaxies in the sample. The panels also report the average HI mass evaluated from the stacks after stacking all the galaxies in each bin, the true average HI mass and the confusion rate.

with smaller stellar masses and lower SFRs are more prone to contamination due to their proximity to nearby galaxies, while larger galaxies with higher SFRs exhibit lower confusion rates.

4.4 Stacking in sSFR bins

In this section we present the stacking results for the specific star-formation rate $\text{sSFR}/\text{yr}^{-1} = M_{\star}/\text{SFR}$ bins. The specific star-formation rate of the galaxies is binned as follows, bin1 $\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$, bin2 $-10 < \log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -9.8$, bin3 $-9.8 < \log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -9.6$ and bin4 $\log_{10}(\text{sSFR}/\text{yr}^{-1}) > -9.6$. The bin limits were picked such that we have enough galaxies per bin to produce stacked spectra with clear signal. Figure 4.10 panel (a) shows the sSFR distribution for the galaxies in our sample. The green vertical lines separate our galaxies in sSFR bins. Panel (b) shows the relation between sSFR and stellar mass of the galaxies in the sample. Most of the large galaxies $\log_{10}(M_{\star}/M_{\odot}) \gtrsim 10$ have lower sSFR when

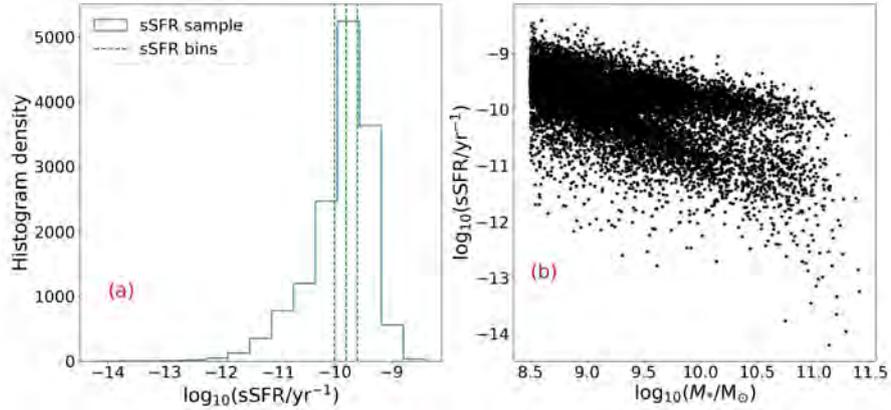


Figure 4.10: Panel (a) shows the specific star-formation rate distribution for our sample. The green vertical dashed lines indicate the specific star-formation rate binning used in this study. Bin 1 consists of 4610 galaxies, bin 2 consists of 2137 galaxies, bin 3 consists of 2936 galaxies and bin 4 consists of 4821 galaxies. Panel (b) shows the relation between M_*/M_\odot and $\text{sSFR}/\text{yr}^{-1}$ for our sample.

compared to the small galaxies $\log_{10}(M_*/M_\odot) < 10$. The correlation between sSFR and galaxy stellar mass indicates that less massive galaxies tend to have higher specific star-formation rates, while more massive galaxies exhibit lower specific star-formation rates. Understanding this relationship provides valuable insights into the processes governing the balance between star formation and stellar mass growth in galaxies.

Figure 4.11 presents the stacking results for the first sSFR bin, $\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$, while Figure 4.12 showcases the stacking results for all the sSFR bins. It is noteworthy that the confusion rates observed in the sSFR bins exhibit distinct patterns. Galaxies in the lowest sSFR bin (bin1: $\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$) demonstrate a high confusion rate, followed by the second sSFR bin (bin2: $-10 < \log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -9.8$) and the third bin (bin3: $-9.8 < \log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -9.6$). On the other hand, the fourth sSFR bin (bin4: $\log_{10}(\text{sSFR}/\text{yr}^{-1}) > -9.6$) exhibits the lowest confusion rate.

These findings are intriguing because, as we observed in the analysis of the stellar mass bins and SFR bins, the confusion rates appeared to be positively correlated. One might expect that if galaxies with larger stellar masses have lower confusion rates (as discussed earlier), then galaxies with lower sSFRs would also be dominated by larger galaxies, resulting in lower confusion rates. However, our results contra-

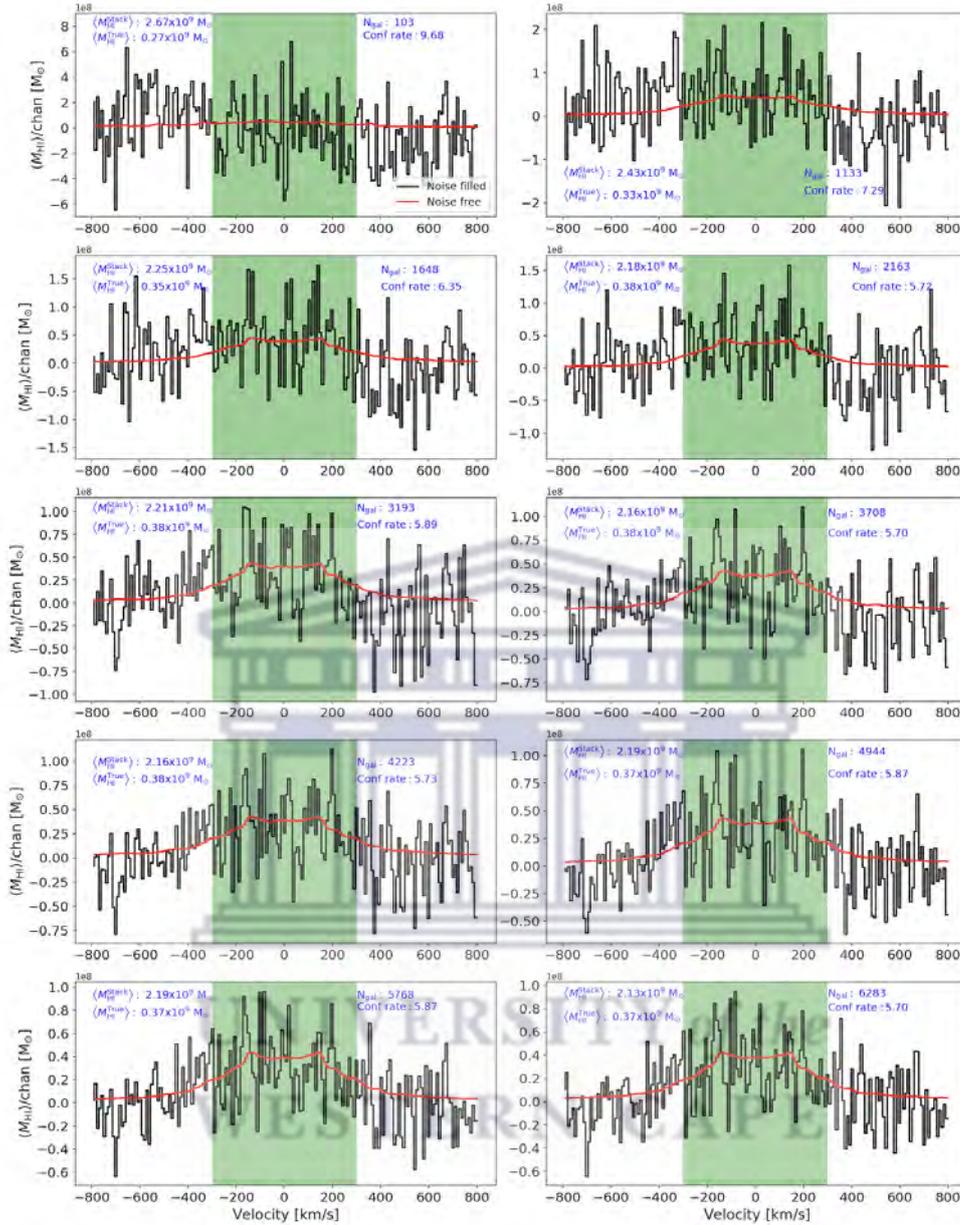


Figure 4.11: The stacking results for the 1st sSFR bin, with $\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$. Each panel shows the successive stacked spectrum after co-adding n-number of galaxies spectra (N_{galx}). The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{galx} , the confusion rate and the signal-to-noise ratio.

dict this expectation. When examining the relationship between sSFRs and stellar masses, as shown in panel (b) of Figure 4.10, it becomes apparent that a range of galaxies, from small to large stellar masses, can exhibit similar sSFRs within each

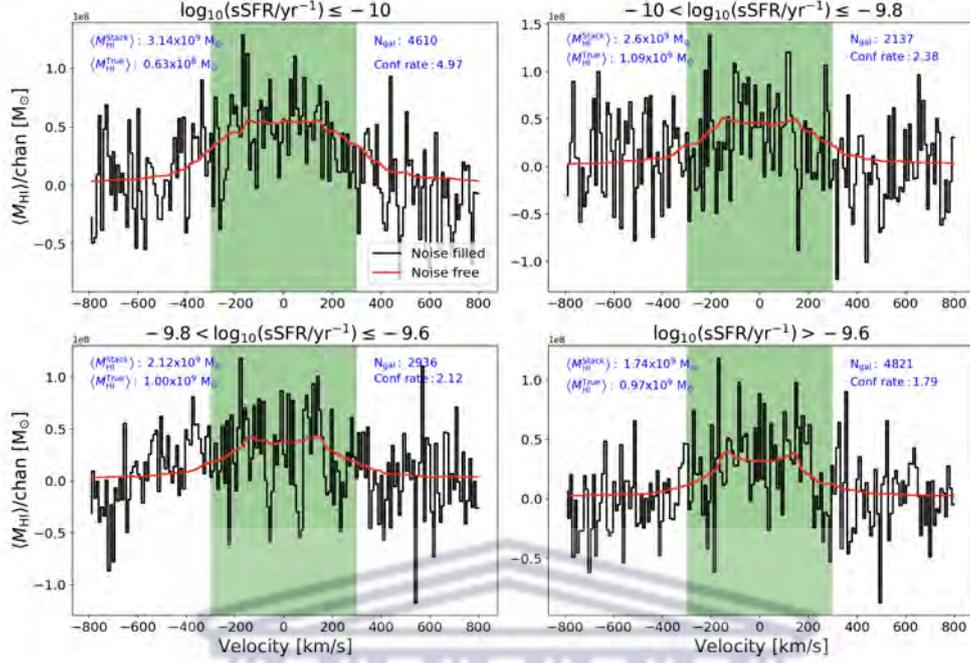


Figure 4.12: Shows the stacking results for all the sSFR bins. The title of each panel indicates the corresponding sSFR bin of the galaxies. The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{gal} , the confusion rate and the signal-to-noise ratio.

sSFR bin. Consequently, we speculate that the anti-correlation between sSFR confusion rates and both the stellar mass bins and SFR bins confusion rates arises from the fact that each sSFR bin encompasses a range of galaxies with varying stellar masses. This interpretation offers a plausible explanation for the observed results.

To further explore the influence of co-adding more galaxy spectra within each bin, Figure 4.13 depicts how the confusion rate changes as more galaxies are stacked. This provides additional insights into the behavior of the stacking process as the number of galaxies included increases.

The analysis of the sSFR bins provides valuable information regarding the relationship between sSFR and confusion rates, offering a unique perspective compared to the stellar mass and SFR bins. The presence of a diverse range of galaxies within each sSFR bin, characterized by different stellar masses but similar sSFRs, contributes to the observed anti-correlation between sSFR confusion rates and the confusion rates in the stellar mass and SFR bins.

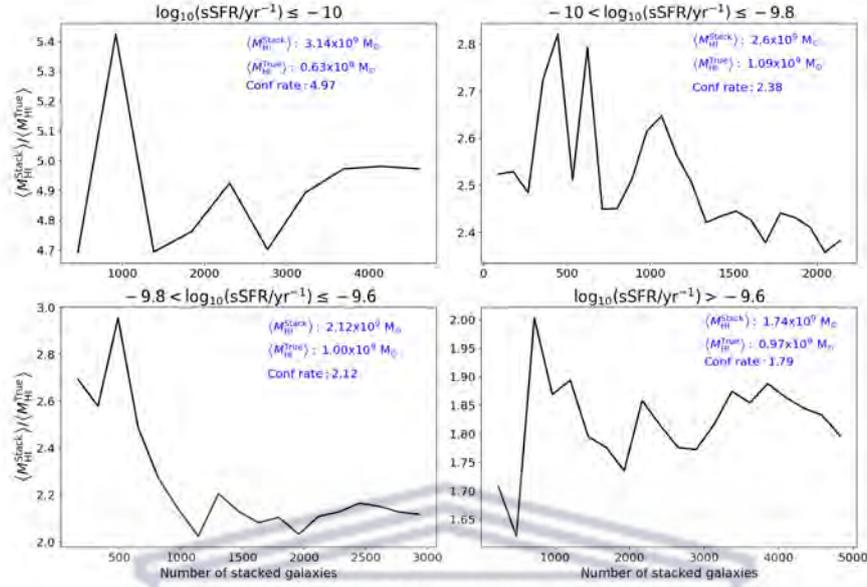


Figure 4.13: The confusion rate for each sSFR bin as a function of the number of stacked galaxies. The title of each panel indicates the corresponding sSFR bin of the galaxies. All the panel shows how the confusion rate changes as we stack more and more galaxies in the sample. The panels also report the average HI mass evaluated from the stacks after stacking all the galaxies in each bin, the true average HI mass and the confusion rate.

sSFR bins	N_{gal}	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	Confusion rate
$\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$	4610	0.63	3.14	4.97
$-10 < \log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -9.8$	2137	1.09	2.60	2.38
$-9.8 < \log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -9.6$	2936	1.00	2.12	2.12
$\log_{10}(\text{sSFR}/\text{yr}^{-1}) > -9.6$	4821	0.97	1.74	1.79

Table 4.4: The resulting parameters for all the sSFR bins.

4.5 Stacking in blue and red galaxies

In this section of the study, we present the stacking results for the blue and red galaxy sub-samples obtained from our HI data cubes. To distinguish between red and blue galaxies, we adopt the method implemented by [Jian et al. \(2018\)](#). In their study, they utilized the HSC CAMIRA cluster catalog and the photo- z galaxy catalog constructed in the HSC wide field (S16A), covering approximately $\sim 174 \text{ deg}^2$, to explore the star formation activity of galaxies in different environments within the redshift range of $0.2 < z < 1.1$. The method they employed to separate the

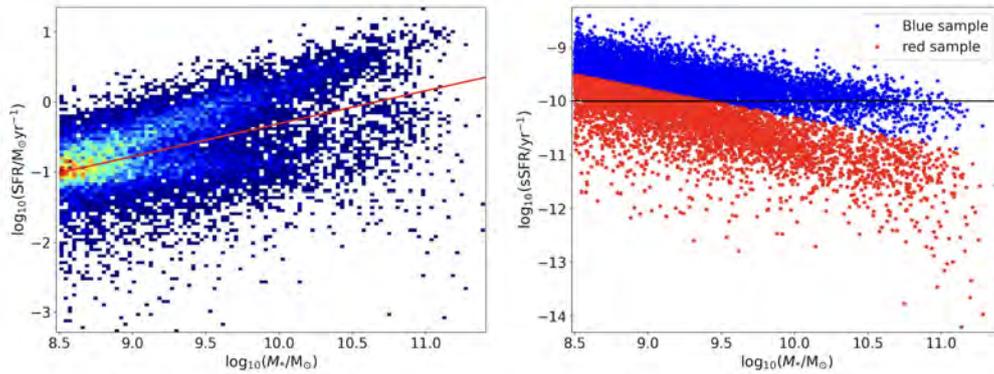


Figure 4.14: Shows red and blue galaxies in our sample separated using the bimodality between blue late-type and red early-type galaxy morphology, as demonstrated by [Jian et al. \(2018\)](#). The red galaxies are represented by red-shaded stars, while the blue galaxies are represented by blue-shaded stars.

star-forming and quiescent populations involves placing the galaxy sample on the SFR- M_* plane and fitting a best-fit line to the data points. This best-fit line effectively separates the star-forming galaxies (above the line) from the quiescent galaxies (below the line), with $\log_{10}(\text{sSFR}/\text{yr}^{-1}) = -10.1$ for $0.2 < z < 0.5$, -10 yr^{-1} for $0.5 < z < 0.8$, and -9.9 yr^{-1} for $0.8 < z < 1.1$.

We followed the same methodology in our study. Figure 4.14 (left panel) displays the SFR- M_* plane for our galaxy sample, with the best-fit line represented by the red curve. Given the redshift range of our galaxies ($0.7 < z < 0.758$), this line effectively separates our sample into star-forming galaxies (above the line) and quiescent galaxies (below the line), with $\log_{10}(\text{sSFR}/\text{yr}^{-1}) = -10$, in $0.5 < z < 0.8$.

The right panel of Figure 4.14 depicts our galaxy sample on the sSFR- M_* plane. By examining the distribution of data points in the plane, a clear bimodality is observed. We observe red galaxies located above the horizontal black line at $\log_{10}(\text{sSFR}/\text{yr}^{-1}) = -10$ and blue galaxies situated below this line. This raises the question of how such a distribution arises. While it is expected that most red-sequence galaxies exhibit minimal or no ongoing star formation, and most blue galaxies are actively forming stars, there exist significant populations of red star-forming galaxies and blue passive galaxies. [Mahajan & Raychaudhury \(2009\)](#) demonstrate that blue inactive galaxies exhibit extremely similar features to their star-forming counterparts, with the exception that their wide range in $H\delta$ equivalent width suggests a recent truncation of star production. The red star-forming galaxies can be further divided into two major groups. The first group comprises large galaxies found in cluster cores, characterized

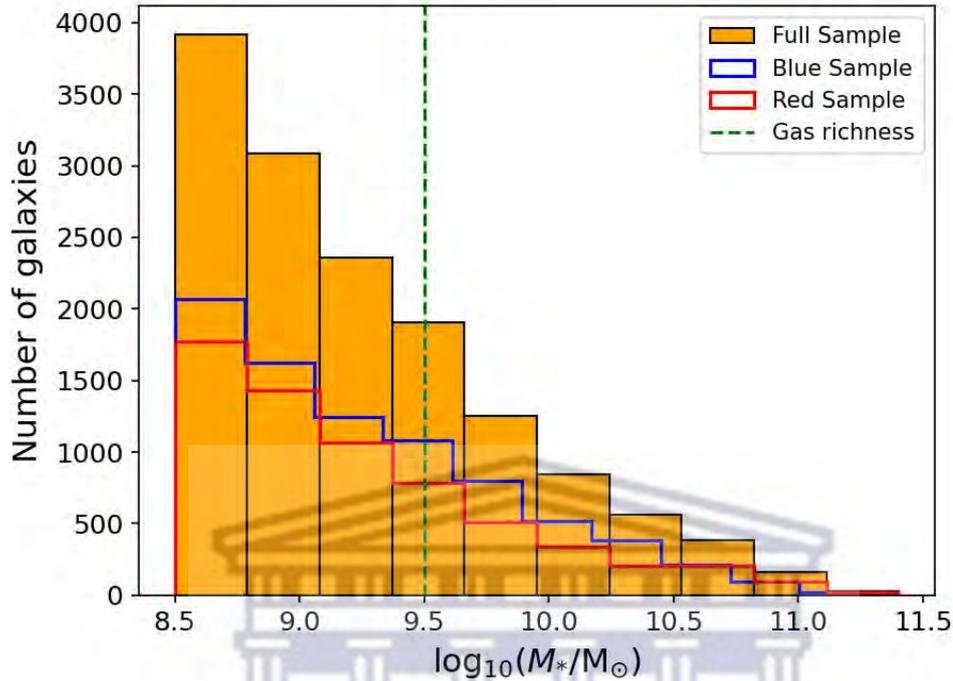


Figure 4.15: Shows the distribution of the galaxy’s colour in stellar mass. The orange histogram represents the full-sample distribution, while the blue unfilled histogram represents the blue-sample distribution and the red unfilled histogram represents the red-sample distribution. The green vertical dashed line highlights the gas-richness threshold (Kannappan et al., 2013; Healy et al., 2019), where galaxies with $\log_{10}(M_{\star}/M_{\odot}) < 9.5$ are gas-rich and galaxies with $\log_{10}(M_{\star}/M_{\odot}) > 9.5$ are gas-poor.

by an aging stellar population yet displaying signs of active star formation (possibly linked to active galactic nuclei). It is evident from spectral indices, stellar and gas-phase metallicities, and mean stellar ages that the remaining red star-forming galaxies’ red colors are primarily attributed to a predominantly metal-rich stellar population.

Figure 4.15 illustrates the distribution of red and blue galaxies in terms of their stellar mass. The orange histogram represents the full sample, while the blue unfilled histogram represents the blue galaxy sub-sample and the red unfilled histogram represents the red galaxy sub-sample. Overall, our sample comprises 6,435 red galaxies and 8,069 blue galaxies. The vertical green dashed line highlights the gas-richness threshold (Kannappan et al., 2013; Healy et al., 2019), with galaxies having $\log_{10}(M_{\star}/M_{\odot}) < 9.5$ classified as gas-rich and galaxies with $\log_{10}(M_{\star}/M_{\odot}) > 9.5$ categorized as gas-poor. The distribution reveals fewer red gas-poor galaxies and a

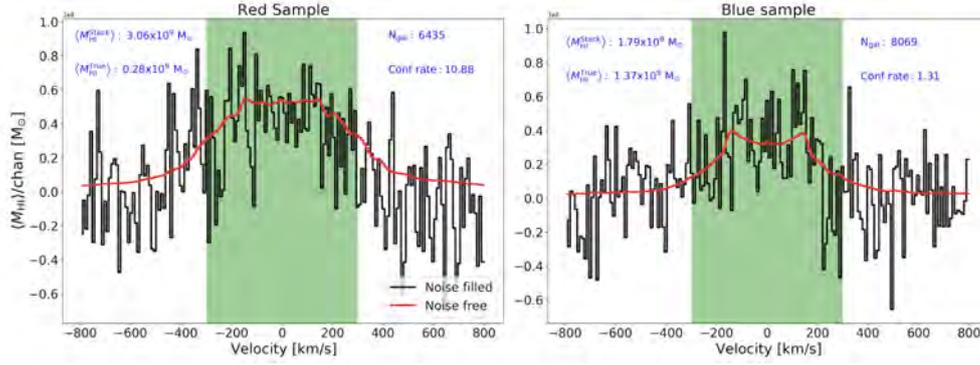


Figure 4.16: The stacking results for the blue and red samples. The title of each panel indicates the corresponding galaxy sample. The black and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-filled and noise-free version of the cubes. Each panel reports the average evaluated HI mass from the stacks, $\langle M_{\text{HI}}^{\text{Stack}}/M_{\odot} \rangle$, the true average HI mass, $\langle M_{\text{HI}}^{\text{True}}/M_{\odot} \rangle$, the number of stacked galaxies, N_{gal} , the confusion rate and the signal-to-noise ratio.

larger number of red gas-rich galaxies. As expected, we observe a larger number of blue gas-rich galaxies and fewer blue gas-poor galaxies. Across nearly all the bins, the distribution indicates that red galaxies are less abundant than blue galaxies, except for $\log_{10}(M_{\star}/M_{\odot}) \gtrsim 10.5$.

We performed stacking analyses on the blue and red galaxy sub-samples using the methodology explained in Chapter 3. The stacking results for the blue and red sub-samples are presented in Figure 4.16. Evidently, red galaxies exhibit a higher confusion rate compared to blue galaxies. Several characteristics of red and blue galaxies contribute to the differences in confusion rates observed when stacking their spectra.

- Stellar populations: Red galaxies typically possess older stellar populations and lower levels of ongoing star formation compared to blue galaxies. The presence of older stellar populations can result in weaker spectral features, making it more challenging to distinguish the true HI signal (i.e. ± 300 km/s or green region) from noise. Consequently, this may lead to a higher confusion rate in the stacked spectra of red galaxies.
- Environmental effects: Red galaxies are frequently found in dense environments such as galaxy clusters or groups, where the presence of neighboring galaxies can contribute to contamination and confusion in the stacked spec-

Colour samples	N_{gal}	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	Confusion rate
Red sample	743	0.28	3.06	10.88
Blue sample	13761	1.37	1.79	1.31

Table 4.5: The resulting parameters for red and blue galaxy samples.

tra. The additional galaxies in proximity to red galaxies can increase the confusion rate in their stacked spectra compared to blue galaxies, which are typically situated in less dense environments.

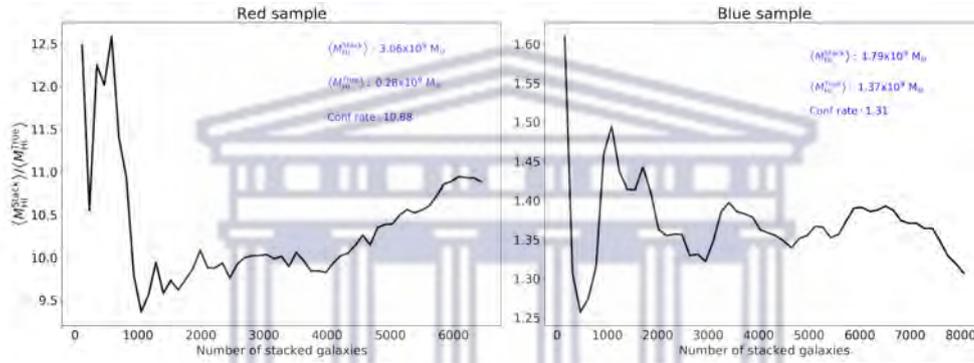


Figure 4.17: The confusion rate for red and blue samples as function of the number of stacked galaxies. The title of each panel indicates the corresponding colour sample of the galaxies. Both panels show how the confusion rate changes as we stack more and more galaxies in the sample. The panels also report the average HI mass evaluated from the stacks after stacking all the galaxies in each sample, the true average HI mass and the confusion rate.

The red sub-sample exhibits a contaminant mass that is approximately eight times greater than that of the blue sub-sample, signifying a significant difference. Furthermore, when examining the characteristic shapes of the stacked spectra, it is evident that the extended wings beyond the velocity range of ± 300 km/s for the red sub-sample are more pronounced and extensive compared to those of the blue sub-sample. This might be an indication that the red sub-sample has a larger number of nearby neighbors (thus more contaminant mass), resulting in longer and larger wings. Additionally, when observing the shapes of the spectra within the velocity range of interest (± 300 km/s), the blue sub-sample displays a double-horn shape, while the red sub-sample exhibits an almost flat shape. The most probable reason for the flat spectrum shape is that one horn of the spectra, which contributes contamination to the red sub-sample and causes the extended wings, falls within the

velocity range of ± 300 km/s. These contaminants accumulate during the stacking process, resulting in higher contaminant mass in the green region.

These implications highlight the importance of considering neighboring galaxies and environmental effects when studying the HI content of galaxies through stacking analyses. The differences observed in confusion rates and spectral shapes between red and blue galaxies provide insights into the complexities of interpreting stacked spectra and understanding the impact of neighboring galaxies on the HI measurements. Figure 4.17 demonstrates how the confusion rate changes as more galaxies are stacked in the red sub-sample (left panel) and the blue sub-sample (right panel). Table 4.5 presents the parameters evaluated from the simulated data (true parameters) as well as the parameters evaluated from the stacks for each color sub-sample.

4.6 Comparing confusion rate for all sub-samples

In this chapter, we perform stacking experiments on galaxies in different sub-samples, specifically targeting sub-samples that are likely to influence the regulation of HI gas content in galaxies. The purpose is to conduct a thorough investigation into how high-redshift stacking experiments may be affected by the incompleteness of optical catalogues. Stacking in various galaxy sub-samples enables us to understand which sub-samples exhibit higher confusion rates and which ones have lower contamination rates during stacking experiments. Although the number of galaxies in each sub-sample bin is not equal, as evident in Table 4.6, we are not concerned with the sample sizes, as previously mentioned, since they have no impact on the confusion rate.

Upon examining all the sub-samples stacked in this chapter, we find that galaxies with small stellar masses ($\log_{10}(M_{\star}/M_{\odot}) < 9.5$, i.e., bin1 and bin2), galaxies with lower star-formation rates ($\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) \leq -0.5$, i.e., bin1 and bin2), galaxies with lower specific star-formation rates ($\log_{10}(\text{sSFR}/\text{yr}^{-1}) \leq -10$), and red galaxies exhibit confusion rates greater than 3. These high confusion rates might be indicating a bias in the catalog towards galaxies with these specific properties. The catalog used in our study consists of high-redshift galaxies and is magnitude-limited, leading to a bias towards these particular types of galaxies. The likely reason for the high confusion rate in these sub-samples is their association with denser environments, where an abundance of nearby galaxies and spatial clustering contribute to

Samples	N_{gal}	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	Confusion rate
All galaxies				
Full sample	14504	0.89	2.34	2.64
Stellar mass, M_{\star}				
bin1	6283	0.37	2.65	5.70
bin2	3978	0.57	2.28	3.98
bin3	2410	1.09	2.65	2.43
bin4	1830	1.53	2.87	1.88
Star-formation rate				
bin1	4931	0.19	3.01	16.03
bin2	4426	0.40	2.26	5.62
bin3	2837	0.91	1.75	1.91
bin4	2310	1.63	2.16	1.32
Specific star-formation rate				
bin1	4610	0.63	3.14	4.97
bin2	2137	1.09	2.60	2.38
bin3	2936	1.00	2.12	2.12
bin4	4821	0.97	1.74	1.79
Colour				
Red	743	0.28	3.06	10.88
Blue	13761	1.31	1.79	1.31

Table 4.6: The resulting parameters for all the sub-samples in this work.

contamination. Additionally, some of these sub-samples are likely located near larger galaxies with high flux densities, further contributing to the higher contamination rates observed.

On the other hand, we find that galaxies with stellar masses ($\log_{10}(M_{\star}/M_{\odot}) > 9.5$, i.e., bin3 and bin4), galaxies with higher star-formation rates ($\log_{10}(\text{SFR}/M_{\odot}\text{yr}^{-1}) > -0.5$, i.e., bin3 and bin4), galaxies with higher specific star-formation rates ($\log_{10}(\text{sSFR}/\text{yr}^{-1}) > -10$, i.e., bin2, bin3, and bin4), and blue galaxies exhibit confusion rates below 3. Therefore, for future studies investigating the HI content of any high-redshift catalog using the stacking method, it is recommended that authors consider stacking sub-samples that yield stacked spectra with confusion rates below 3. This might be because the high-redshift catalog is not biased towards these specific sub-samples, as demonstrated in this study.

5 HI content of galaxies

5.1 Group vs Non-Group galaxies

Studying the HI gas content of galaxies in different environments is crucial for understanding how the surrounding conditions influence galaxy evolution. In this part of the study, we aim to investigate the impact of different environments on the HI gas content of galaxies in our sample. To classify galaxies into group and non-group sub-samples, we employed the friends-of-friends algorithm (for example: [Janowiecki et al., 2020](#); [Liu et al., 2008](#); [Robotham et al., 2011](#); [Rhee et al., 2023](#)). This algorithm allows us to identify galaxies that belong to galaxy groups or clusters based on their spatial proximity and shared velocities.

By dividing our sample based on their environment, we can explore how the presence of neighboring galaxies, the density of the environment, and other factors associated with galaxy groups affect the HI gas content of galaxies. This analysis enables us to assess whether environmental processes, such as interactions, tidal forces, and gas stripping, play a significant role in shaping the HI properties of galaxies.

Understanding the relationship between environment and HI gas content provides valuable insights into the mechanisms driving galaxy evolution. It helps us decipher the influence of environmental factors on gas accretion, star formation, gas depletion, and the overall transformation of galaxies over cosmic time.

The friends-of-friends algorithm connects the galaxy's physical quantities, such as position ([Huchra & Geller, 1982](#)). If the galaxies are in the same group they must be spatially linked and consequently they will be defined as friends-of-friends. The friends-of-friends algorithm has two criteria ([Liu et al., 2008](#)):

$$D_{i,j} \leq D_L, \quad (5.1)$$

$$V_{i,j} \leq V_L, \quad (5.2)$$

that are used to conclude if two galaxies are physically linked, where $D_{i,j}$ is the

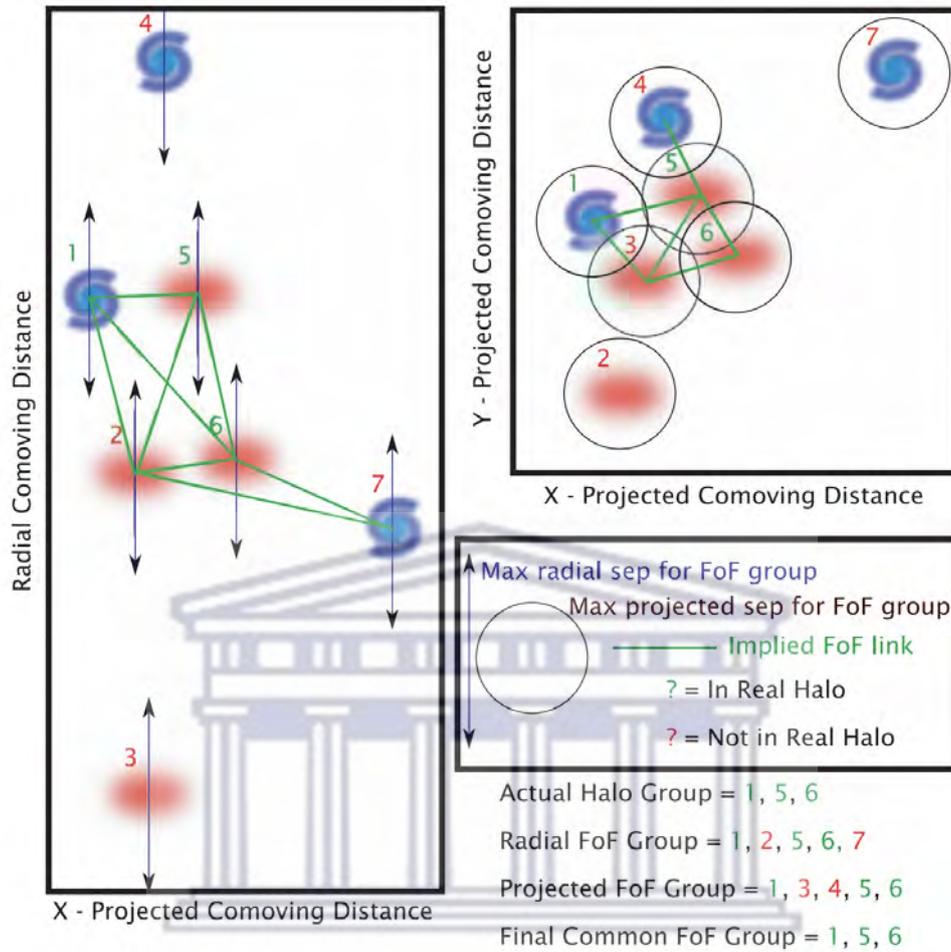


Figure 5.1: Schematic that demonstrates how the friends-of-friends algorithm works when linking galaxies on the redshift survey data (Janowiecki et al., 2020; Liu et al., 2008; Robotham et al., 2011; Rhee et al., 2023). The galaxies are shown along the line of sight on the left panel, while the right panel shows the same set of galaxies projected on the sky. The radial separation together with the projected separation are used to link galaxies via friends-of-friends algorithm image credit to Robotham et al. (2011).

separation distance between galaxies i, j in the projected direction and $V_{i,j}$ is the separation distance of galaxies i, j in the line-of-sight direction. D_L and V_L are the linking lengths of galaxies i, j and they are commonly functions of redshift and other local physical properties. Figure 5.1 is the schematic of how the algorithm works (Robotham et al., 2011). If we take the galaxy labeled 1 to be the host galaxy (i.e. target galaxy), the galaxies labeled 1, 2, 5, 6, 7 will fulfill the radial friends-of-friends group criteria, ($V_{i,j} \leq V_L$). The galaxies labeled 1, 3, 4, 5, 6 will fulfill the projected friends-of-friends criteria, ($D_{i,j} \leq D_L$). As aforementioned the galaxies must fulfill

both criteria for them to be considered to be in a group, therefore only galaxies labeled 1, 5, 6 fulfill both criteria, hence they form a group.

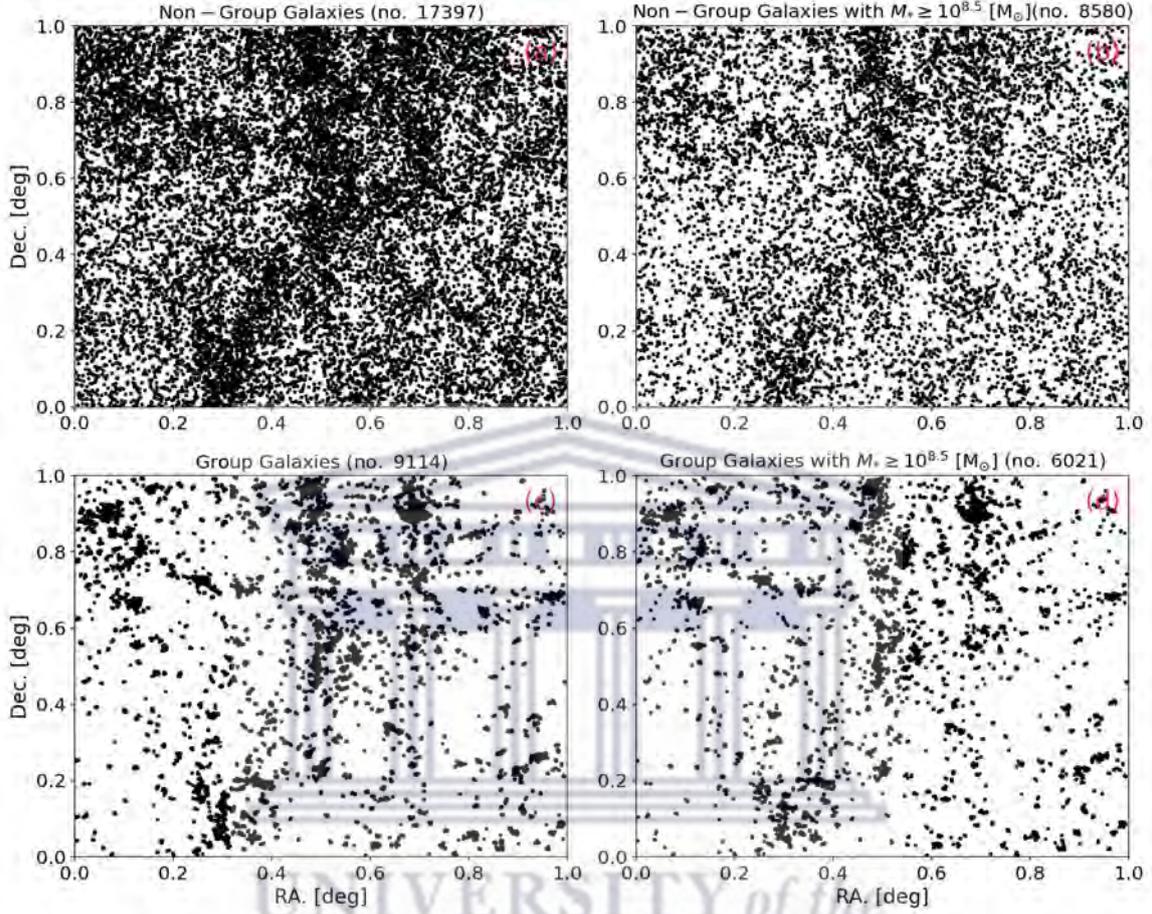


Figure 5.2: RA and Dec - projected co-moving distance that links galaxies of our sample via friends-of-friends algorithm. Panel (a) show the RA- and Dec-projection of the non-group galaxies (galaxies that do not meet both friends-of-friends algorithm criteria), and panel (b) show the same non-group galaxies after flagging out all the galaxies with $M_*/M_\odot < 10^{8.5}$, because the catalog is incomplete there. Panel (c) and (d) respectively show RA- and Dec-projection of the group galaxies before and after flagging out galaxies with $M_*/M_\odot < 10^{8.5}$.

After following the friends-of-friends procedure we were able to divide our sample into non-group and group galaxy sub-samples. Shown in figure 5.2 are the non-group and group galaxy sub-samples. Liu et al. (2008); Robotham et al. (2011); Janowiecki et al. (2020); Rhee et al. (2023) defined a galaxy to be in a group if it has two or more group members that fulfills both friends-of-friends criteria. In this work we use the same way to define a galaxy to be part of the group or not. Panel (a) and (c) shows the RA- and Dec projected co-moving distance for 17397 non-group and 9114 group

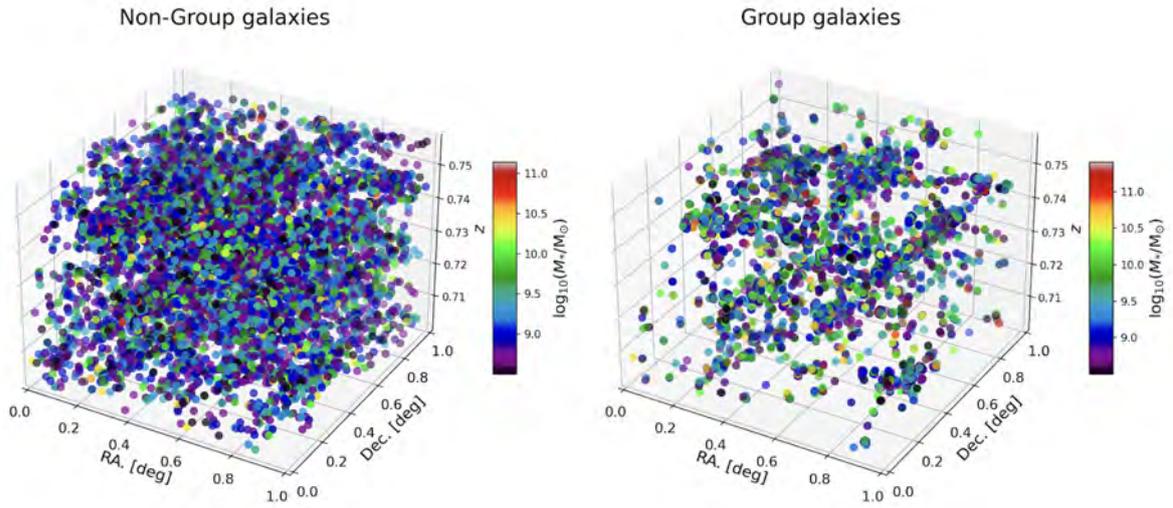


Figure 5.3: Shows the 3 dimensional cube of the non-group and group galaxies. The title of each panel indicates the corresponding sub-sample. The panels show the galaxies along the line of sight and their projection on the sky. Galaxies in the cubes are colour coded with stellar masses. Both panels are populated with galaxies with $M_*/M_\odot < 10^{10}$.

galaxies, respectively, before flagging out galaxies with $M_*/M_\odot < 10^{8.5}$ because the catalog is incomplete below this mass limit as aforementioned. Panel (b) and (d) shows the RA- and Dec-projected co-moving distance for 8580 non-group and 6021 group galaxies, respectively, after flagging out galaxies with $M_*/M_\odot < 10^{8.5}$.

Figure 5.3 shows the 3 dimensional cube of the non-group (left panel) and group galaxies (right panel). The x-, y- and z-axis of the cubes, respectively, represents the RA [deg], Dec [deg] and z -redshift. The circles inside the cubes represent galaxies in each of the two sub-samples and they are colour coded with their stellar mass $\log_{10}(M_*/M_\odot)$. Both sub-samples are mostly populated with galaxies with $\log_{10}(M_*/M_\odot) < 10$, that is mainly because our simulation has few galaxies with $\log_{10}(M_*/M_\odot) \geq 10$.

Figure 5.4 illustrates the properties of non-group and group galaxies. Panel (a) displays the distribution of non-group and group galaxies in terms of stellar mass. It is evident that the non-group galaxies dominate below $\log_{10}(M_*/M_\odot) \sim 10$, while the group galaxies become more prevalent at $\log_{10}(M_*/M_\odot) \geq 10$. Furthermore, a significant portion of non-group galaxies is found below $\log_{10}(M_*/M_\odot) \sim 9.5$, indicating a higher gas content in the non-group sample compared to the group sample.

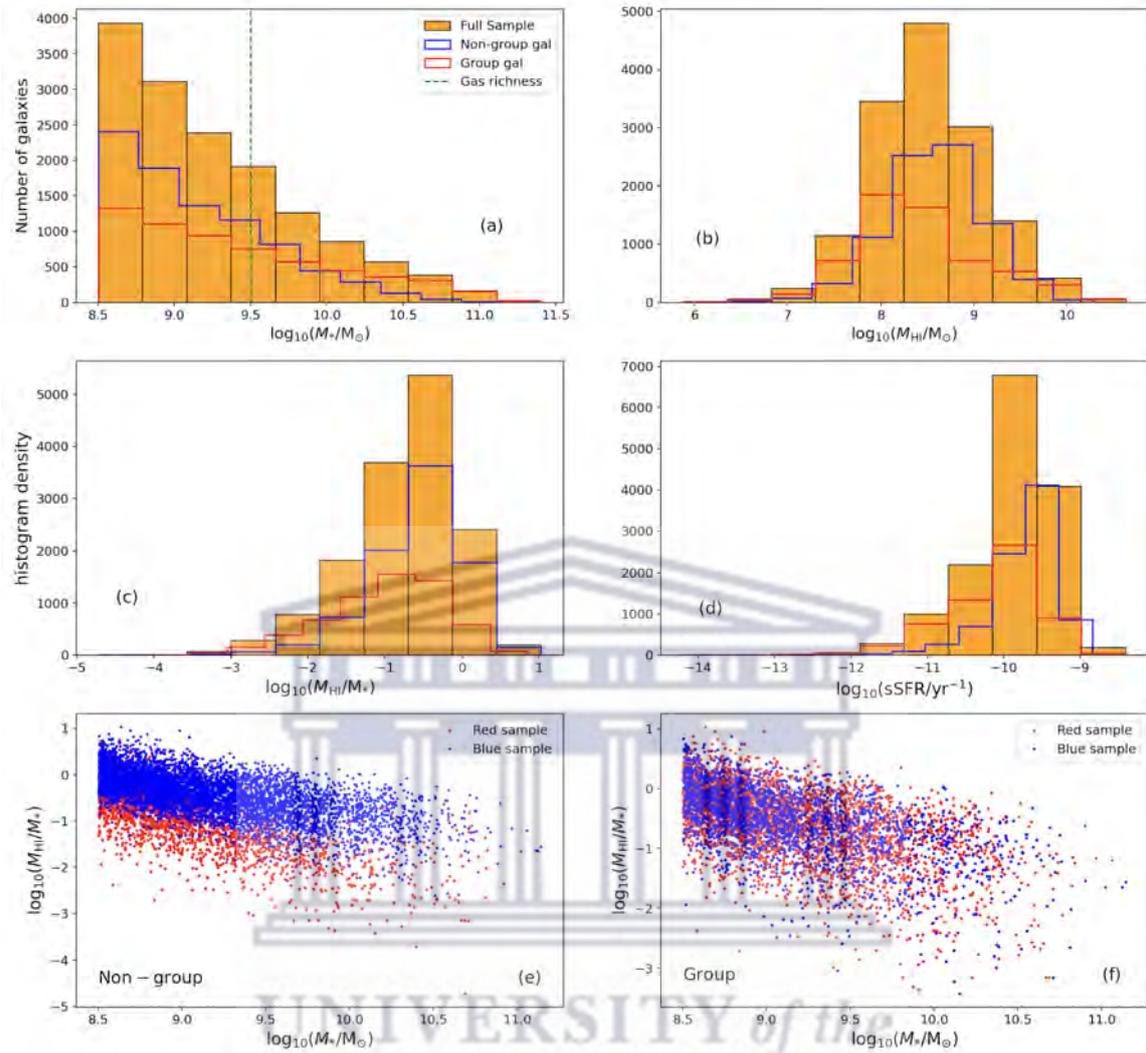


Figure 5.4: Evaluated galaxy properties for the non-group and group galaxies. In panels (a)-(d) the orange histogram represent the full sample. The non-group sample is presented by the blue-unfilled histogram, while group sample is presented by the red histogram. Panel (a) shows the stellar mass distribution. In panel (b) we show HI mass distribution, while in panel (c) and (d) we show the distribution of HI gas fraction and specific star-formation rate. Panel (e) and (f) show the HI mass vs stellar mass relation for the non-group and group galaxies, respectively. In panel (e) and (f) galaxies are colour coded according to their colour.

In panel (b), the HI mass distribution of both environments is depicted. The vertical green-dashed line represents the threshold for "gas-richness," where galaxies below this line are considered gas-rich and those above are gas-poor. The distribution clearly shows that the majority of HI gas is concentrated in the non-group galaxies. Panel (c) showcases the distribution of HI gas fraction in non-group and group galaxies. As mentioned earlier, the non-group galaxies are richer in gas compared

to the group galaxies, and this distribution serves as evidence that most of the gas is contained within the non-group galaxies.

Moving to panel (d), the sSFR distribution for both environments is displayed. It is noticeable that group galaxies exhibit higher sSFR than non-group galaxies at $\log_{10}(\text{sSFR}/M_{\odot}\text{yr}^{-1}) < -10.5$. Consequently, this suggests that non-group galaxies are relatively inactive in their environment, as they possess small stellar masses, a higher gas content, but low sSFR. Panels (e) and (f) present the HI gas fraction vs. stellar mass relation for non-group and group galaxies, respectively, with the color indicating their classification as blue or red galaxies. These panels reveal that the non-group sample primarily consists of blue galaxies (6028 blue and 2552 red galaxies), while the group sample is predominantly composed of red galaxies (2107 blue and 3914 red galaxies).

After dividing our sample into sub-samples we stacked their HI emission line for the two sub-samples in order to study the HI gas content of the galaxies in different environments. Figure 5.5 shows the stacking results for the two sub-samples. The stacked spectrum is generated following the procedure described in chapter 3. Panel (a) is the resulting co-added spectrum for the Non-group galaxies, while panel (b) is the resulting co-added spectrum for the group galaxies. Each panel reports the number of galaxies stacked and the confusion rate in each sub-sample.

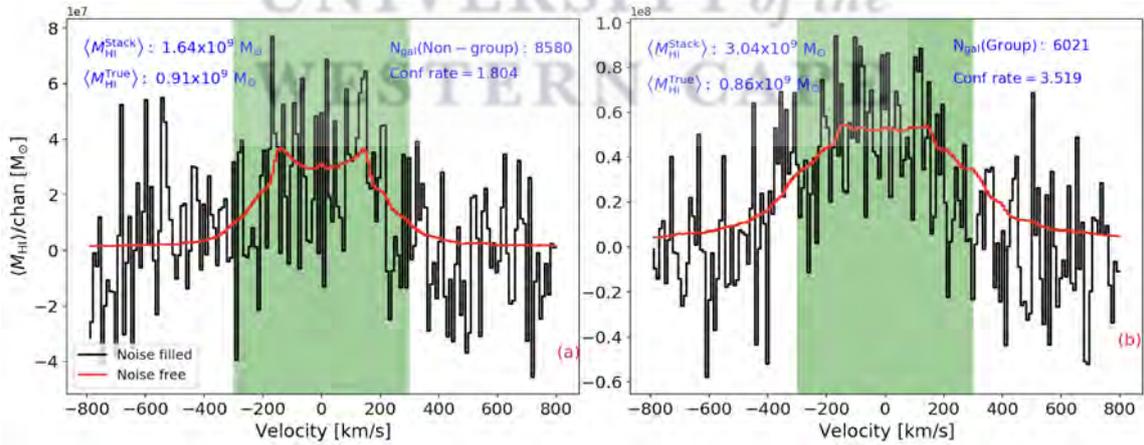


Figure 5.5: Stacking results for non-group (panel (a)) and group galaxies (panel (b)). The red and blue curve respectively, represent the stacked spectrum from the sub-volumes extracted from the noise-free and noise-filled versions of our cubes. Both panels report the number of galaxies stacked together and the confusion rate. Other resulting parameters are presented in table 5.1.

The non-group galaxies are HI richer compared to group galaxies as shown in table 5.1. This is in agreement with results from Rhee et al. (2023); Brown et al. (2017); Stevens et al. (2019). Rhee et al. (2023) presents the early science results from Deep Investigation of Neutral Gas Origins (DINGO), using the Australian Square Kilometre Array Pathfinder (ASKAP) (Duffy et al., 2012; Ball, 2009). They study the environmental impacts on HI gas content of galaxies in different environments at redshift range $0.04 < z < 0.09$ and found that an average HI mass of group central galaxies is larger than those of satellite and isolated galaxies, but with a lower HI gas fraction. Brown et al. (2017) examine environment-driven gas depletion in satellite galaxies. They use galaxies at redshift range $0.02 \leq z \leq 0.05$ from the Sloan Digital Sky survey (SDSS), with HI data from the Arecibo Legacy Fast ALFA survey (Giovanelli et al., 2005; Haynes et al., 2018) and found that central galaxies are more HI-rich than satellites. While Stevens et al. (2019) examine the influence of the environment on the cold-gas properties of galaxies at redshift $z = 0$ within the TNG100 cosmological, magnetohydrodynamic simulation, part of the illustrisTNG suite and found that satellites are ~ 3 times HI poorer than centrals of the same stellar mass.

We note that when we look at the evaluated average HI mass for the stacks, $\langle M_{\text{HI}}^{\text{Stack}} \rangle$, the group galaxies appear to be HI more massive than non-group galaxies. These results did not surprise us, because the stacking results suffer from confusion and that results with the overestimates in the evaluated average HI mass in both sub-samples. It turns out that group galaxies have a high confusion rate that is why we see these results. The reason group galaxies have high confusion rate is because they have a lot of nearby galaxies and thus that results in the high contamination rate on the extracted host galaxy sub-volumes. If our analysis were based only on the average HI mass evaluated from the stacking results our findings would be in contrary with the literature (for example: Rhee et al., 2023; Brown et al., 2017; Stevens et al., 2019) unless adequate confusion correction method is used (for example: Elson et al., 2016, 2019b; Healy et al., 2019; Healy, 2021; Sinigaglia et al., 2022). The non-group sub-sample average HI gas fraction (i.e. $f_{\text{HI}}^{\text{True}}$ and $f_{\text{HI}}^{\text{Stack}}$) is higher than the group galaxy. The lower average HI gas fraction for group sub-sample probably arises from the fact that they have higher stellar mass and lower average HI mass.

To investigate this behavior in detail, we investigate how the HI gas content of galaxies in different environments will differ with physical properties. We did that

Sub-samples	N_{gal}	$\langle M_{\star} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	$f_{\text{HI}}^{\text{True}}$	$f_{\text{HI}}^{\text{Stack}}$	Confusion rate
Non-group	7743	3.91	0.91	1.64	0.23	0.42	1.80
Group	6858	7.91	0.86	3.04	0.11	0.39	3.52

Table 5.1: The resulting parameters from the simulated data (true results) and from the stacked spectrum (stack results).

investigation using the stellar mass and specific star formation rate as in [Rhee et al. \(2023\)](#). Figure 5.6 shows the stellar mass of galaxies, M_{\star}/M_{\odot} , plotted against the specific star formation rate, $\text{sSFR}/\text{yr}^{-1}$, and it is colour coded with $M_{\text{HI}}^{\text{True}}$. The horizontal black dashed line separates the blue to the red galaxies, with $\text{sSFR}/\text{yr}^{-1} \geq 10^{-10}$ in $0.5 < z < 0.8$. The galaxies with $\text{sSFR}/\text{yr}^{-1} < 10^{-10}$ are considered to be red galaxies and those with $\text{sSFR}/\text{yr}^{-1} \geq 10^{-10}$ are the blue galaxies ([Jian et al., 2018](#)). The non-group sub-sample is populated with blue galaxies, while the group sample is populated with red galaxies. The non-group sub-sample has fewer red galaxies than the group sub-sample. In both sub-samples the galaxies with $\log_{10}(M_{\text{HI}}/M_{\odot}) > 8.5$ are populated on and above the star-formation main sequence (i.e. the blue solid line). We divided both non-group and group sub-samples into galaxy sub-classes. The sub-classes include star-formation main sequence (SFMS), starburst (SB), transition zone (TZ), and quiescent (QS) galaxies. The blue solid line represents the main sequence, it was generated by fitting equation 5.3 from [Janowiecki et al. \(2020\)](#) to the sample:

$$\log \frac{\text{SFR}_{\text{MS}}}{M_{\star}} (\text{yr}^{-1}) = m_{\text{SFMS}} \left(\log \frac{M_{\star}}{M_{\odot}} - 9 \right) + b_{\text{SFMS}}, \quad (5.3)$$

where , m_{SFMS} , is the slope of the line and b_{SFMS} is the normalisation constant.

The black solid lines below and above the SFMS are ± 0.3 dex ($\sim 1\sigma$) away from SFMS. The galaxies within this range are considered to be the star-formation main sequence galaxies (SFMS) and the galaxies above SFMS are defined to be starbursts (SB). To determine quiescent (QS) galaxies, we again adopt from [Janowiecki et al. \(2020\)](#) equation 5.4:

$$\Delta \text{SFMS} = \log \frac{\text{SFR}}{M_{\star}} (\text{yr}^{-1}) - \log \frac{\text{SFR}_{\text{MS}}}{M_{\star}} (\text{yr}^{-1}), \quad (5.4)$$

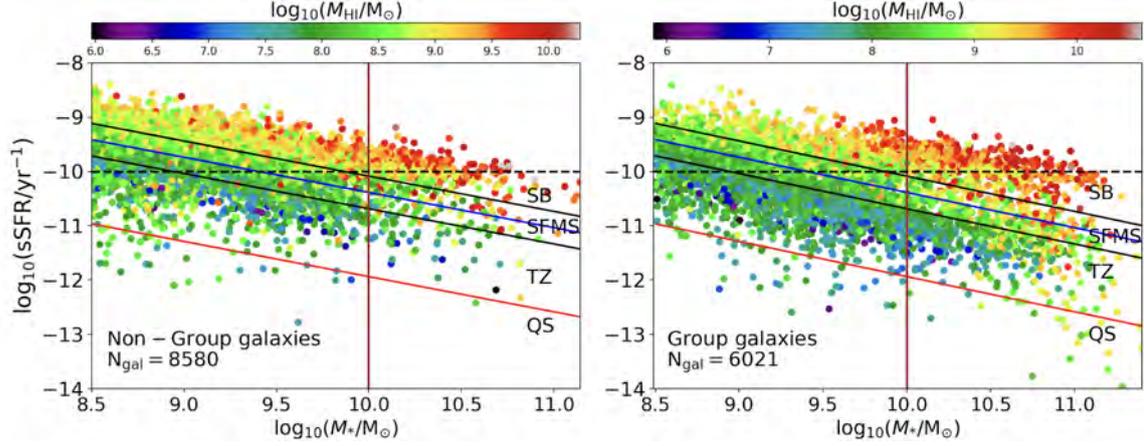


Figure 5.6: The stellar mass to specific star-formation rate plane for non-group and group galaxies. Sub-samples are colour coded with their true HI masses. Galaxies in this plane are separated into different sub-classes. The blue line represents the star-formation main sequence (SFMS) and all the galaxies within the black line above and below the blue line are regarded to be SFMS galaxies (for example [Janowiecki et al., 2020](#); [Rhee et al., 2023](#)). Galaxies above the SFMS are defined to be the starburst (SB) and the galaxies below the red line are quiescent (QS) galaxies, while galaxies that lie above QS and below the SFMS are the transition zone (TZ) galaxies. The horizontal black dashed line separates the blue to the red galaxies, with $\text{sSFR}/\text{yr}^{-1} \geq 10^{-10}$ in $0.5 < z < 0.8$. The galaxies with $\text{sSFR}/\text{yr}^{-1} < 10^{-10}$ are considered to be red galaxies and those with $\text{sSFR}/\text{yr}^{-1} \geq 10^{-10}$ are the blue galaxies ([Jian et al., 2018](#)). The vertical brown-solid line separates SB and SFMS in stellar mass at $\log_{10}(M_{\star}/M_{\odot}) = 10$.

and ΔSFMS is determined to be -1.55 dex and the galaxies below that line are the QS galaxies. The galaxies that lie between the SFMS and the QS are the transition zone galaxies (TZ). We also divided the SFMS and SB in galaxy stellar mass, with $M_{\star}/M_{\odot} < 10^{10}$ (i.e. SB(low) and SFMS(low)) and $M_{\star}/M_{\odot} \geq 10^{10}$ (i.e. SB(high) and SFMS(high)) to investigate the dependence of the HI properties on stellar mass. We stacked all the sub-samples to study the HI richness and HI gas content of these sub-classes. [Figure 5.7](#) and [5.8](#) shows the resulting stacks for non-group and group sub-classes, respectively. On each panel of the figures, the number of stacked galaxies and the confusion rate are reported.

Apparently the results in [table 5.2](#) agree with the global results. Indeed the average HI mass for the non-group sub-classes is higher than the group sub-classes. The $\langle M_{\star} \rangle$ for SB and SFMS for non-group sub-classes is higher than the group sub-classes, whereas the TZ and QS sub-classes have low $\langle M_{\star} \rangle$. From these results the non-group sub-sample is more dominant in HI gas reservoirs than group sub-sample.

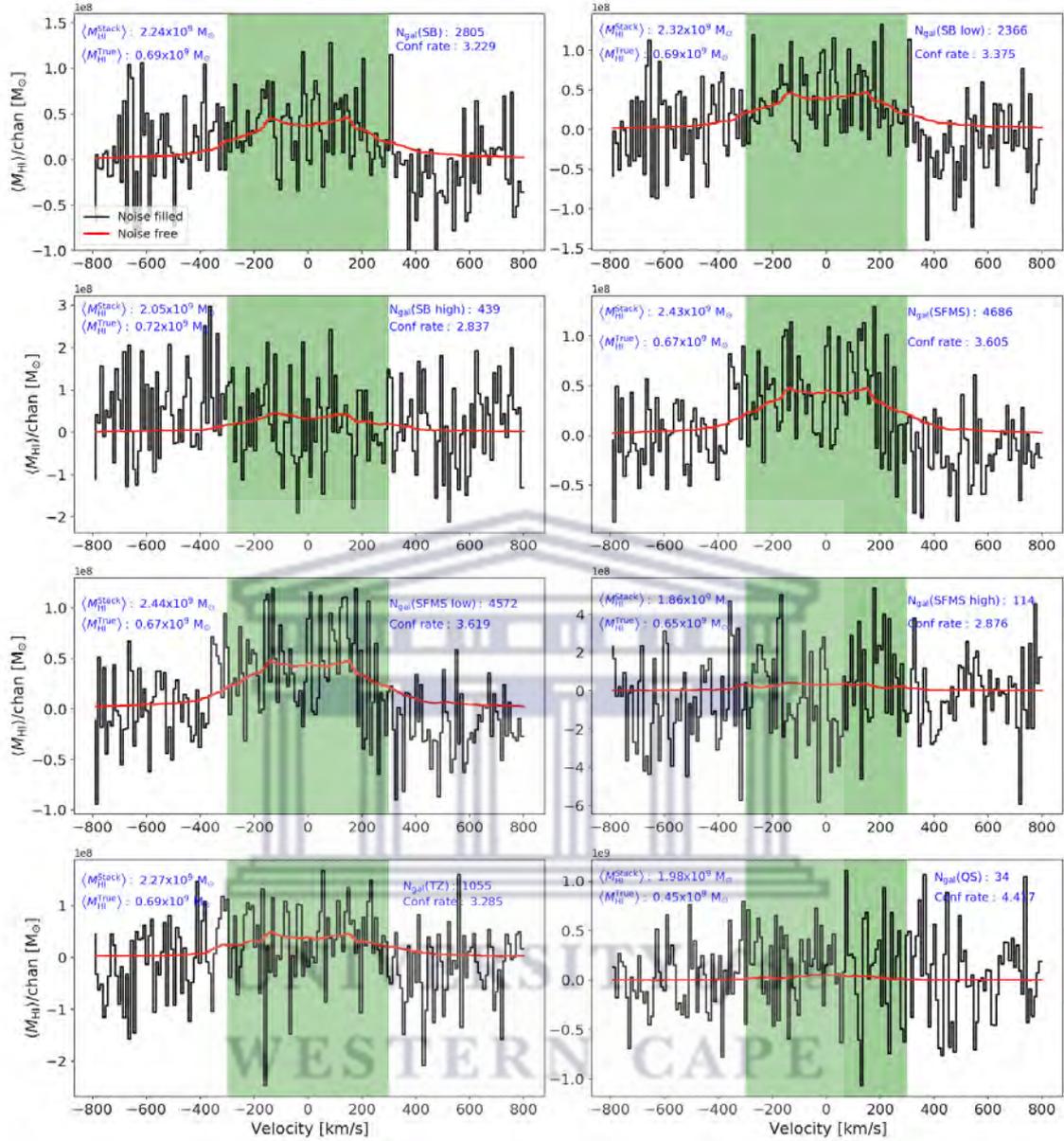


Figure 5.7: Stacking results for all the sub-classes of the non-group sub-sample. The red and black curve represent the stacked spectrum from the sub-volumes extracted from noise-free and noise-filled versions of our cubes, respectively. All the panels report the number of stacked galaxies in each sub-class and the confusion rate. Other parameters of the sub-classes are presented in table 5.2.

The lack of the HI gas in group sub-sample indicates that the group environment is affecting the gas loss from the galaxies (Kilborn et al., 2009). The probable cause of this effect is the tidal stripping of HI gas from disk galaxies as they are early processed in group environments. Because the SB and SFMS of non-group sub-sample has high $\langle M_{\star} \rangle$, this resulted in these sub-classes having low HI gas fraction

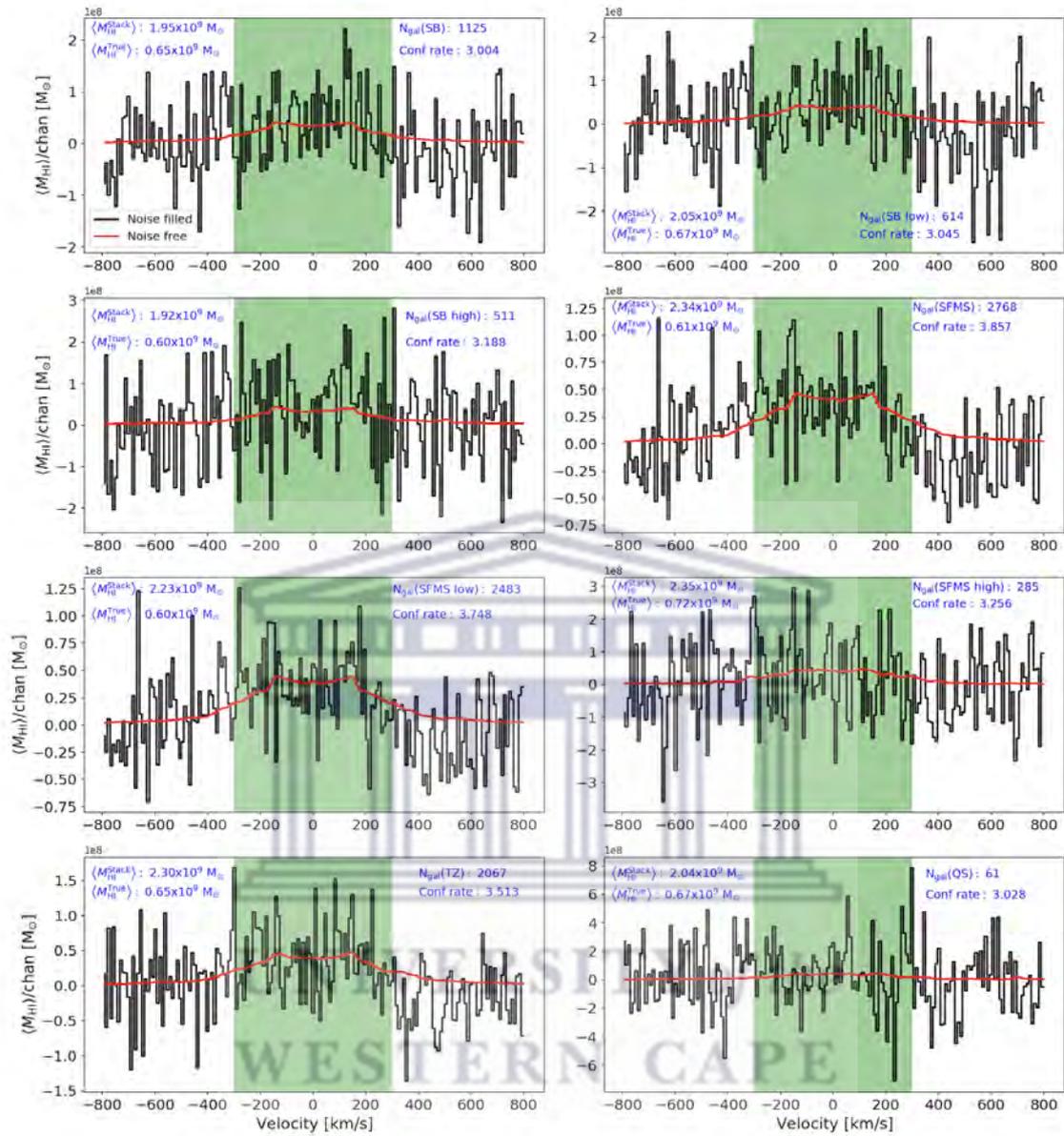


Figure 5.8: Stacking results for all the sub-classes of the group sub-sample. The red and black curve represent the stacked spectrum from the sub-volumes extracted from noise-free and noise-filled version of our cubes, respectively. All the panels report the number of stacked galaxies in each sub-class and the confusion rate. Other parameters of the sub-classes are presented in table 5.2.

when compared to SB and SFMS sub-classes of the group sub-sample. The TZ and QS sub-classes of the group sub-sample have low HI gas fraction because they have high $\langle M_{\star} \rangle$ than the same sub-classes in the non-group sub-sample. Indeed, most of the sub-classes of the group sub-sample have low HI gas fraction than non-group sub-classes, that is because they have higher $\langle M_{\star} \rangle$. These results agree with (for

example: [Brown et al., 2017](#); [Stevens et al., 2019](#)), while they are in contrast with [Rhee et al.](#) (for example: [2023](#)). The reason our results are in contrast with [Rhee et al. \(2023\)](#) is because they separated their group sub-sample into group centrals and isolated group centrals.

Sub-samples	N_{gal}	$\langle M_{\star} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{True}} \rangle$ [$10^9 M_{\odot}$]	$\langle M_{\text{HI}}^{\text{Stack}} \rangle$ [$10^9 M_{\odot}$]	$f_{\text{HI}}^{\text{True}}$	$f_{\text{HI}}^{\text{Stack}}$	Confusion rate
Non-group							
SB	2769	3.64	0.69	2.24	0.19	0.62	3.23
SB(low)	2215	3.64	0.69	2.32	0.19	0.64	3.38
SB(high)	554	3.66	0.72	2.05	0.20	0.56	2.84
SFMS	4094	3.68	0.67	2.43	0.18	0.66	3.61
SFMS(low)	3992	3.69	0.67	2.44	0.18	0.66	3.62
SFMS(high)	102	3.16	0.65	1.86	0.20	0.59	2.88
TZ	855	3.20	0.69	2.27	0.22	0.71	3.29
QS	25	2.01	0.45	1.98	0.22	0.99	4.42
Group							
SB	1161	3.21	0.65	1.95	0.20	0.61	3.00
SB(low)	765	3.26	0.67	2.05	0.21	0.63	3.05
SB(high)	396	3.11	0.60	1.92	0.19	0.62	3.19
SFMS	3360	3.48	0.61	2.34	0.18	0.67	3.86
SFMS(low)	3063	3.42	0.60	2.23	0.17	0.65	3.75
SFMS(high)	297	4.05	0.72	2.35	0.19	0.58	3.26
TZ	2267	3.84	0.65	2.30	0.17	0.60	3.51
QS	70	3.32	0.67	2.04	0.20	0.61	3.03

Table 5.2: Present the resulting parameters for non-group and group sub-samples sub-classes.

When we compare sub-classes in each environment we find that SB is HI richer than SFMS, TZ and QS in both environments. In the group sub-sample the SB also has a HI gas fraction higher than SFMS, TZ and QS. These results agree with (for example: [Rhee et al., 2023](#)). Interestingly, in the non-group sub-sample the SB sub-class has lower HI gas fraction than TZ and QS. The reason for this is that SB has higher $\langle M_{\star} \rangle$. When comparing the sub-classes SB and SFMS separated in stellar mass, the $\langle M_{\text{HI}} \rangle$ is correlated with the $\langle M_{\star} \rangle$, while $\langle f_{\text{HI}} \rangle$ is anti-correlated with the $\langle M_{\star} \rangle$ in all cases. We further computed the average star-formation rate, $\langle \text{SFR}/M_{\odot}\text{yr}^{-1} \rangle$, and the average time depletion timescale, $\langle t_{\text{dep}}/\text{Gyr} \rangle = \langle M_{\text{HI}} \rangle / \langle \text{SFR} \rangle$ presented in table 5.3. Furthermore, we compared the $\langle \text{SFR} \rangle$ and $\langle t_{\text{dep}} \rangle$ for the sub-samples SB and SFMS separated in M_{\star} . In both group and non-group environments the sub-classes have high $\langle \text{SFR} \rangle$ and longer $\langle t_{\text{dep}} \rangle$, this is the indication that the galaxies are

less active in their respective environments and they take longer period to deplete their HI reservoir, except for one sub-class SB(high) of the non-group sub-sample. The SB(high) sub-classes of the non-group sub-sample have low $\langle \text{SFR} \rangle$ and shorter $\langle t_{\text{dep}} \rangle$ compared to the other sub-classes. This means the galaxies in SB(high) are highly active star-forming galaxies and they will probably deplete their HI reservoirs in a very short period unless they get gas accretion to sustain their HI reservoirs and to keep their high $\langle \text{SFR} \rangle$ (Chowdhury et al., 2022).

Sub-samples	$\langle \text{SFR} \rangle$ [$M_{\odot} \text{yr}^{-1}$]	$\langle t_{\text{dep}}^{\text{True}} \rangle$ [Gyr]	$\langle t_{\text{dep}}^{\text{Stack}} \rangle$ [Gyr]
Non-group			
SB(low)	0.34	1.98	6.03
SB(high)	0.27	2.26	7.20
SFMS(low)	0.32	1.87	6.99
SFMS(high)	0.33	2.21	7.21
Group			
SB(low)	0.35	1.96	6.63
SB(high)	0.39	1.85	5.25
SFMS(low)	0.35	1.97	7.12
SFMS(high)	0.34	1.83	5.27

Table 5.3: Present additional resulting parameters for the non-group and group galaxies sub-classes separated in stellar mass, at $\log_{10}(M_{\star}/M_{\odot}) = 10$.

5.2 Confusion rate in Group and Non-Group galaxies

In addition to the findings presented in chapter 4, we further divided our sample into group and non-group sub-samples and performed stacking to investigate the impact of optical catalog incompleteness on high-redshift stacking experiments. The stacking results for the sub-samples are summarized in table 5.1. Our analysis revealed that group galaxies exhibit a significantly higher confusion rate compared to non-group galaxies. Specifically, the confusion rate for group galaxies was found to be very high, at = 3.52. This high confusion rate can be attributed to the presence of numerous nearby galaxies that contaminate the flux of group galaxies, leading to an increased level of confusion in their stacked spectra.

Conversely, the non-group sub-sample, consisting of individuals who do not belong to any group, demonstrated a low confusion rate of 1.8. This lower confusion rate

can be attributed to the fact that non-group galaxies are situated in less densely clustered spatial environments, resulting in fewer neighboring galaxies that introduce additional flux to their extracted sub-volumes. Consequently, the influence of nearby galaxies on the sub-volumes of non-group galaxies is diminished, leading to a reduced level of confusion in clustered environments.

These findings indicate that our catalog, as well as other high-redshift catalogs, is biased towards group galaxies. Therefore, caution should be exercised when stacking galaxies in dense group environments in future studies examining the HI content of high-redshift catalogs using stacking methods. The results underscore the importance of implementing proper confusion corrections to account for contamination effects.



6 Correcting for confusion in the stacked spectrum

As emphasized throughout the previous chapters, the stacking technique is a cost-effective and powerful tool employed by researchers to study the average HI mass of undetected galaxy samples (for example: Zwaan, 2000; Lah et al., 2009; Geréb et al., 2015). While the HI stacking technique offers numerous advantages, accurately deriving galaxy properties from stacked spectra proves challenging in practice due to confusion. When extracting the sub-volume of a target galaxy, it is highly unlikely that the sub-volume will exclusively contain the flux from that particular galaxy (Elson et al., 2016). Contamination from nearby galaxies is inevitable, and thus, the confusion must be corrected for in the stacked spectrum before conducting any quantitative analysis.

In section 6.1 of our study, we employ the confusion correction method introduced by Chowdhury et al. (2022) to address the issue of confusion rate in our stacked spectra. Our aim is to demonstrate the limitations of traditional methods of confusion correction and highlight the importance of accurate modeling of confusion rate. We show that the traditional methods are inadequate in effectively addressing the issue of confusion in stacked spectra.

Subsequently, in section 6.2, we computed the average HI depletion timescale using a similar approach to Chowdhury et al. (2022). Our objective was to compare their results with our own findings and ascertain whether their evaluated values were overestimated. By performing the computations, we aimed to contribute to the understanding of the HI depletion timescale and provide insights into potential discrepancies in previous studies.

6.1 Correcting for confusion by fitting and subtracting the second-order polynomial on the stacks

In their study, Chowdhury et al. (2022) utilized the upgraded Giant Metrewave Radio Telescope (GMRT) (Swarup et al., 2000; Gupta et al., 2017; Kale, Kale) to investigate the relationship between the average HI mass of star-forming galaxies at redshifts $z = 0.74 - 1.45$ and their average stellar mass, $\langle M_{\star} \rangle$. They employed stacking techniques on HI emissions from their galaxy sample. Prior to the stacking

experiments, they flagged out red galaxies, AGNs, and galaxies with $M_*/M_\odot < 10^9$, resulting in 11,419 main-sequence star-forming galaxies remaining. The remaining sample was further divided into sub-samples based on galaxy stellar mass, with $M_*/M_\odot < 10^{10}$ and $M_*/M_\odot > 10^{10}$. They performed stacking on these stellar mass sub-samples and attempted to correct for the confusion rate on the stacks.

To correct for the confusion rate, [Chowdhury et al. \(2022\)](#) fitted a second-order polynomial to the stacked spectra or stacked HI sub-cubes and subtracted it from the stack. The second-order polynomial was only fitted to the wings of the stacked spectra, excluding the central velocity range ± 250 km/s to avoid fitting the polynomial to the signal. [Figure 6.1](#) illustrates the stacking results after fitting and subtracting the second-order polynomial. Their results demonstrate clear detections of HI 21 cm emission and rest-frame 1.4 GHz continuum emission, indicating the presence of neutral hydrogen gas and ongoing star formation processes within the studied galaxies. After correcting for the confusion rate, they computed the average HI masses and average HI depletion timescales. The results of their calculations are presented in [figure 6.2](#). They found that galaxies with $M_*/M_\odot > 10^{10}$ have HI reservoirs that can sustain their average star formation rates (SFRs) for only 0.86 Gyr, indicating the need for gas accretion to maintain their HI reservoirs and sustain their high SFRs.

In our study, we conducted a follow-up on the work of [Chowdhury et al. \(2022\)](#) using simulated LADUMA HI data cubes within a different redshift range, specifically $0.7 < z < 0.758$, which is distinct from the redshift range used in [Chowdhury et al. \(2022\)](#) (i.e., $z = 0.74 - 1.45$). We pursued this investigation because we had concerns about the method they employed to correct for the confusion rate in their stacks. As mentioned previously, it is crucial for the confusion correction method to effectively reduce the confusion rate on the stacks to avoid overestimation of the average HI gas properties.

Our hypothesis was that "fitting the second-order polynomial on the stacked spectrum and/or stacked sub-cubes and excluding the central velocity range ± 250 km/s on the stacks will not sufficiently reduce the confusion rate on the stacks, resulting in a confusion rate greater than > 1.5 ." To emulate their experiment, we flagged out 18,201 galaxies with $M_*/M_\odot < 10^9$, leaving us with 8,354 blue and red galaxies with $M_*/M_\odot \geq 10^9$, as shown in [figure 6.3](#). Additionally, we flagged out 3,684 red galaxies, as illustrated in [figure 6.3](#), and were left with 4,670 blue star-forming

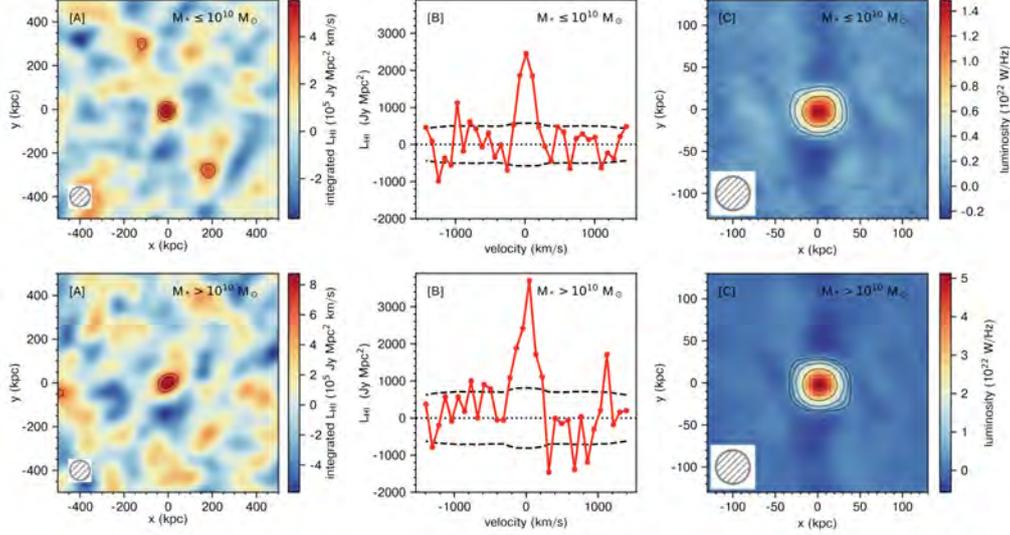


Figure 6.1: The stacked HI 21 cm emission images, HI 21 cm emission spectra, and stacked rest-frame 1.4 GHz luminosity images of star-forming galaxies at $z \sim 1$. These results of Chowdhury et al. (2022) demonstrate clear detections of HI 21 cm emission and rest-frame 1.4 GHz continuum emission, indicating the presence of neutral hydrogen gas and ongoing star formation processes within the studied galaxies.

galaxies with $M_*/M_\odot \geq 10^9$.

Next, we divided our sample into stellar mass sub-samples, specifically $M_*/M_\odot < 10^{10}$ and $M_*/M_\odot > 10^{10}$. Using the stacking method described in chapter 3, we stacked our sub-samples. The stacking results for the sub-sample with $M_*/M_\odot < 10^{10}$ and $M_*/M_\odot > 10^{10}$ are depicted in figures 6.4 and 6.5, respectively. The resulting stacks from the noise-free and noise-filled cubes are represented by the blue and black spectra, respectively. Panel (a) in both figures provides information on the sub-sample's M_* range, the number of galaxies in the sub-sample, and the confusion rate from the stacks. We fitted the second-order polynomial to our stacked spectrum, indicated by the red curve in panel (a) of both figures. Similar to Chowdhury et al. (2022), we excluded the velocity range ± 250 km/s during the fitting process. We then subtracted the second-order polynomial from the stacked spectrum, and the resulting spectrum is displayed in panel (b) of both figures. The confusion rate calculated after the correction is reported in panel (b) of both figures. Table 6.1 presents all the average quantities we computed for the sub-samples.

After following the correction method proposed by Chowdhury et al. (2022), we were able to reduce the confusion rate from 2.921 to 1.614 for the sub-sample with $M_*/M_\odot < 10^{10}$, and from 4.333 to 2.356 for the sub-sample with $M_*/M_\odot > 10^{10}$.

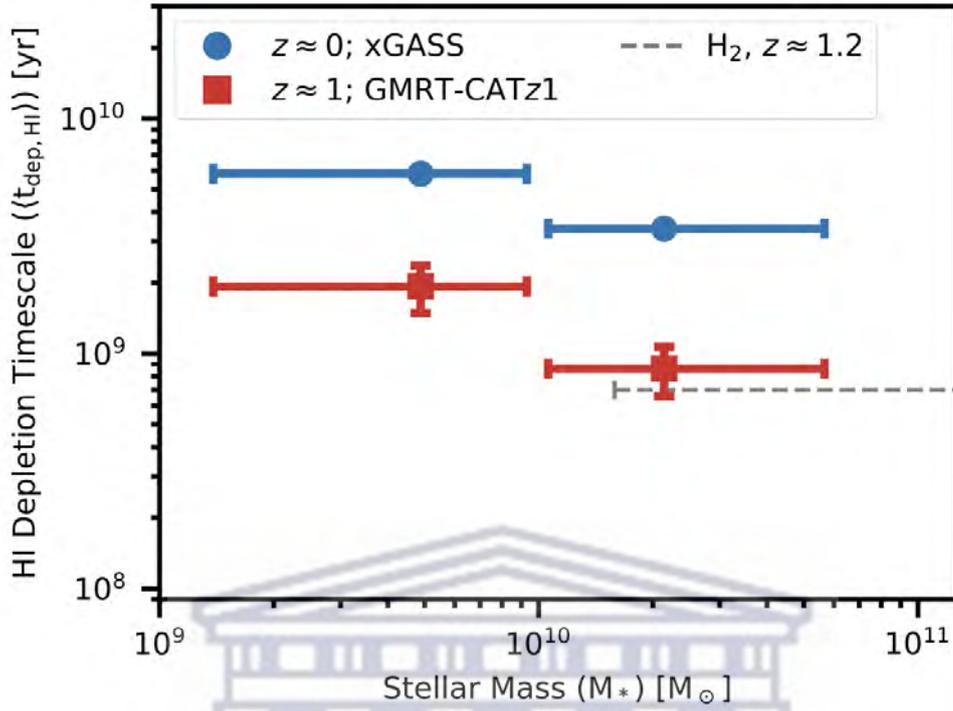


Figure 6.2: The average HI depletion timescale, $\langle t_{\text{dep, HI}} \rangle = \langle M_{\text{HI}} \rangle / \langle \text{SFR} \rangle$, of star-forming galaxies at $z \sim 1.1$, plotted against the average stellar mass from [Chowdhury et al. \(2022\)](#). The red circles show our measurements of the average HI depletion timescale for the two stellar mass sub-samples. The blue points show the average HI depletion timescale, $\langle t_{\text{dep, HI}} \rangle = \langle M_{\text{HI}} \rangle / \langle \text{SFR} \rangle$, for two sub-samples of xGASS galaxies at $z \sim 0$ ([Catinella et al., 2018](#)), each with the same stellar-mass distribution as the corresponding DEEP2 sub-sample. The dashed gray line shows the H_2 depletion timescale of star-forming galaxies at $z \sim 1.2$ ([Tacconi et al., 2013](#)).

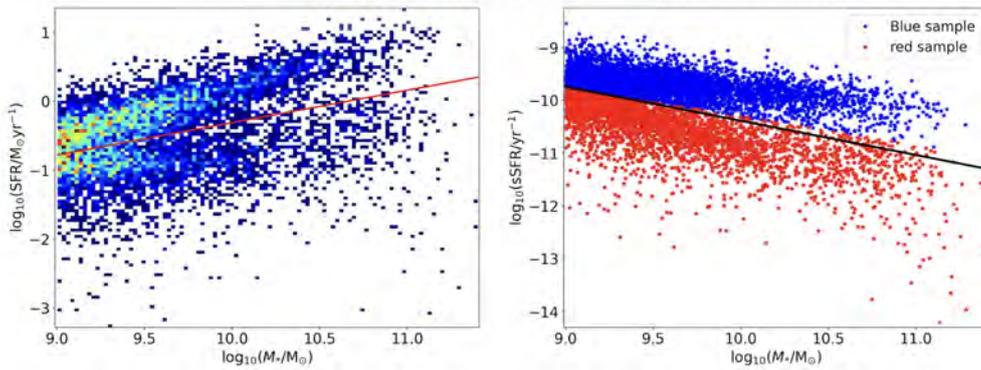


Figure 6.3: The remaining sample after flagging out all the galaxies with $M_*/M_\odot < 10^9$ on our sample as [Chowdhury et al. \(2022\)](#) on theirs. The red galaxies are represented with the red-shaded star, while blue galaxies are represented with blue-shaded stars.

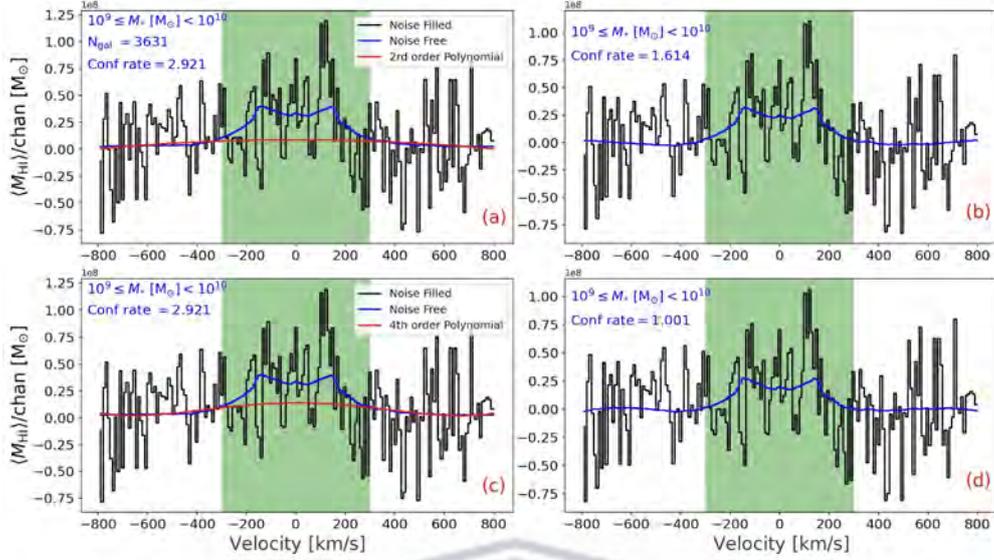


Figure 6.4: The stacking results for the first sub-sample, with $10^9 \leq M_*/M_\odot < 10^{10}$. In all the panels the blue and black curves represent the stacked spectrum from sub-volumes extracted from the noise-free and noise-filled versions of our cubes. The red-curve in panel (a) represents the second order polynomial that was fitted to and subtracted from the stacked spectrum. Panel (a) also reports the sub-sample stacked, the number of galaxies stacked in this sub-sample and the confusion rate after stacking the galaxies. Panel (b) shows the stacked spectrum after the second order polynomial is subtracted from the stacks. It also reports the confusion rate after the second order polynomial was subtracted from the stacks. In panel (c) the red curve represents the fourth order polynomial that was fitted to and subtracted from the stacked spectrum. Panel (d) shows the stacked spectrum after the fourth order polynomial is subtracted from the stacks. It also reports the confusion rate after the fourth order polynomial was subtracted from the stacks. The additional computed parameters from these results are presented in table 6.1.

Although we managed to reduce the confusion rate by subtracting the second-order polynomial from the stacks, we observed that a significant amount of confusion still remained in the stacks. To further demonstrate the inadequacy of using a polynomial to correct for the confusion rate, we decided to fit a higher-order polynomial, specifically a fourth-order polynomial, to the stacked spectrum.

Similar to the second-order polynomial fitting, we excluded the velocity range ± 250 km/s when fitting the fourth-order polynomial to the stacks. In figures 6.4 and 6.5, panel (c), the red curve represents the fitted fourth-order polynomial. We then subtracted the polynomial from the stacks, and the resulting stacks are shown in panel (d) of both figures. The confusion rate after this correction is reported in panel

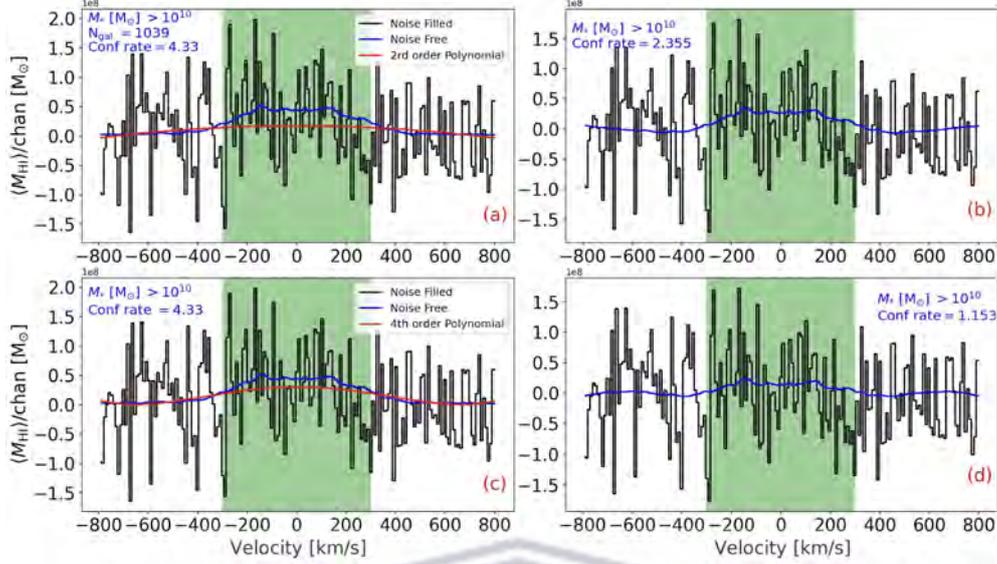


Figure 6.5: The stacking results for the first sub-sample, with $M_*/M_\odot > 10^{10}$. In all the panels the blue and black curves represents the stacked spectrum from sub-volumes extracted from the noise-free and noise-filled versions of our cubes. The red-curve in panel (a) represent the second order polynomial that was fitted to and subtracted from the stacked spectrum. Panel (a) also reports the sub-sample stacked, the number of galaxies stacked in this sub-sample and the confusion rate after stacking the galaxies. Panel (b) shows the stacked spectrum after the second order polynomial is subtracted from the stacks. It also reports the confusion rate after the second order polynomial was subtracted from the stacks. In panel (c) the red curve represents the fourth order polynomial that was fitted to and subtracted from the stacked spectrum. Panel (d) shows the stacked spectrum after the fourth order polynomial is subtracted from the stacks. It also reports the confusion rate after the fourth order polynomial was subtracted from the stacks. The additional computed parameters from these results are presented in table 6.1.

(d) as well. Table 6.1 presents the evaluated average HI mass after subtracting the polynomial, along with the confusion rate after the correction. It is important to note that the fourth-order polynomial correction method seems "well" in terms of reducing the confusion rate. For the sub-sample with $M_*/M_\odot < 10^{10}$, the confusion rate was reduced from 2.921 to 1.001, and for the sub-sample with $M_*/M_\odot > 10^{10}$, the confusion rate was reduced from 4.33 to 1.153. However, it is worth noting that this method comes with a drawback—the signal-to-noise ratio is now significantly low, and the evaluated $\langle M_{\text{HI}} \rangle$ from the stacks is dominated by noise, this means we lost signal.

The results obtained in this study, even after applying the correction method intro-

	Low M_* sub-sample	High M_* sub-sample
N_{gal}	6172	1362
Average stellar mass, $\langle M_* \rangle$, ($10^9 M_\odot$)	3.13	27.89
Average true HI mass, $\langle M_{\text{HI}}^{\text{True}} \rangle$, ($10^9 M_\odot$)	0.79	0.63
Average evaluated HI mass, $\langle M_{\text{HI}}^{\text{Stack}} \rangle$, ($10^9 M_\odot$)	2.31	2.71
Confusion rate, $\langle M_{\text{HI}}^{\text{Stack}} \rangle / \langle M_{\text{HI}}^{\text{True}} \rangle$	2.92	4.33
Average evaluated HI mass after subtracting the 2nd-order polynomial, $\langle M_{\text{HI}}^{\text{Stack-2nd}} \rangle$, ($10^9 M_\odot$)	1.28	1.47
Confusion rate, $\langle M_{\text{HI}}^{\text{Stack-2nd}} \rangle / \langle M_{\text{HI}}^{\text{True}} \rangle$	1.61	2.36
Average evaluated HI mass after subtracting the 4th-order polynomial, $\langle M_{\text{HI}}^{\text{Stack-4th}} \rangle$, ($10^9 M_\odot$)	0.79	0.72
Confusion rate, $\langle M_{\text{HI}}^{\text{Stack-4th}} \rangle / \langle M_{\text{HI}}^{\text{True}} \rangle$	1.00	1.15

Table 6.1: The resulting parameters for both sub-samples before and after fitting to and subtracting from the stacks the second and fourth order polynomials.

duced by [Chowdhury et al. \(2022\)](#), indicate that the method still leaves the stacks with a high confusion rate. A high confusion rate in the stacked spectrum implies that any quantitative analysis conducted using these stacks will likely lead to overestimation. Based on our findings, we recommend that future studies explore alternative correction methods, such as those proposed by [Fabello \(2012\)](#); [Jones \(2016\)](#); [Elson et al. \(2016\)](#), rather than relying solely on fitting and subtracting a second-order polynomial from the stacks.

For instance, [Fabello \(2012\)](#) considered any galaxy located within a half beam width and within the velocity range ± 300 km/s of the target galaxy sub-volume to contribute additional flux that must be removed or subtracted. [Elson et al. \(2016\)](#) demonstrated a high level of accuracy and reliability in determining the amount of contaminating emission from nearby galaxies to the host galaxy sub-volume. They decomposed the sub-volume into emission contributions from nearby galaxies and the host galaxy, allowing for the subtraction of the emission contributed by neigh-

boring galaxies from the sub-volume.

By exploring alternative correction methods, researchers can improve the accuracy and reliability of confusion correction in stacking experiments, ultimately leading to more robust and trustworthy results.

6.2 HI depletion timescale

In this section, we calculate the average HI depletion timescale $\langle t_{\text{dep}} \rangle = \langle M_{\text{HI}} \rangle / \langle \text{SFR} \rangle$ for different average HI masses presented in Table 6.1. The results of the calculated $\langle t_{\text{dep}} \rangle$ are shown in Figure 6.6, where the error bars on the y-axis represent the statistical uncertainties of the measurements. These uncertainties were computed using the Jackknife re-sampling method, following the approach of Sahinler & Topuz (2007); Yu (2002); Shao & Tu (2012).

To compute the uncertainties, we randomly sampled 75 percent of the data from each sub-sample and re-sampled it using the Jackknife re-sampling method. From the re-samples, we calculated the mean standard error for all data points shown in Figure 6.6.

Table 6.2 presents the $\langle \text{SFR} \rangle$ for each sub-sample and the resulting average HI depletion timescale before and after subtracting the second and fourth-order polynomials from the stacks. The blue shaded triangles represent the true results of the average HI depletion timescale for the sub-samples.

From the figure, it is evident that the results obtained from the average evaluated HI mass from the stacks (represented by black crosses) significantly overestimate the true results of the average HI depletion timescale, mainly due to confusion. Even after correcting for confusion by fitting and subtracting the second-order polynomial, the average HI depletion timescale is still overestimated by a factor > 1.5 for the sub-samples (represented by red shaded circles). This underscores the importance of adequate confusion correction methods to avoid overestimation in quantitative analyses.

The fourth-order polynomial confusion correction method appears to work "well" but it should be noted that the subtraction of the fourth-order polynomial results in a loss of signal. The average HI depletion timescale after subtracting the fourth-order

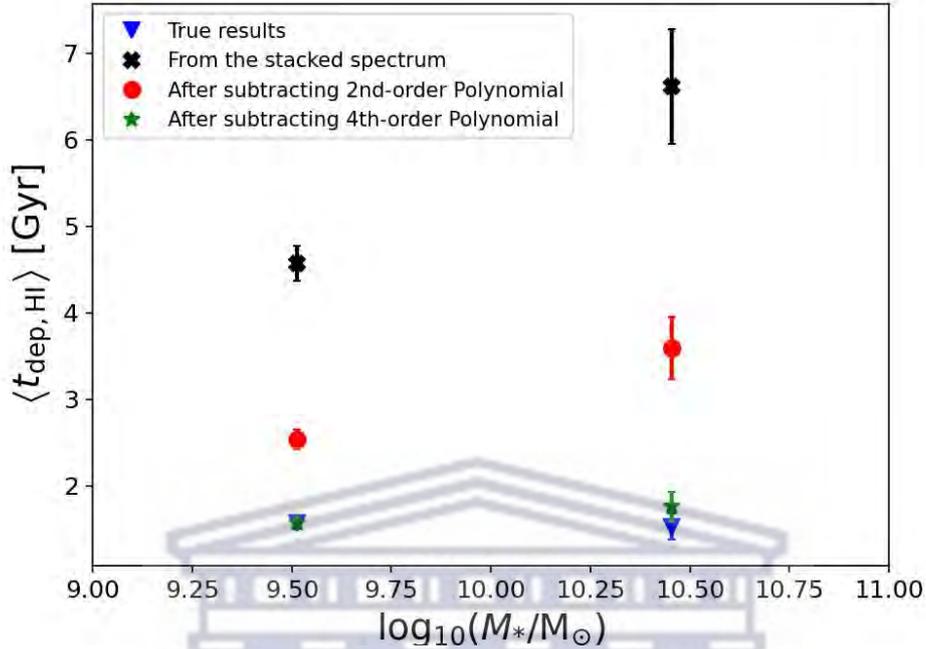


Figure 6.6: The average HI depletion timescale $\langle t_{\text{dep}} \rangle = \langle M_{\text{HI}} \rangle / \langle \text{SFR} \rangle$ plotted against stellar mass of galaxies for both stellar mass sub-samples, with $M_*/M_\odot < 10^{10}$, and with $M_*/M_\odot > 10^{10}$ at redshift range $z = 0.7 - 0.758$. The black crosses represent the average HI depletion timescale computed from the resulting stacks before correcting for contamination rate. The red-shaded circles represent the average HI depletion timescale computed from the resulting stacks after fitting to and subtracting from the stacks the second order polynomial. The green-shaded stars represent the average HI depletion timescale computed from the resulting stacks after fitting to and subtracting from the stacks the fourth order polynomial. The error bars on y-axis indicate the statistical uncertainties on the measurements, there are computed using jackknife re-sampling statistic (for example: Sahinler & Topuz, 2007; Yu, 2002; Shao & Tu, 2012) with 75 percent of the sample randomly selected and re-sampled. We found a very short average HI depletion timescale, = 1.761 Gyr, for star-forming galaxies with $M_*/M_\odot > 10^{10}$ at $z = 0.7 - 0.758$.

polynomial is represented by green shaded stars in the figure.

Our findings suggest that the results obtained by Chowdhury et al. (2022) after fitting and subtracting the second-order polynomial from the stacks are overestimated, and they are even worse than the results we obtained using the same method. This difference in results can be attributed to the fact that Chowdhury et al. (2022) worked with galaxies at higher redshifts ($z = 0.74 - 1.45$), which have a higher confusion rate compared to our redshift bin ($z = 0.7 - 0.758$). After correcting for con-

	Low M_* sub-sample	High M_* sub-sample
Average star-formation rate $\langle \text{SFR} \rangle$ ($M_\odot \text{yr}^{-1}$)	0.50	2.45
Average true HI depletion timescale $\langle t_{\text{dep}}^{\text{True}} \rangle$ (Gyr)	1.57	1.53
Average evaluated HI depletion timescale $\langle t_{\text{dep}}^{\text{Stack}} \rangle$, (Gyr)	4.58	6.62
Average evaluated HI depletion timescale after subtracting the 2nd-order polynomial, $\langle t_{\text{dep}}^{\text{Stack-2nd}} \rangle$, (Gyr)	2.53	3.60
Average evaluated HI depletion timescale after subtracting the 4th-order polynomial, $\langle t_{\text{dep}}^{\text{Stack-4th}} \rangle$, (Gyr)	1.57	1.76

Table 6.2: Present the average SFR for both sub-samples and the average HI depletion timescale for before and after correcting for the confusion rate.

fusion using a second-order polynomial, we found that galaxies with $M_*/M_\odot > 10^{10}$ have HI reservoirs that can sustain their SFRs for a much shorter period, lower by a factor of approximately ≥ 2.356 compared to the average HI depletion timescale found in Chowdhury et al. (2022). This indicates that the average HI depletion timescale reported in Chowdhury et al. (2022) is overestimated by a factor > 2.356 , as the confusion rate increases with higher redshifts and the cubes used in their study cover a redshift range higher than ours. Consequently, a larger amount of gas accretion is required to sustain their HI reservoirs and maintain their high SFRs at those redshifts.

7 HI scaling relations

Examining the relation of the HI content with other physical properties of galaxies is important in understanding the physical processes regulating galaxy evolution (for example: [Brown et al., 2015](#); [Healy et al., 2019](#); [Catinella et al., 2010, 2018](#)). Solid scaling relations at redshift $z \gtrsim 0$ between the HI content of star-forming galaxies and their stellar mass (M_* , [Huang et al., 2012](#); [Maddox et al., 2021](#)), star-formation rate (SFR, [Feldmann, 2020](#); [Sinigaglia et al., 2022](#)) and specific star-formation rate (sSFR, [Sinigaglia et al., 2022](#); [Rhee et al., 2023](#)), between others, have been discovered by large-scale HI galaxy surveys. Figure 7.1 shows the $\log_{10}(M_{\text{HI}})$ - $\log_{10}(M_*)$ relation for MIGHTEE-HI Early Science galaxies HI detections, spanning redshift $0 < z < 0.084$. Black points mark individual galaxy. The red curve represent the median and $1\text{-}\sigma$ uncertainty of the relation from the MIGHTEE-HI data, while the grey curve is the same relation derived from the ALFALFA survey ([Maddox et al., 2021](#)).

In Figure 7.1, the MIGHTEE-HI and ALFALFA relations show good agreement at intermediate masses. However, at higher stellar masses ($\log_{10}(M_*/M_\odot) \gtrsim 9.5$), a decrease in the MIGHTEE-HI data is observed. [Maddox et al. \(2021\)](#) speculate that this decrease might be due to the limited volume probed in Early Science. Another notable feature is the apparent break at $\log_{10}(M_*/M_\odot) \sim 9$, which was also identified by [Maddox et al. \(2015\)](#). Interestingly, this break has not been observed in any of the available simulations. The scaling relations presented in [Sinigaglia et al. \(2022\)](#) and [Rhee et al. \(2023\)](#) are depicted in panels of Figure 7.2. A detailed discussion on these relations will be provided when presenting the results obtained in this part of the work later in the chapter.

The primary focus of this study is to investigate the scaling relations in LADUMA high-redshift galaxies. The research begins by exploring how the confusion rate from stacked spectra can negatively impact the scaling relations of galaxies if left uncorrected. Subsequently, a comparison is made between the scaling relations obtained in this study and those derived by [Sinigaglia et al. \(2022\)](#) and [Rhee et al. \(2023\)](#). Both [Sinigaglia et al. \(2022\)](#) and [Rhee et al. \(2023\)](#) employ the spectral stacking procedure to investigate the HI scaling relations. [Sinigaglia et al. \(2022\)](#) use the MIGHTEE-HI Early Science data cubes ([Jonas & Team, 2016](#)), obtained with the MeerKAT radio telescope, covering a redshift range of $0.23 < z < 0.49$.

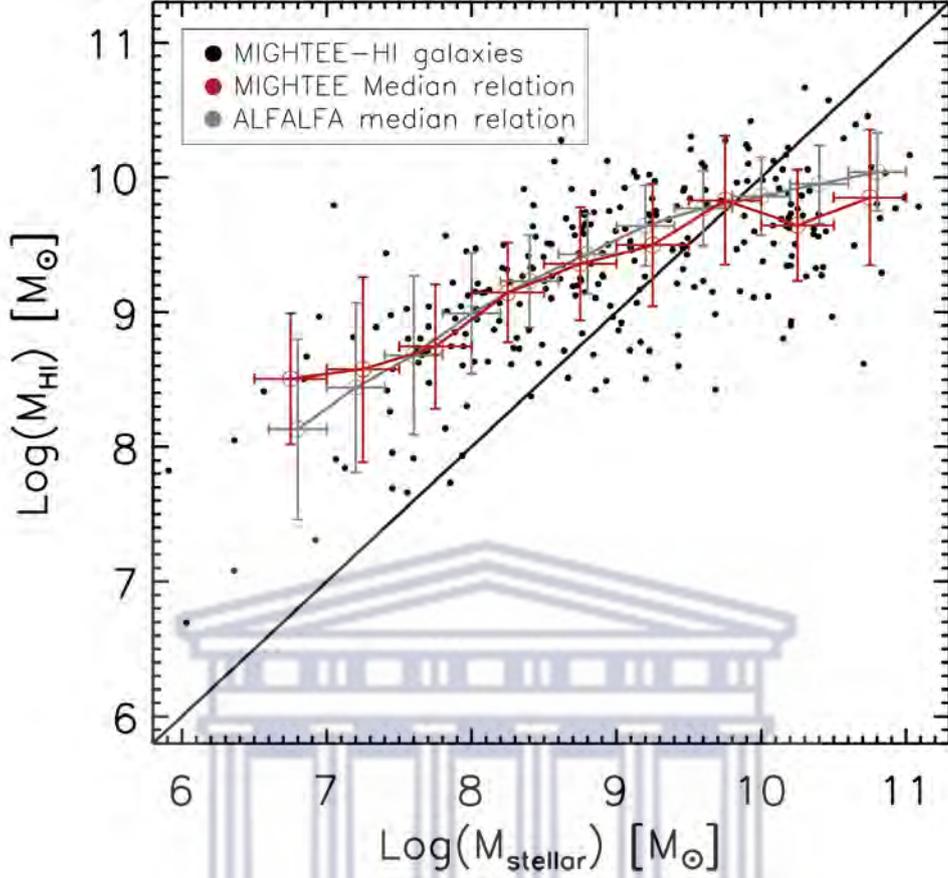


Figure 7.1: the $\log_{10}(M_{\text{HI}})$ - $\log_{10}(M_{\star})$ relation for MIGHTEE-HI Early Science galaxies HI detections, spanning redshift $0 < z < 0.084$. Black points mark individual galaxy. The red curve represent the median and $1\text{-}\sigma$ uncertainty of the relation from the MIGHTEE-HI data, while the grey curve is the same relation derived from the ALFALFA survey. Image credit to Maddox et al. (2021).

On the other hand, Rhee et al. (2023) utilize the Deep Investigation of Neutral Gas Origins (DINGO) data from the Australian Square Kilometer Array Pathfinder (ASKAP), covering a redshift range of $0.04 < z < 0.09$. It is important to note that the redshift ranges of the DINGO early science data and the MIGHTEE-HI early science data cubes differ from those of the LADUMA simulations employed in this study.

In this study, we focus on investigating the HI scaling relations for star-forming galaxies within our sample. To identify star-forming galaxies, we utilize the bimodality between blue late-type and red early-type galaxy morphology, as demonstrated by Jian et al. (2018). Specifically, we classify the blue galaxies in our sample as star-forming galaxies. We examine the scaling relations between different sub-samples

and produce the scaling relations by stacking the HI line spectra for stellar mass bins (Figure 4.4), star-formation rate bins (Figure 4.8), and specific star-formation rate bins (Figure 4.12) presented in Chapter 4.

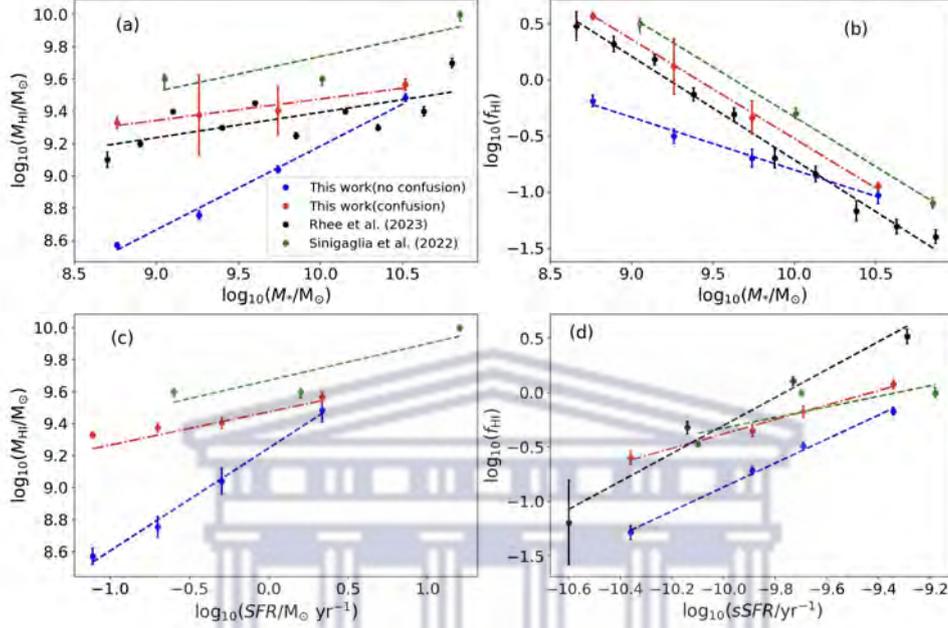


Figure 7.2: Star forming galaxy HI scaling relations: $\log_{10}(M_{\text{HI}})$ - $\log_{10}(M_{\star})$ panel (a), $\log_{10}(f_{\text{HI}})$ - $\log_{10}(M_{\star})$ panel (b), $\log_{10}(M_{\text{HI}})$ - $\log_{10}(\text{SFR})$ panel (c) and $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ panel (d). Blue and red circles represents our scaling relations produced from the simulated data (using true galaxy properties) and produced from the stacking results, respectively. Black circles represents scaling relations at $z \sim 0.7$ produced by Rhee et al. (2023) at redshift range $0.04 < z < 0.09$. The green circles represents scaling relations produced by Sinigaglia et al. (2022) at redshift range $0.23 < z < 0.49$. The dashed lines are the fitted lines to the data points and the colour of each line corresponds to the colour on the data points that it is fitted on. The error bars on the blue and red data points we computed using Jackknife re-sampling methods.

Figure 7.2 presents the scaling relations derived in this work. Panel (a) displays the $\log_{10}(M_{\text{HI}})$ - $\log_{10}(M_{\star})$ relation, panel (b) shows the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(M_{\star})$ relation, panel (c) presents the $\log_{10}(M_{\text{HI}})$ - $\log_{10}(\text{SFR})$ relation, and panel (d) exhibits the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ relation. The blue circles represent the results obtained from the LADUMA input catalogue that is made up of the true galaxy properties, representing the true galaxy properties, while the red circles indicate the results obtained from the LADUMA spectral stacked spectra. The black circles correspond to the scaling relations from Rhee et al. (2023), and the green circles represent the scaling relations from Sinigaglia et al. (2022).

To fit a linear model in $\log - \log$ space ($\log_{10}(A) = \alpha \log_{10}(B) + \beta$) to the blue, red, black, and green data points, we utilize the least-squares-fitting method. Furthermore, we compute the uncertainties for the resulting parameters (α and β) using Markov chain Monte Carlo (MCMC) methods (for example: [Geman & Geman, 1984](#); [Chen et al., 2011](#)). The dashed lines in Figure 7.2 represent the fitted lines to the data points, with each line’s color corresponding to the color of the data points it was fitted on. The resulting parameters and their uncertainties are presented in Table 7.1.

	LADUMA with no confusion	LADUMA with confusion	Rhee et al. (2023) parameters	Sinigaglia et al.(2022) parameters
$\log_{10}(M_{\text{HI}})-\log_{10}(M_{\star})$				
α	0.53 ± 0.04	0.13 ± 0.01	0.16 ± 0.02	0.08 ± 0.07
β	3.89 ± 0.25	8.14 ± 0.23	7.82 ± 0.12	0.13 ± 0.01
$\log_{10}(f_{\text{HI}})-\log_{10}(M_{\star})$				
α	-0.47 ± 0.04	-0.87 ± 0.05	-0.92 ± 0.03	-0.92 ± 0.07
β	3.90 ± 0.42	8.15 ± 0.43	8.52 ± 0.30	8.88 ± 0.73
$\log_{10}(M_{\text{HI}})-\log_{10}(\text{SFR})$				
α	0.64 ± 0.06	0.16 ± 0.05	-	0.26 ± 0.04
β	9.25 ± 0.11	9.49 ± 0.10	-	9.65 ± 0.44
$\log_{10}(f_{\text{HI}})-\log_{10}(\text{sSFR})$				
α	1.10 ± 0.02	0.66 ± 0.03	1.28 ± 0.02	0.40 ± 0.07
β	10.17 ± 0.20	6.26 ± 0.28	12.55 ± 0.17	3.56 ± 0.69

Table 7.1: Best-fit parameters for the scaling relations computed in this work, in [Rhee et al. \(2023\)](#) and in [Sinigaglia et al. \(2022\)](#).

The $\log_{10}(M_{\text{HI}})-\log_{10}(M_{\star})$ scaling relation shown in panel (a) of Figure 7.2 reveals that the scaling relation derived from the simulated data (representing the true galaxy properties) has a steeper slope compared to the one produced from the stacked spectrum. At low stellar masses, $\log_{10}(M_{\star}/M_{\odot}) \lesssim 9.5$, the simulated data exhibits a lower HI mass compared to the results from the stacked spectrum, differing by approximately 1 dex. However, as we reach higher stellar masses, around $\log_{10}(M_{\star}/M_{\odot}) \sim 10$, the HI mass from the simulations and the stacked spectrum converge, with the stacked spectrum results being around 0.1 dex higher. This suggests that the stacked spectrum tends to produce higher HI mass results than those obtained from the simulated data. Considering the inadequacy of traditional meth-

ods used in the literature to correct for confusion rate in the stacked spectrum, as discussed in Chapter 6, it becomes evident that there is still missing information in our current understanding of galaxy evolution based on stacking experiments. The results obtained from stacking experiments are prone to overestimation due to contamination, especially in the HI mass of low stellar mass galaxies.

It is worth noting that most of the low stellar mass galaxies in our sample are non-group galaxies and exhibit a blue color (as shown in Chapter 5). Therefore, to gain a better understanding of the evolution of low stellar mass galaxies, it would be more reliable to study the HI content of non-group or blue galaxies. These types of galaxies produce stacked spectra with lower confusion rates, enabling more precise estimates of the HI content in low stellar mass galaxies.

In panel (b), the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(M_{\star})$ relation, the results from the stacked spectrum demonstrate a steeper slope compared to the simulation results. At low stellar masses, $\log_{10}(M_{\star}/M_{\odot}) \leq 9.5$, the stacked spectrum shows a higher $\log_{10}(f_{\text{HI}})$ by ~ 0.5 dex. However, as we move to higher stellar masses, around $\log_{10}(M_{\star}/M_{\odot}) \sim 10.5$, the two curves converge.

Panel (c) and panel (d) focus on the star-formation properties. In panel (c), the $\log_{10}(M_{\text{HI}})$ - $\log_{10}(\text{SFR})$ relation exhibits a comparable normalization constant for the stacked spectrum results and the simulation results, but the simulation results display a steeper slope compared to the stacked spectrum results. Panel (d) presents the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ relation, where the stacked spectrum results demonstrate a steeper slope compared to the simulation results. At low specific star-formation rates (sSFR), $\log_{10}(\text{sSFR}) \lesssim -10.20$, the stacked spectrum relation exhibits a higher HI gas fraction compared to the relation produced from the simulations by approximately 0.7 dex. However, as we move to higher sSFR, $\log_{10}(\text{sSFR}) \gtrsim -10$, the divergence between the HI gas fraction for the two gradually decreases.

The deviation observed in the scaling relations derived from the stacked spectrum results compared to the simulated data (representing the true galaxy properties) underscores the importance of accurately correcting for confusion or contamination rates when conducting quantitative analyses using stacked spectrum results. The discrepancy observed in the scaling relations between the stacked spectrum and simulated data in this work is primarily caused by confusion in the stacked spectrum.

Now, let us compare our scaling relations, both from the stacking results and true

galaxy properties, to those presented in Rhee et al. (2023) and Sinigaglia et al. (2022). In all the scaling relations, the slopes obtained from Rhee et al. (2023) and Sinigaglia et al. (2022) are consistent with our scaling relations derived from the stacked spectrum, except for the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ relation. The slopes of the scaling relations produced from the simulations (true galaxy properties) are generally steeper than those from Rhee et al. (2023) and Sinigaglia et al. (2022), again with the exception of the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ scaling relation. For the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ relation, our scaling relation from the simulation exhibits a slope that agrees with Rhee et al. (2023) and is steeper than Sinigaglia et al. (2022).

The interesting aspect of our findings is that, in most of our scaling relations (such as panel (a), (b), and (c)), the scaling relations from the literature align with our scaling relations derived from the stacked spectrum, despite the presence of confusion. The most likely reason for this behavior is that they did not precisely correct for the confusion rate in their stacks (as discussed, for example, in Chowdhury et al. (2022)). Consequently, the results they obtained might be overestimated due to contamination in the stacks, thus leading to agreement between their scaling results and our results that incorporate confusion. In Rhee et al. (2023) and in Sinigaglia et al. (2022) they claim to have estimated that confusion is negligible, that is why their results are in agreement with our contaminated scaling relations. To correct for contamination Sinigaglia et al. (2022) followed Elson et al. (2016), they decomposed the spectrum extracted for target galaxy into contributions from the host galaxy and contributions from nearby contaminating galaxies. Thereafter they subtracted the average contribution from nearby contaminating galaxies from the stacked spectrum.

We agree that Elson et al. (2016) method is a good method and we would like to advise authors to utilise this method. One of the most advantageous things about the method is that one can be able to test or to check if they precisely corrected for contamination. Elson et al. (2016) method involves using Obreschkow & Meyer (2014) flux-limited mock galaxy catalogue, which is based on S³ – SAX to inject galaxies with realistic HI masses and clustering into blank synthetic data cube matching the same angular and spectral size as your observations. Since this method involves injecting galaxies with realistic HI masses, it is easy to test or to check after implementing the contamination correcting method on the stacked spectrum if the average realistic HI mass you injected is reproduced by the stacked spectrum without any over-estimations or underestimations. Therefore we speculate

that [Sinigaglia et al. \(2022\)](#) did not really do this test as it is not mentioned in their work and thus their results are in agreement with ours with contamination. They did not notice that their results are over-estimated.

This highlights the importance of precisely correcting for the confusion rate in stacked spectra when comparing and interpreting scaling relations from different studies. Failure to account for confusion can lead to biased and overestimated results, potentially impacting our understanding of galaxy evolution based on stacking experiments.

Another important consideration is that our work is conducted at different redshift intervals compared to [Rhee et al. \(2023\)](#) and [Sinigaglia et al. \(2022\)](#). Redshift plays a significant role in galaxy evolution, and different redshift ranges encompass distinct physical processes and environments. The observed variations in scaling relations between studies at different redshifts can be attributed to various factors.

Evolutionary effects come into play as galaxy properties, such as gas content and star formation rates, evolve with redshift. By studying galaxies at different redshift intervals, we capture different stages of their evolutionary history, leading to variations in the scaling relations and providing insights into the evolution of galaxy properties. Redshift-dependent selection biases may arise due to the specific redshift intervals chosen for galaxy selection. Different redshift ranges can favor the inclusion of certain galaxy types or environments, leading to variations in the scaling relations. It is crucial to account for these biases when comparing scaling relations across different redshift studies.

Environmental effects also contribute to the observed variations in scaling relations. The influence of the cosmic environment on galaxy properties can vary with redshift. By studying scaling relations at different redshift intervals, we gain insights into how environmental factors impact the observed relationships, including galaxy interactions, mergers, and gas accretion. Redshift-dependent instrumental effects should also be considered. Different observational instruments and techniques used in studies at different redshifts can introduce variations in the measured galaxy properties. Instrumental characteristics, such as sensitivity, resolution, and calibration, can impact the derived scaling relations. Comparing scaling relations across different redshift studies allows us to better understand potential systematic effects and assess the robustness of the observed trends.

By considering the implications of working at different redshift intervals, we can account for the complex interplay between galaxy properties, evolution, and the cosmic environment. This approach enhances our understanding of the physical processes shaping the observed scaling relations and facilitates comparisons across studies conducted at various redshift ranges.



8 Summary and Future outlook

This chapter provides the summary of the results and the conclusions in section 8.1, it follows by providing an outlook for the future work in section 8.2.

8.1 Summary

In this study, our main objective was to investigate the impact of selection effects introduced by input optical catalogues on high-redshift stacking experiments. To accomplish this, we employed the HI stacking technique to analyze the HI content of LADUMA synthetic data cubes generated in the work by Elson et al. (2016). These data cubes consisted of simulated galaxy models within the redshift range of $0.7 < z < 0.758$ and included 26,511 HI galaxy emissions.

To ensure our analysis focused on the appropriate high-redshift portion of the LADUMA cubes, we specifically selected the data cubes that best represented this region. The user-specified inputs utilized to generate the cubes can be found in Table 2.2. Considering that the S³ – SAX catalogue is complete down to $M_*/M_\odot \sim 10^{8.5}$, we conducted our stacking experiments exclusively on galaxies with stellar masses of $M_*/M_\odot \geq 10^{8.5}$.

The main results derived from our investigation can be summarized as follows:

- **Selection Effects:** By stacking galaxies in different sub-samples, we examined the influence of selection effects introduced by the input optical catalogues. Our analysis revealed that the high-redshift catalogues will exhibit biases towards specific galaxy types. These included galaxies with lower stellar masses ($\log_{10}(M_*/M_\odot) < 9.5$), lower star-formation rates ($\log_{10}(\text{SFR}/M_\odot\text{yr}^{-1}) < -0.5$), lower specific star-formation rates ($\log_{10}(\text{sSFR}/\text{yr}^{-1}) < -10$), galaxies found in groups, and red galaxies. It is important for future studies to be cautious when stacking these galaxy types, as the inherent biases in high-redshift catalogues can lead to higher confusion rates in stacked spectra. Adequate correction methods should be employed to avoid potential overestimation in subsequent quantitative analyses.
- **Environmental Impact:** Using stacking methods, we investigated the impact of

the environment on HI gas properties by examining two galaxy sub-samples: group galaxies and non-group galaxies. Non-group galaxies were found to possess higher HI richness and HI gas fractions compared to group galaxies. Furthermore, we observed correlations between HI properties and star-formation/stellar mass properties within the sSFR-stellar mass plane. Active star-forming sub-classes exhibited higher HI masses and HI gas fractions compared to other sub-classes, while lower stellar mass sub-classes displayed higher HI gas fractions and longer depletion timescales. These results suggest that lower mass galaxies are relatively inactive within their respective environments and are at an earlier stage of evolution.

- **Confusion Rate Correction:** Our stacking experiments focused on star-forming galaxies in two different stellar mass sub-samples: $\log_{10}(M_{\star}/M_{\odot}) < 10$ and $\log_{10}(M_{\star}/M_{\odot}) > 10$. To correct for confusion rates in the stacked spectra, we implemented a correction method proposed by [Chowdhury et al. \(2022\)](#), which had previously been used for GMRT star-forming galaxies. However, our investigation revealed that this method resulted in stacked spectra with confusion rates greater than 1.5. Consequently, the corrected HI depletion timescales were overestimated by a factor greater than 1.5. Given these findings, we strongly criticize the use of this method and recommend against its implementation in future studies. It is crucial to employ accurate and robust methods for correcting confusion rates to ensure precise estimation of galaxy properties.
- **Scaling Relations:** For the scaling relations that we measured in this work, we find that our results produced from the stacked spectrum that has confusion are in agreement with the results from DINGO early science data and MITHEE-HI early science data cubes stacking experiments by [Rhee et al. \(2023\)](#) and [Sinigaglia et al. \(2022\)](#), respectively. Except the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ scaling relation. For the $\log_{10}(f_{\text{HI}})$ - $\log_{10}(\text{sSFR})$ relation, we find that our results from the simulations (true galaxy properties) are in agreement with [Rhee et al. \(2023\)](#). We argue that the contrast between our scaling relations results from the simulations (true galaxy properties) and the results of [Rhee et al. \(2023\)](#) and [Sinigaglia et al. \(2022\)](#) is due to the fact that they did not precisely correct for confusion rate on their work. The other reason is that our work is conducted at different redshift intervals compared to [Rhee et al. \(2023\)](#) and [Sinigaglia et al. \(2022\)](#).

8.2 Future outlook

In this section, we provide an outlook on potential future directions and areas for further research. Here are some suggestions:

- **Refining Selection Effects:** Further investigations should focus on refining the understanding and correction of selection effects introduced by input optical catalogues. Developing more sophisticated methods to account for biases and improve the accuracy of stacking experiments will enhance the reliability of the results.
- **Improved Confusion Rate Correction:** It is crucial to develop more accurate and robust methods for correcting confusion rate in stacked spectra. This will ensure more precise estimates of galaxy properties and avoid overestimation in quantitative analyses.
- **Redshift Dependency:** Exploring the redshift dependency of scaling relations and studying the evolution of these relations over different redshift ranges will provide valuable insights into galaxy evolution and the impact of cosmic time.
- **Integration of Multi-wavelength Data:** Integrating data from multiple wavelengths, such as optical, infrared, and radio, will provide a more comprehensive understanding of galaxy properties. Incorporating multi-wavelength information into stacking analyses will enhance our knowledge of galaxy evolution and scaling relations.
- **Larger and Deeper Surveys:** Expanding the sample size and depth of surveys will improve statistical significance and enable more detailed investigations. Larger and deeper surveys will provide a better understanding of scaling relations and their dependence on various galaxy properties.

By addressing these aspects in future research, we can advance our understanding of selection effects, improve confusion rate correction, and gain deeper insights into galaxy properties and their evolution.

In conclusion, this study highlights the importance of considering selection effects and accurately correcting for confusion rate in high-redshift stacking experiments. It

emphasizes the need for careful interpretation of stacking results and provides suggestions for future research to refine methodologies and enhance our understanding of galaxy properties.



Bibliography

- Ball L., 2009, ATNF Memo
- Barnes D. G., et al., 2001, [Monthly Notices of the Royal Astronomical Society](#), 322, 486
- Bekki K., Couch W. J., Drinkwater M. J., Gregg M. D., 2001, [The Astrophysical Journal](#), 557, L39
- Bertone S., De Lucia G., Thomas P. A., 2007, [Monthly Notices of the Royal Astronomical Society](#), 379, 1143
- Bigiel F., Blitz L., 2012, [The Astrophysical Journal](#), 756, 183
- Bigiel F., Leroy A., Walter F., Brinks E., De Blok W., Madore B., Thornley M. D., 2008, [The Astronomical Journal](#), 136, 2846
- Bird S., Garnett R., Ho S., 2017, [Monthly Notices of the Royal Astronomical Society](#), 466, 2111
- Blyth S., 2014, in American Astronomical Society Meeting Abstracts# 223. pp 404–03
- Blyth S., et al., 2016, MeerKAT Science: On the Pathway to the SKA; Sissa Medialab: Trieste, Italy, p. 4
- Bower R. G., Benson A., Malbon R., Helly J., Frenk C., Baugh C., Cole S., Lacey C. G., 2006, [Monthly Notices of the Royal Astronomical Society](#), 370, 645
- Broeils A. H., Rhee M.-H., 1997, [Astronomy and Astrophysics](#), 324, 877
- Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, [Monthly Notices of the Royal Astronomical Society](#), 452, 2479
- Brown T., et al., 2017, [Monthly Notices of the Royal Astronomical Society](#), 466, 1275
- Catinella B., Giovanelli R., Haynes M. P., 2006, [The Astrophysical Journal](#), 640, 751

- Catinella B., et al., 2010, [Monthly Notices of the Royal Astronomical Society](#), 403, 683
- Catinella B., et al., 2018, [Monthly Notices of the Royal Astronomical Society](#), 476, 875
- Chen S., Schwarz D. J., 2016, [Astronomy & Astrophysics](#), 591, A135
- Chen S., Dick J., Owen A. B., 2011, | 10.1214/10-AOS831
- Chengalur J. N., Braun R., Wieringa M., 2001, [arXiv preprint astro-ph/0104331](#)
- Chilingarian I., Cayatte V., Revaz Y., Dodonov S., Durand D., Durret F., Micol A., Slezak E., 2009, [Science](#), 326, 1379
- Chowdhury A., Kanekar N., Chengalur J. N., 2020, [The Astrophysical Journal Letters](#), 900, L30
- Chowdhury A., Kanekar N., Chengalur J. N., 2022, [The Astrophysical Journal Letters](#), 931, L34
- Crain R. A., et al., 2016, [Monthly Notices of the Royal Astronomical Society](#), 464, 4204
- Davé R., 2008, [Monthly Notices of the Royal Astronomical Society](#), 385, 147
- Davé R., Finlator K., Oppenheimer B. D., 2011, [Monthly Notices of the Royal Astronomical Society](#), 416, 1354
- De Lucia G., Blaizot J., 2007, [Monthly Notices of the Royal Astronomical Society](#), 375, 2
- Delhaize J., Meyer M., Staveley-Smith L., Boyle B., 2013, [Monthly Notices of the Royal Astronomical Society](#), 433, 1398
- Di Cintio A., Tremmel M., Governato F., Pontzen A., Zavala J., Bastidas Fry A., Brooks A., Vogelsberger M., 2017, [Monthly Notices of the Royal Astronomical Society](#), 469, 2845
- Duffy A. R., Meyer M. J., Staveley-Smith L., Bernyk M., Croton D. J., Koribalski B. S., Gerstmann D., Westerlund S., 2012, [Monthly Notices of the Royal Astronomical Society](#), 426, 3385

- Elmegreen B., 1993, *The Astrophysical Journal*, 411, 170
- Elson E., Blyth S., Baker A., 2016, *Monthly Notices of the Royal Astronomical Society*, 460, 4366
- Elson E., Kam S., Chemin L., Carignan C., Jarrett T., 2019a, *Monthly Notices of the Royal Astronomical Society*, 483, 931
- Elson E., Baker A., Blyth S., 2019b, *Monthly Notices of the Royal Astronomical Society*, 486, 4894
- Fabello S., 2012, PhD thesis, München, Ludwig-Maximilians-Universität, Diss., 2012, doi:10.5282/edoc.14509
- Fabello S., Catinella B., Giovanelli R., Kauffmann G., Haynes M. P., Heckman T. M., Schiminovich D., 2011, *Monthly Notices of the Royal Astronomical Society*, 411, 993
- Feldmann R., 2020, *Communications Physics*, 3, 226
- Geman S., Geman D., 1984, *IEEE Transactions on pattern analysis and machine intelligence*, pp 721–741
- Geréb K., Morganti R., Oosterloo T., Hoppmann L., Staveley-Smith L., 2015, *Astronomy & Astrophysics*, 580, A43
- Giovanelli R., Haynes M. P., 2002, *The Astrophysical Journal*, 571, L107
- Giovanelli R., et al., 2005, *The Astronomical Journal*, 130, 2613
- Glowacki M., Elson E., Davé R., 2020, *Monthly Notices of the Royal Astronomical Society*, 498, 3687
- Gupta Y., et al., 2017, *Current Science*, pp 707–714
- Haynes M. P., et al., 2018, *The Astrophysical Journal*, 861, 49
- Healy J., 2021, | 10.33612/diss.177293235
- Healy J., Blyth S., Elson E., van Driel W., Butcher Z., Schneider S., Lehnert M., Minchin R., 2019, *Monthly Notices of the Royal Astronomical Society*, 487, 4901
- Holwerda B., Blyth S.-L., Baker A., et al., 2011, *Proceedings of the International Astronomical Union*, 7, 496

- Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012, [The Astrophysical Journal](#), 756, 113
- Huchra J., Geller M., 1982, [The Astrophysical Journal](#), 257, 423
- Janowiecki S., Catinella B., Cortese L., Saintonge A., Wang J., 2020, [Monthly Notices of the Royal Astronomical Society](#), 493, 1982
- Jian H.-Y., et al., 2018, [Publications of the Astronomical Society of Japan](#), 70, S23
- Jonas J., Team M., 2016, [MeerKAT Science: On the Pathway to the SKA](#), p. 1
- Jones M., 2016
- Kale R.,
- Kannappan S. J., et al., 2013, [The Astrophysical Journal](#), 777, 42
- Katz N., Keres D., Dave R., Weinberg D. H., 2003, [The IGM/Galaxy Connection: The Distribution of Baryons at \$z=0\$](#) , pp 185–192
- Kennicutt Jr R. C., 1998, [The astrophysical journal](#), 498, 541
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, [Monthly Notices of the Royal Astronomical Society](#), 363, 2
- Kilborn V. A., Forbes D. A., Barnes D. G., Koribalski B. S., Brough S., Kern K., 2009, [Monthly Notices of the Royal Astronomical Society](#), 400, 1962
- Kim S., Jeong H., Rey S.-C., Lee Y., Lee J., Joo S.-J., Kim H.-S., 2020, [The Astrophysical Journal](#), 903, 65
- Lagos C. d. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, [Monthly Notices of the Royal Astronomical Society](#), 418, 1649
- Lagos C. D. P., Baugh C. M., Zwaan M., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A., Van Kampen E., 2014, [Monthly Notices of the Royal Astronomical Society](#), 440, 920
- Lah P., et al., 2007, [Monthly Notices of the Royal Astronomical Society](#), 376, 1357
- Lah P., et al., 2009, [Monthly Notices of the Royal Astronomical Society](#), 399, 1447
- Landy S. D., Szalay A. S., 1993, [The Astrophysical Journal](#), 412, 64

- Leroy A. K., Walter F., Brinks E., Bigiel F., De Blok W., Madore B., Thornley M., 2008, [The astronomical journal](#), 136, 2782
- Leroy A. K., Walter F., Schruba A., Collaboration H., et al., 2012, in American Astronomical Society Meeting Abstracts# 219. pp 346–03
- Liu H. B., Hsieh B., Ho P. T., Lin L., Yan R., 2008, [The Astrophysical Journal](#), 681, 1046
- Maddox N., Hess K. M., Obreschkow D., Jarvis M., Blyth S.-L., 2015, Monthly Notices of the Royal Astronomical Society, 447, 1610
- Maddox N., et al., 2021, [Astronomy & Astrophysics](#), 646, A35
- Mahajan S., Raychaudhury S., 2009, Monthly Notices of the Royal Astronomical Society, 400, 687
- Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, [The Astrophysical Journal](#), 723, 1359
- McGaugh S. S., Schombert J. M., Bothun G. D., De Blok W., 2000, [The Astrophysical Journal](#), 533, L99
- Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, [Publications of the Astronomical Society of Australia](#), 34
- Obreschkow D., 2009, Technical report, The Cosmic Evolution of Atomic and Molecular Hydrogen in Galaxies. University of Oxford
- Obreschkow D., Meyer M., 2014, [arXiv preprint arXiv:1406.0966](#)
- Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009a, [The Astrophysical Journal](#), 698, 1467
- Obreschkow D., Heywood I., Klöckner H.-R., Rawlings S., 2009b, [The Astrophysical Journal](#), 702, 1321
- Obreschkow D., Klöckner H.-R., Heywood I., Levrier F., Rawlings S., 2009c, [The Astrophysical Journal](#), 703, 1890
- Popping G., Somerville R. S., Trager S. C., 2014, [Monthly Notices of the Royal Astronomical Society](#), 442, 2398

- Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, [The Astrophysical Journal](#), 635, 123
- Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., Hulst T. v. d., 2013, [Monthly Notices of the Royal Astronomical Society](#), 435, 2693
- Rhee J., et al., 2023, [Monthly Notices of the Royal Astronomical Society](#), 518, 4646
- Robotham A. S., et al., 2011, [Monthly Notices of the Royal Astronomical Society](#), 416, 2640
- Sahinler S., Topuz D., 2007, [Journal of Applied Quantitative Methods](#), 2, 188
- Shao J., Tu D., 2012, [The jackknife and bootstrap](#). Springer Science & Business Media
- Sinigaglia F., et al., 2022, [The Astrophysical Journal Letters](#), 935, L13
- Stevens A. R., et al., 2019, [Monthly Notices of the Royal Astronomical Society](#), 483, 5334
- Swarup G., Ananthakrishnan S., Subrahmanya C., Rao A., Kulkarni V., Kapahi V., 2000, [GEOPHYSICAL MONOGRAPH-AMERICAN GEOPHYSICAL UNION](#), 119, 297
- Tacconi L., et al., 2013, [The Astrophysical Journal](#), 768, 74
- Trachternach C., De Blok W., McGaugh S., Van der Hulst J., Dettmar R.-J., 2009, [Astronomy & Astrophysics](#), 505, 577
- Tully R. B., Fisher J. R., 1977, [Astronomy and Astrophysics](#), 54, 661
- Tumlinson J., et al., 2011, [Science](#), 334, 948
- Verheijen M., 2001, in [Gas and Galaxy Evolution](#). p. 573
- Walter F., Brinks E., De Blok W., Bigiel F., Kennicutt R. C., Thornley M. D., Leroy A., 2008, [The Astronomical Journal](#), 136, 2563
- Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., N. Chengalur J., 2016, [Monthly Notices of the Royal Astronomical Society](#), 460, 2143
- Yu C. H., 2002, [Practical Assessment, Research, and Evaluation](#), 8, 19

Zwaan M. A., 2000

Zwaan M., Van Dokkum P., Verheijen M., 2001, [Science](#), 293, 1800

Zwaan M. A., Meyer M., Staveley-Smith L., Webster R., 2005, [Monthly Notices of the Royal Astronomical Society: Letters](#), 359, L30

de Blok W., Fraternali F., Heald G., Adams E., Bosma A., Koribalski B., et al., 2015, [arXiv preprint arXiv:1501.01211](#)

