University of Cape Town

Thesis presented for the degree of Master of Science

in the Department of Astronomy

Using HI stacking to study galaxy properties in the nearby universe

(Looking for needles in the HI-stack)

Author:

Julia L. Healy

Supervisors:

Dr S.L Blyth

Dr E. Elson

October 2016

Plagiarism Declaration

I, Julia Healy, know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Abstract

Neutral atomic hydrogen (HI) is the raw fuel from which the star-forming molecular gas forms and is therefore an important tracer of galaxy evolution. Due to the intrinsic faintness of the HI emission line (observed at rest at 21 cm), galaxies beyond a few hundred megaparsecs are difficult to observe directly with current radio telescopes. However, in the next year, MeerKAT and other SKA pathfinder telescopes will begin operating and enable deeper, large surveys (e.g. LADUMA) of neutral gas in galaxies. HI Stacking is an observational technique that will be highly exploited to learn about the HI content of galaxies that are not directly detected. Stacking involves combining the HI spectra of all the galaxies in a distant sample, thereby generating a high signal-to-noise measure of their average HI content.

This work presents a new Python-based package capable of stacking HI galaxy spectra. This package will be used to stack the HI spectra of high-redshift galaxies observed with the MeerKAT telescope. In this work the package is applied to a sample of galaxies observed as part of the Nançay Interstellar Baryon Legacy Extragalactic Survey (NIBLES, van Driel et al. (2016)) to learn more about the gas properties of galaxies in the local universe. Using the stacking technique, we are able to recover the average HI mass of different galaxy populations for which there was no HI directly detected. In order to obtain the average gas properties that best represent the overall galaxy populations, we also stack both the HI detected and non-detected spectra from the NIBLES survey. We find that our gas fraction vs. stellar mass distribution results agree well with previous stacking experiments (Brown et al., 2015; Catinella et al., 2010; Fabello et al., 2011a) and the NIBLES sample enables us to probe an order of magnitude lower in stellar mass. We find a dependence on the underlying stellar mass distribution for our gas fraction vs. $NUV - r$ colour, especially when comparing to Brown et al. (2015) and Fabello et al. (2011a).

Acknowledgements

Undertaking MSc research is a collaborative effort, and as such, this work would not have been possible without the support of a couple of key people. Most importantly, I would like to express my deep appreciation and gratitude to my supervisors Dr Sarah Blyth and Dr Ed Elson. Their guidance and support over the last two years has been nothing short of incredible. Ed and Sarah, I am endlessly thankful for your seemingly unending patience, especially when it came to those early drafts.

Lauren, Tam, Naomi, Kerry, and Jamie, thank you for your friendship and company as we discussed anything and everything (occasionally including our work) over a drink and food at the pub. Jamie, thank you for being my sounding board and letting me talk at you. Lauren, I appreciate you letting me rest on the couch in your office when I needed a break. I am thankful to all my friends who have been very patient with me as I finish up this thesis.

To my dearest family, none of this would have become a reality without your love and support. Adam, thank you for always keeping me humble and teaching me how to question properly. William, without you I think I would have starved. Willo, thank you for putting up with me when the times got stressed and for keeping me fed and happy – your hugs do solve a number of problems. Mom and Dad, your continuing love and support as I follow my passions means the world to me. Dad, thank you for all your help on both the software development and all the proofreading.

I would like to thank SKA SA for providing the funding that made this work possible. This work has made use of data from the NIBLES survey. I would to thank Wim van Driel and Zhon Butcher for providing access to the data. I would also like to thank Zhon for all his help as I worked to understand the data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

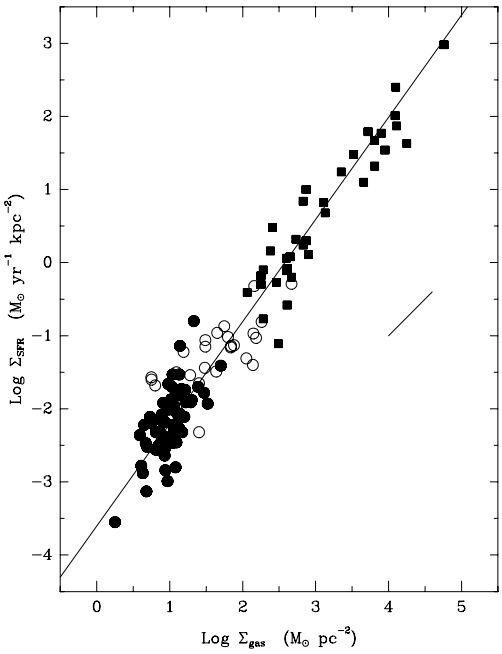
1.1 The role of gas in galaxy evolution

How galaxies evolve over cosmic time is currently a key area of research in Astrophysics. The evolution of galaxies is most easily traced by the formation of stars. Through the star formation process, the interstellar medium of galaxies is enriched with metals and energy from the explosive processes at the end of stellar evolution; in some cases the stellar evolution process may eject gas thus enriching the intergalactic medium (IGM) (de Blok et al., 2015). In low-mass galaxies with shallow gravitational potential wells, the metals ejected from supernovae can be expelled into the IGM (Mac Low & Ferrara, 1999). Neutral atomic hydrogen (HI) is the raw fuel for star formation, and can be traced by its emission line which has a rest wavelength of 21 cm.

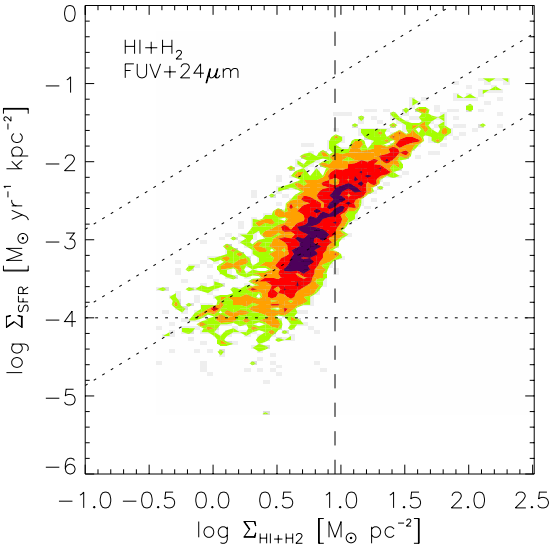
The Kennicutt-Schmidt Law is an empirical law that relates the star formation rate surface density to the gas surface density on global (Figure 1.1a) and kilo-parsec (Figure 1.1b) scales (Kennicutt, 1989; Kennicutt, Jr., 1998; Bigiel et al., 2008; Leroy et al., 2012). Figure 1.1 highlights the connection between the gas and star formation. However understanding exactly what conditions are needed for stars to form is not yet clear due to, in part, the limitations of current technology (current radio telescopes do not have the ability to resolve HI emission at parsec scales in external galaxies). It is known that the neutral hydrogen plays a crucial role in star formation by providing the raw fuel. In order to properly understand how galaxies grow in mass, it is essential to understand what processes govern the supply of neutral hydrogen to galaxies, and how that supply affects star formation.

The processes by which gas, more specifically the HI, enters galaxies from the IGM, and the manners in which it is subsequently expelled by star formation back into the IGM, is usually referred to as the "gas cycle". It is thought that gas accretion from the IGM along the filaments of the cosmic web is how galaxies are able to sustain the observed star formation rates (e.g. Sancisi et al. (2008)).

Simulations indicate that there may be two main modes of gas accretion, "cold mode" and "hot mode" accretion (Kereš et al., 2005). In "cold mode" accretion, the gas flows into the dark matter halo along filaments of the cosmic web. However in "hot mode" accretion, gas is accreted "quasi-spherically" (Kereš et al., 2005) and is shock-heated as it enters the halo before cooling and condensing to form stars. Kereš et al. (2005) found



(a) *The Global Kennicutt-Schmidt Law for a 61 nearby spiral galaxies. The solid circles indicate normal disk galaxies, and the squares represent starburst galaxies (Kennicutt, Jr., 1998, Fig. 6).*



(b) *The combined data from seven spiral galaxies at 750 pc resolution. This figure shows star formation rate surface density (Σ_{SFR} [$M_\odot \cdot yr^{-1} \cdot kpc^{-2}$]) as a function of the total gas surface density (Σ_{gas} [$M_\odot \cdot pc^{-2}$]) (Bigiel et al., 2008, Figure 8).*

Figure 1.1: *Figure 1.1a shows the relationship between gas surface density and star formation rate surface density on a global scale. The relationship for star formation rate surface density on kiloparsec scales is shown for the total gas surface density in Figure 1.1b.*

that "cold mode" accretion dominates for low mass galaxies (thus is the dominant gas accretion process for galaxies at $z \gtrsim 3$), while "hot mode" dominates the high mass galaxies making this mode the dominant process in the low density environments at low redshift. Observations support the Kereš et al. (2005) findings that "cold mode" accretion is not the dominant process at low redshift as the observed accretion is an order of magnitude too low to support the ongoing star formation (Sancisi et al., 2008; Putman et al., 2012).

1.2 Evolution of Observable Galaxy Properties

Large optical surveys such as the Sloan Digital Sky Survey (York et al., 2000, SDSS) have made it possible to study the relationship between optical galaxy properties and galaxy evolution. Figure 1.2 shows the colour-magnitude (CM) diagram for a selection of SDSS galaxies. The blue disk-dominated galaxies occupy the bottom of the diagram, while the red, more spheroidal galaxies occupy the top of the diagram. It was found later when studying the CM diagram for galaxies of all morphologies, that the colour distribution and can be approximated by the sum of two Gaussian functions (Strateva et al., 2001). The explanation given for the bimodality is that there are two distinct galaxy populations which are governed by different evolutionary processes (e.g. Baldry et al. (2004) [B04]; Bell et al. (2004), Taylor et al. (2014), etc.). Using a sample of more than 200 000 galaxies with photometry and spectroscopy from SDSS, B04 sought to trace the CM bimodality without separating the sample into different morphologies. They found that the sum of a

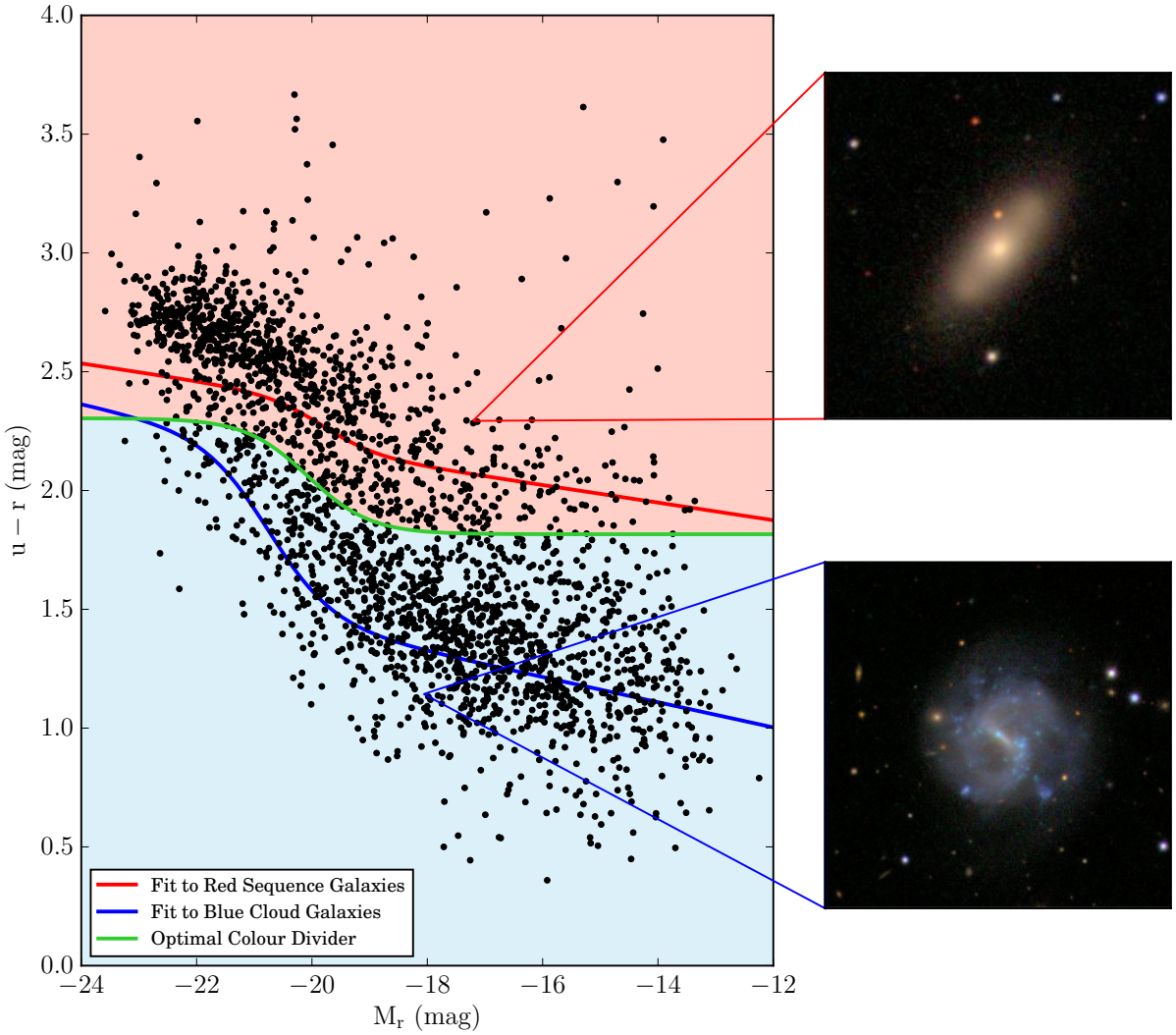


Figure 1.2: *Optical Colour-Magnitude Diagram, the data points are taken from a sample of SDSS galaxies and the fitted lines are taken from the work of B04. The u- and r-band magnitudes have been corrected for galactic extinction.*

straight line and a hyperbolic tan function trace the distributions of blue and red galaxies (indicated by the red and blue lines in Figure 1.2). The optimal divide between these two distributions is shown by the green line in Figure 1.2 and is given by,

$$\mathcal{T}(M_r) = p_0 + p_1(M_r + 20) + q_0 \tanh\left(\frac{M_r - q_1}{q_2}\right), \tag{1.1}$$

where $p_0 = 2.06$, $p_1 = 0$, $q_0 = -0.244$, $q_1 = -20.07$, and $q_2 = 1.09$ for the optimal colour divider (Baldry et al., 2004, Eq. 9).

The question of how galaxies evolve on the CM diagram remains open-ended. The region between the red and blue populations, the so-called "green valley", is thought to be a transitional region for galaxies. Initially, galaxies were thought to evolve from the "blue cloud" through the green valley to the "red sequence". This was supported by the fact that very few green valley objects have been observed, and those that have been observed have lower star formation rates than the main sequence galaxies of the same stellar mass (M$_\star$). Thus green valley objects typically have lower specific star formation

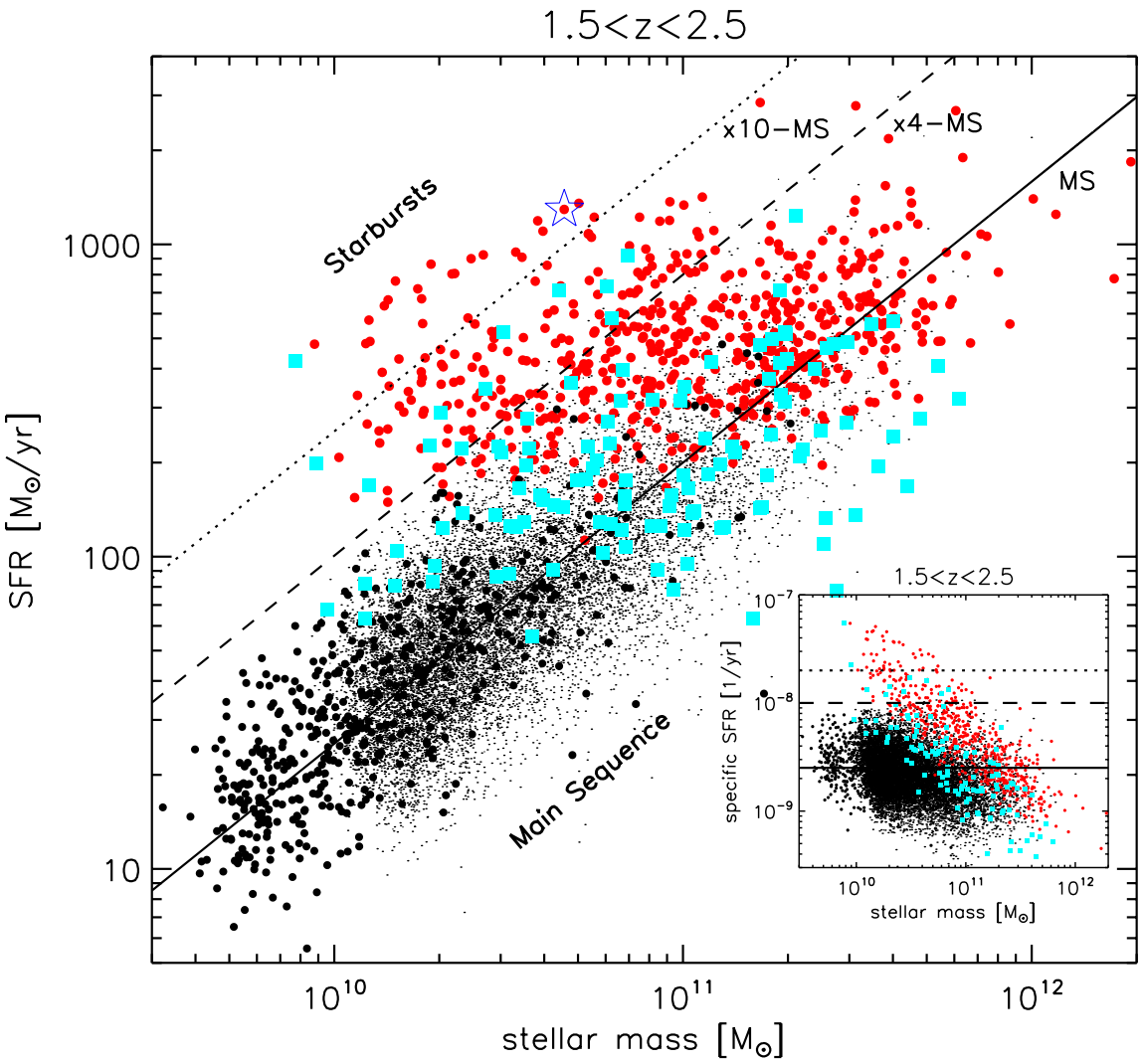


Figure 1.3: *The galaxy main sequence (indicated by the solid line labelled MS) is populated by regular star-forming galaxies. Starburst galaxies occupy the space above the main sequence, while the area below the main sequence is occupied by galaxies in the green valley, and even lower, the red non star-forming galaxies. The solid, dashed, and dotted lines trace the constant specific star formation rate. The subfigure in the bottom right corner shows the specific star formation rate per year for the same sample as a function of stellar mass. (Rodighiero et al., 2011, Figure 1)*

rates (sSFR) which is defined as the ratio of the star formation rate (SFR) to stellar mass (SFR/$M_\star$). Figure 1.3 shows the galaxy main sequence for a sample of galaxies at high redshift. The galaxy main sequence is sometimes referred to as the star-forming galaxy main sequence as it is defined by the correlation between star formation rate and stellar mass for star-forming galaxies. The galaxy main sequence is defined as the constant specific star formation rate (sSFR) which traces star-forming galaxies for stellar mass vs star formation rate (SFR). Galaxies lying below the main sequence have a lower sSFR which can be an indicator of quenched star formation (Schawinski et al., 2014, and references therein).

Schawinski et al. (2014, S14) used morphological classifications from Galaxy Zoo to separate the early-type (typically red) and late-type (typically blue) galaxies to study

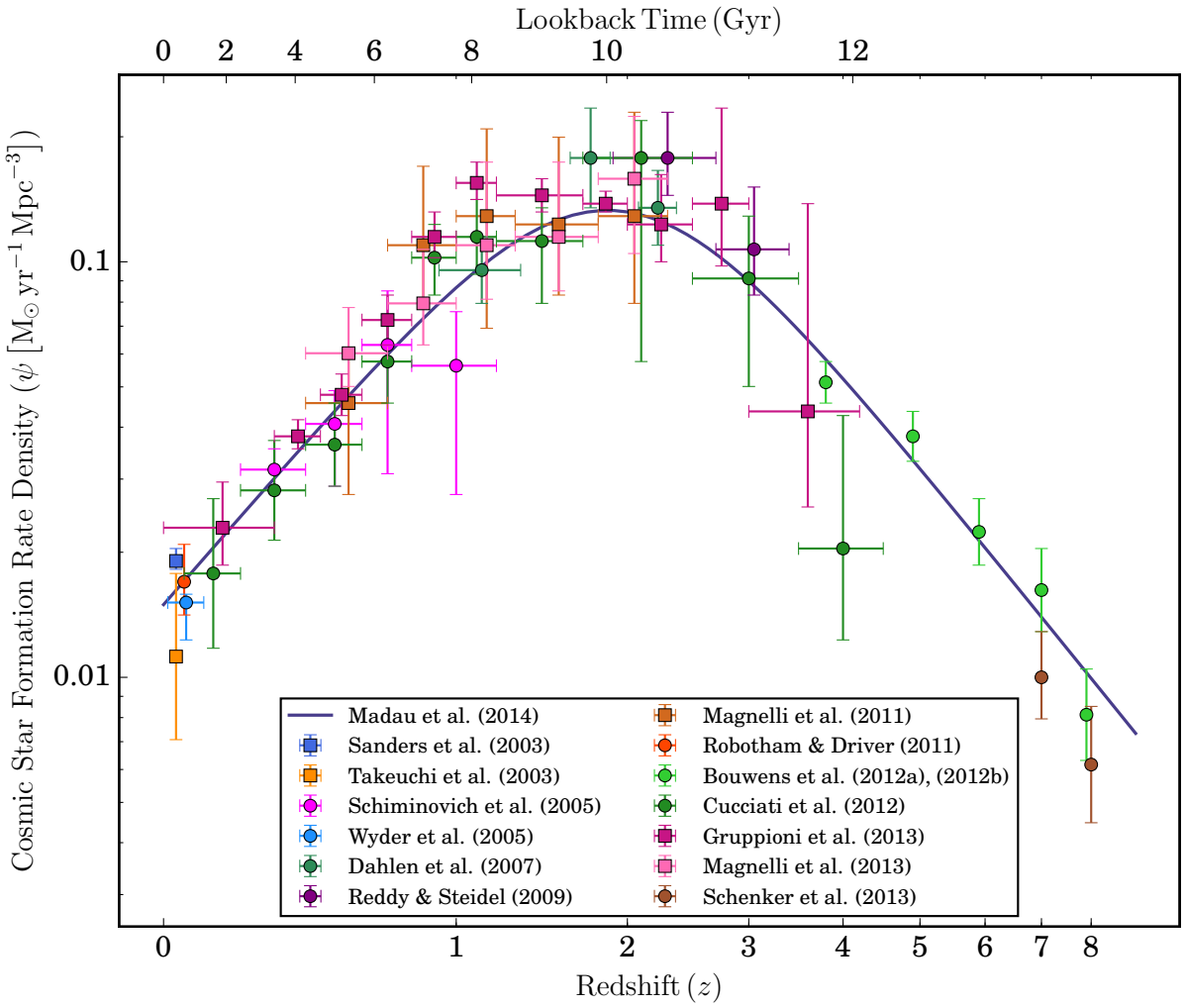


Figure 1.4: *Cosmic Star Formation Rate Density as a function of redshift and look-back time. The two symbols in the plot above indicate the type of measurement: the square points indicate infrared measurements, and the round points indicate ultraviolet measurements that have been corrected for dust attenuation. The colours of each data point correspond to the author responsible for the measurements; the authors are listed in the legend above. The data used to create this plot are taken from Madau & Dickinson (2014, Table 1.)*

their evolution independently. S14 found that both early- and late-type galaxies enter the green valley as their sSFR decreases, but the decrease in sSFR occurs at different rates – quickly for early-type galaxies, and more slowly for late-type galaxies. S14 concluded that each population follows a different evolutionary track due to the different time scales over which star formation is quenched. Thus, considering the green valley as a transition zone for all types of galaxy morphologies is an over simplification.

Kannappan et al. (2013) argue that while the current models of galaxy evolution support galaxy mergers as a way to build the observed large scale structure, the models do not explain the observed abundance of disk-dominated galaxies over bulge-dominated galaxies. The bimodality observed in the CM diagram by Baldry et al. (2004, etc.) occurs at a characteristic mass of $M_\star \sim 10^{10.5}\,M_\odot$ (Kauffmann et al., 2004). According to Kannappan et al. (2013) there is not just the bimodality scale, but also what they call the "gas-richness threshold" which signals the transition in galaxies from gas-rich to gas-poor

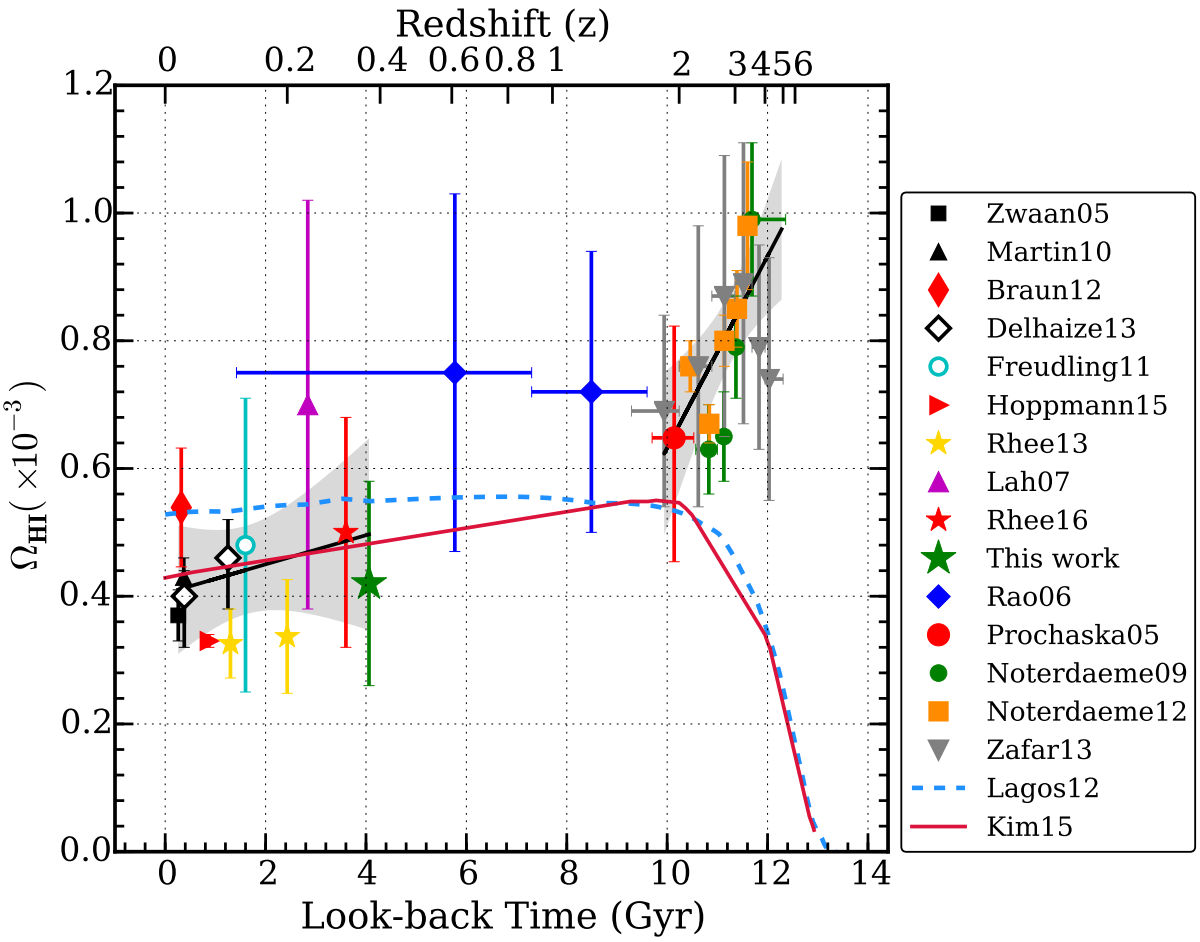


Figure 1.5: *Cosmic* HI *density ($\Omega_{\rm HI}$) measurements collated by Rhee et al. (2016, Figure 11). There are three different methods of determining the* HI *density shown in this plot: absorption measurements (DLAs and MgII absorbers) (Noterdaeme et al., 2009, 2012; Zafar et al., 2013; Prochaska et al., 2005; Rao et al., 2006),* HI *stacking (Rhee et al., 2016; Delhaize et al., 2013; Rhee et al., 2013; Lah et al., 2007), the remaining $\Omega_{\rm HI}$ measurements in this plot are made by directly detecting the* HI *in galaxies (Zwaan et al., 2005; Martin et al., 2010; Braun, 2012; Hoppmann et al., 2015). The colours of each data point correspond to the author responsible for the measurements; the authors are listed in the legend above.*

which corresponds to $M_\star \sim 10^{9.5-10}\,M_\odot$. Exploration of the galaxy properties for samples that span both threshold scales is needed to understand why disk-like galaxies dominate, as both scales are evidenced by changes in both gas-richness and morphology.

The improvement in recent years of ultraviolet (UV) and infrared (IR) facilities such as the Galaxy Evolution Explorer (UV; GALEX) and the Spitzer Space Telescope (IR) have allowed the star formation rates of galaxies to be studied as a function of cosmic time. The results of these studies (shown in Figure 1.4) show a smooth increase in the star formation rate from redshift (z) of 8 to a peak at $z \sim 2$ and then a decrease by an order of magnitude to the value at the present time. Figure 1.4 shows the evolution of the cosmic star formation rate density as a function of both redshift and look-back time where the UV measurements are indicated by the round data points while the square data points indicate the measurements made in the IR regime.

The cosmic neutral hydrogen density ($\Omega_{\rm HI}$) is the neutral gas analogue to the cosmic

star formation rate density, and is an important parameter in galaxy evolution models. Since HI provides raw fuel for star formation, knowing the $\Omega_{\rm HI}$ at different epochs gives insight into how much gas was available for star formation. However obtaining measures of the $\Omega_{\rm HI}$ beyond the local universe ($z \sim 0$) is challenging due to the long observing hours required with current telescopes (Zwaan et al., 1997, 2003, 2005; Martin et al., 2010; Freudling et al., 2011; Braun, 2012; Hoppmann et al., 2015). At high redshifts ($z > 0.5$), indirect methods of detecting the presence of HI must be used. For $z > 2$, $\Omega_{\rm HI}$ can be indirectly measured from Damped Lyman-α (DLA) systems. DLA systems are absorption systems – the presence of HI is inferred from the broad absorption features in the Lyman-α forest of quasar spectra (e.g. Prochaska et al. (2005); Noterdaeme et al. (2009); Sánchez-Ramírez et al. (2015)). Below $z < 2$, the Lyman-α lines are observed in the UV part of the spectrum, which makes it difficult to observe using ground-based facilities. Rao et al. (2006) find that the MgII absorption doublet is a good alternative for tracing low redshift DLA systems as it appears in the optical part of the spectrum for $z > 0.11$. At intermediate redshifts ($0.1 < z < 0.5$), HI stacking analyses have been used with increasing frequency to study $\Omega_{\rm HI}$ (?Delhaize et al., 2013; Rhee et al., 2013, 2016). Stacking is a statistical technique that can be used to obtain a measure of the HI content for a sample of galaxies for which there may be no individual detections in the HI emission spectra. The technique uses the optical positions and redshifts to co-add the HI spectra in order to create an average HI spectrum from which the average HI mass per galaxy can be obtained (this technique is discussed in detail in Chapter 2).

?? shows the measurements of $\Omega_{\rm HI}$ made to date. $\Omega_{\rm HI}$ has been well sampled at both high ($z > 2$) redshift and low ($z \sim 0$) redshift, and while there are measurements at the intermediate redshifts ($0.1 < z < 1.4$), these measurements are not sufficient to constrain the evolution of $\Omega_{\rm HI}$ due to the large measurement uncertainties (this region is shaded blue in ??). Attempts have been made to quantify the evolution of $\Omega_{\rm HI}$, however conclusions will need to await higher precision measurements at intermediate redshifts (Sánchez-Ramírez et al., 2015).

1.3 HI in the Local Universe

Radio Astronomy and the study of the HI content of the universe is a relatively young area of astronomical research. The first radio waves from an extraterrestrial source were discovered by Karl Jansky in the 1930s. Jansky associated the radio signal with the dust and gas located at the centre of the galaxy. In 1945, H.C. Van de Hulst determined that the hyperfine transition in the electronic ground state of a neutral hydrogen atom produces a photon which corresponds to a wavelength of 21cm (1420 MHz) which lies in the radio range. The spin of the electron flips to a lower energy state releasing a photon of 5.87 μeV in the process (corresponding to a wavelength of 21 cm). The hyperfine transition has an extremely low transition probability, however due to the very high abundance of HI in the universe and the very low density of the interstellar gas, the transition occurs often enough that it can be used to trace the HI content of galaxies. Due to the low flux density of the HI emission line, sensitive instruments are required to detect it.

Various HI surveys of the local universe have been carried out over the last few decades using single dish telescopes to survey large areas on the sky, and synthesis telescopes (radio interferometers) to carry out detailed, resolved studies of select samples of nearby galax-

ies. Large single dish radio telescopes are very sensitive instruments, and are thus very useful for studying the global HI properties of galaxies but they lack the spatial resolution needed resolve galaxies (except for very nearby galaxies). The 305 m Arecibo telescope has the ability to resolve sources larger than 3.5′. Synthesis radio telescopes have the spatial resolution necessary for detailed studies of the kinematics and the distribution of HI within galaxies, for example the Jansky Very Large array operating in the A configuration can resolve structures larger than 1.3″[1] for very nearby galaxies. Detailed studies of the HI kinematics of galaxies include the Westerbork HI Survey of Irregular and Spiral Galaxies (WHISP; van der Hulst et al. (2001)) using the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands. WHISP was a targeted survey of 400 early- and late-type galaxies, using the HI as a tracer to study the structure and kinematics of the galaxies in the sample. The HI Nearby Galaxy Survey (THINGS; Walter et al. (2008)) looked at the spatially resolved HI in a sample of 30 nearby galaxies using the Very Large Array in the USA. Due to the limitations of current technology requiring long observation periods, the aforementioned surveys have typically been limited to $D < 15\,\mathrm{Mpc}$. Large blind HI single dish surveys such as the HI Parkes All Sky Survey (HIPASS; Barnes et al. (2001)) and the Arecibo Legacy Fast ALFA[2] survey (ALFALFA; Giovanelli et al. (2005)) have provided a picture of the HI around galaxies for a broad range of masses. HIPASS (Meyer et al., 2004) and its northern extension (Wong et al., 2006) were able to detect the HI in more than 4300 and 1000 galaxies respectively, wwhile the second release ($\alpha.70$) (Haynes et al., 2011) of the more sensitive ALFALFA data contains HI detections for more than 20 000 galaxies.

Due to sensitivity limits, the blind HI surveys are typically biased toward gas-rich blue cloud galaxies (Huang et al., 2012). Huang et al. (2012) go on to show that more than 80% of galaxies in an HI-selected HI sample are classified blue by Baldry et al. (2004), while only 67% of an optically-selected HI sample lie in the same region of the CM diagram. To properly understand the role of HI in the evolution of galaxies, it is important to also study the gas-poor galaxies to gain more insight into what drives them to be gas-poor. Recent HI surveys such as the GALEX Arecibo SDSS Survey (GASS; Catinella et al. (2010)) have carried out targeted HI observations for a sample of galaxies selected from the overlap of SDSS and GALEX. One of the main aims of GASS was to explore the $\mathrm{M_{HI}}$-to-$\mathrm{M_\star}$ (f_{HI}) scaling relations for a sample of galaxies with $\mathrm{M_\star} > 10^{10}\,\mathrm{M_\odot}$. The gas scaling relations are a useful tool to study various galaxy properties as a function of how gas-rich or gas-poor a galaxy may be. Blue, star-forming (typically gas-rich) galaxies have been found to have a higher f_{HI} than red, passive galaxies (typically gas-poor). A more recent survey studied the HI content as a function of stellar mass ranging from $10^6\,\mathrm{M_\odot} < \mathrm{M_\star} < 10^{12}\,\mathrm{M_\odot}$; this survey, the Nançay Interstellar Baryon Legacy Extragalactic Survey (NIBLES) was conducted using the 100 m Nançay Radio Telescope in France (van Driel et al., 2016). The details of NIBLES are discussed in detail in Chapter 4.

One of the greatest successes of HIPASS was the measurement of the HI Mass Function (Zwaan et al., 2003) for the local Universe. Figure 1.6 shows the more recently measured HIMF derived from the ALFALFA $\alpha.40$ catalogue (Haynes et al., 2011). The HI Mass Function (HIMF) is the HI equivalent of the optical galaxy luminosity function. The HIMF in conjunction with the luminosity function is required to link galaxy formation models to the observed local universe. The luminosity function provides a measure of the

[1] `https://science.nrao.edu/facilities/vla/docs/manuals/oss/performance/resolution`
[2] Arecibo L-band Feed Array

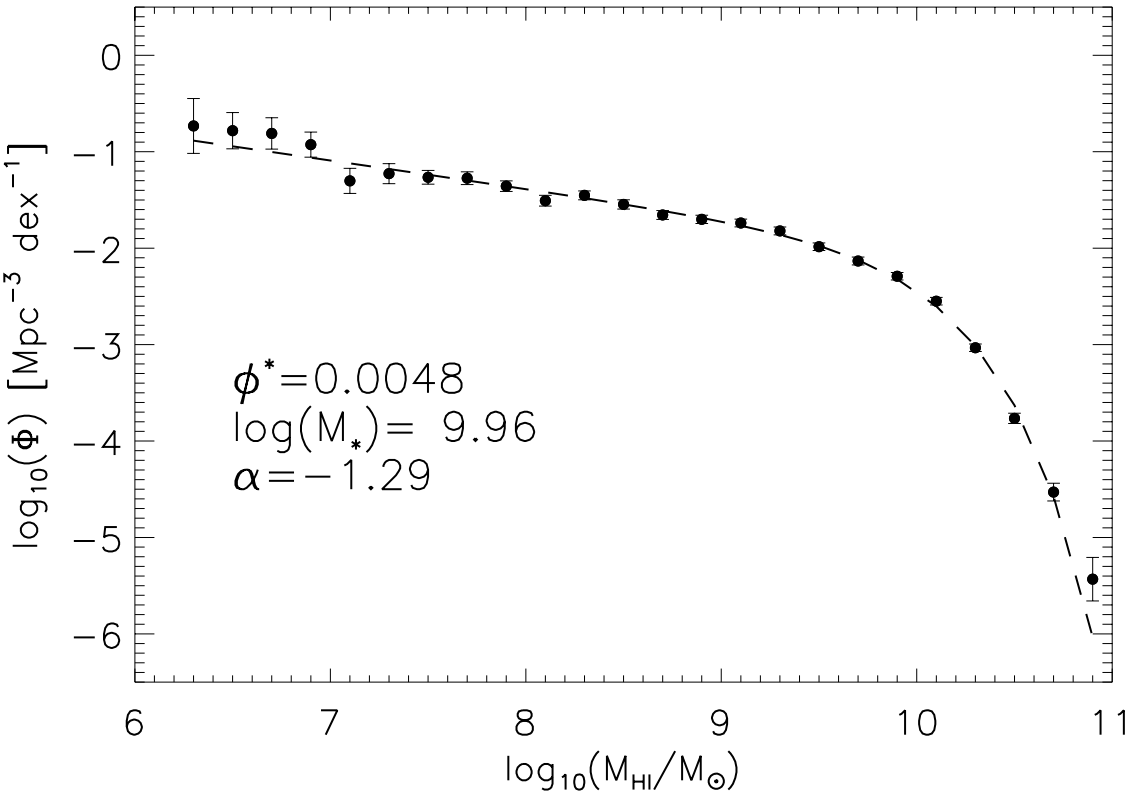


Figure 1.6: *The ALFALFA HIMF calculated from the α.40 catalogue. The best-fit Schechter (1976) function is plotted using the dashed lines, and the best-fit parameters are listed on the plot. ϕ^* is the normalisation factor, $\log(M_*)$ is the galaxy characteristic mass and defines the "knee" of the function. The slope of the low-mass end of the function is defined by α (Haynes et al., 2011, Fig. 15).*

ways in which stellar mass is distributed among galaxies, while the HIMF contains the information about the distribution of the mass of the gas which provides the raw fuel for the formation of stars (Zwaan et al., 2003). The HIMF is a way to measure the amount of HI in the universe as it gives the number density of galaxies of a particular mass, and is characterised by the Schechter (1976) function below,

$$\theta(\mathrm{M_{HI}})\mathrm{dM_{HI}} = \theta^* \left(\frac{\mathrm{M_{HI}}}{\mathrm{M^*_{HI}}}\right)^{\alpha} \exp\left(-\frac{\mathrm{M_{HI}}}{\mathrm{M^*_{HI}}}\right) \mathrm{dM_{HI}}. \tag{1.2}$$

There are a number of important quantities that can be obtained from the HIMF; Ω_{HI} is calculated by integrating the HIMF. The slope of the faint end of the HIMF is an input to the galaxy formation models (Zwaan et al., 1997) as it contains information about the dwarf galaxy population. The "knee" of the function defines the characteristic galaxy mass ($\mathrm{M^*_{HI}}$) above which the volume density of galaxies drops off exponentially.

The Arecibo Ultra Deep Survey (AUDS; Freudling et al. (2011); Hoppmann et al. (2015)) used the Arecibo telescope to detect HI out to $z \sim 0.1$, the highest redshift probed with HI emission at that time. AUDS is one of the most sensitive blind single dish surveys to date. The main aim was to probe the HI content around $z \sim 0.1$. Hoppmann et al. (2015) measured the HIMF using 102 detected galaxies. Due to the wide spectral range of the AUDS data, Hoppmann et al. (2015) were able to look at both the HIMF

and $\Omega_{\rm HI}$ in a number of different redshift bins using the same dataset. They found that there is little evolution in the redshift range studied. In particular the HIMF calculated for $z > 0.06$ is equivalent to the HIMF for the local universe which they claim is evidence that there has been no evolution in the HIMF for the last Gyr (Hoppmann et al., 2015).

Three recent surveys have made use of existing radio observatories to observe HI emission for $z > 0.2$.

1. The HIGHz survey (Catinella & Cortese, 2014) conducted using the 305 m Arecibo Telescope observed the HI in 39 high stellar mass ($M_\star > 10^{10}\,M_\odot$) galaxies.

2. The Blind Ultra Deep HI Environmental Survey (BUDHIES; Verheijen et al. (2007); Jaffé et al. (2013)) used the WSRT to probe the structure of two Abell clusters and to study the gas content of galaxies within the cluster environment.

3. The COSMOS[3] HI Large Extragalactic Survey (CHILES; Fernández et al. (2013), in progress), is currently being conducted using the recently upgraded Jansky Very Large Array. CHILES currently holds the record for imaging HI at the highest redshift at $z = 0.376$ (Fernández et al., 2016). CHILES aims to probe the HI content to high redshift ($z \sim 0.5$), and will be one of the first surveys to do so (Fernández et al., 2016).

AUDS, HIGHz, BUDHIES, and CHILES are currently probing higher redshifts, but due to the faintness of the HI emission and the long observing times required, only the highest mass galaxies can be directly detected. This means that most of the HIMF cannot be directly probed, thus statistical techniques such as stacking become an important tool to study the HI properties of the undetected galaxies, thereby allowing us to probe in a much more complete manner the full galaxy population for a particular cosmological volume. As a result of the low HI emission detection numbers, the above mentioned survey teams have used techniques such as HI stacking to obtain average measures of the HI content in their galaxy samples.

1.4 HI Stacking

Stacking is a statistical technique that takes a sample of spectra that are predominantly classified as non-detections to create a high S/N spectrum representative of the total HI mass in a galaxy sample, and hence also the average HI mass of the sample. Non-detections are the HI spectra of galaxies for which the no HI was directly detected. The stacking technique requires the spectra to be aligned in the rest frame before the sample is co-added to create the average spectrum. Stacking makes it possible to extract average information from a sample of galaxy spectra that would not have been useful otherwise. This method is discussed in detail in Chapter 2.

Published stacking studies have focused on four main areas: HI content of galaxies in dense environments and how the gas content relates to other observables, gas content of active galaxies, the relations between HI and various star stellar mass/star formation indicators, and using stacking to measure $\Omega_{\rm HI}$ at low to intermediate redshifts ($z < 0.4$).

[3]Cosmic Evolution Survey

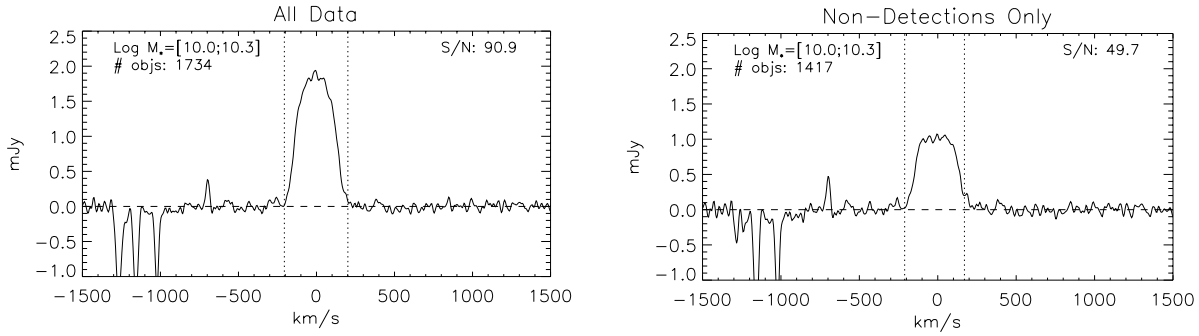


Figure 1.7: *Fabello et al. (2011a, Figure 6) demonstrates that stacked spectra produced from samples containing both detections and non-detections are not dominated by the flux from the galaxies for which there was a direct* HI *detection.*

Some notable results include the contributions from three different groups that have successfully used stacking to determine $\Omega_{\rm HI}$ (**?**Delhaize et al., 2013; Rhee et al., 2013, 2016). The $\Omega_{\rm HI}$ measurements made in these studies are indicated by the round data points in **??**. The results from these studies are consistent with previous measurements of $\Omega_{\rm HI}$ made from direct detections or intensity mapping. However, as is evident in **??**, the error bars on the measurements are still too large to constrain the evolution of $\Omega_{\rm HI}$.

CHILES have presented some early stacking results from the pilot study (Fernández et al., 2013), they found an average HI mass of $(1.8 \pm 0.3) \times 10^9\,{\rm M}_\odot$ for a sample of 80 galaxies located at redshifts in the range $0.12 < z < 0.13$. The BUDHIES survey (Jaffé et al., 2016) used HI stacking to study the gas properties of galaxies in cluster environments. They found that the galaxies located in the cluster core are more gas-poor than those located on the edge of the cluster region.

Other results include the work by Fabello et al. (2011a) and Brown et al. (2015) on exploring the relationship between HI mass and stellar mass (HI to stellar mass ratio or gas fraction). Fabello et al. (2011a) stacked a sample of ~ 5000 galaxies that had both GALEX and SDSS imaging with ${\rm M}_\star > 10^{10}\,{\rm M}_\odot$. By separating the sample into early-type and late-type galaxies, they found that the gas fraction correlates better with NUV $- r$ colour – this result was confirmed by Brown et al. (2015) who used a larger sample with ${\rm M}_\star > 10^9\,{\rm M}_\odot$.

Initially only the non-detected HI spectra were used in stacking analyses. More recently, with the abundance of spectra available through surveys like ALFALFA, stacking analyses have included both the detected and non-detected spectra. If available, including the detected spectra in the analysis provides a more complete average of the sample; it should be noted that in a sample of both detected and non-detected HI spectra, the detected spectra make up a small percentage of the total number of spectra. In the sample of galaxies selected from both GALEX and ALFALFA, Brown et al. (2015) find fewer than 20% of the HI spectra are classified as detections. Figure 1.7 shows the comparison between two stacked spectra. The left panel shows the stacked profile created using both HI detections and non-detections, while the panel shows the stacked spectrum created using only HI non-detections. This comparison highlights the fact that stacking both detections and non-detections together is not dominated by the HI flux from the detections (Rhee

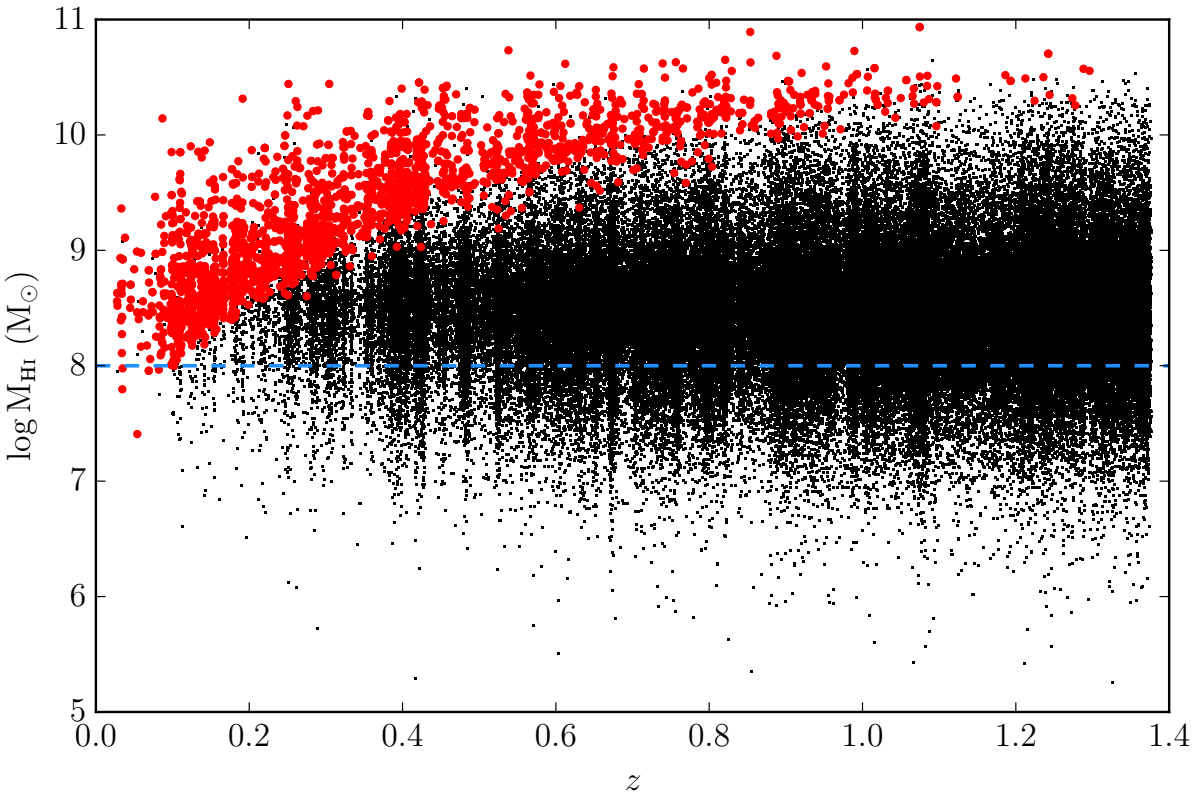


Figure 1.8: *The data used for this plot is taken from the S^3-SAX catalogue (Obreschkow et al., 2009a,b,c), the simulations used to create the catalogue are complete for* HI *mass greater than* $\log M_{HI} = 8$ *as indicated by the dashed blue line. The red data points highlight the galaxies for which there will be a 5σ detection, while the black points show the galaxies for which there will be no direct detection of* HI.

et al. (2013) found that the detected HI spectra of late-type galaxies contribute only 44% to the average spectrum, while intermediate-type HI detections contribute only 24% to the average stacked spectrum). It is clear that the non-detected spectra also contribute significantly to the stacked spectrum.

1.5 Probing HI Emission at Intermediate Redshift

Directly detecting HI with current radio telescopes beyond $z \sim 0.25$ is challenging due to the long observing times required. Next generation telescopes such as the Square Kilometre Array (SKA) and SKA Pathfinders (APERTIF in the Netherlands, MeerKAT in South Africa, and ASKAP in Australia) have been designed with greater fields of view than current telescopes, to enable them to observe more objects in much less time.

Looking At the Distant Universe with the MeerKAT Array (LADUMA; Holwerda et al. (2011)) is one of the approved Large Survey Projects on the upcoming MeerKAT telescope. It will be one of the deepest HI surveys ever performed, observing thousands of galaxies in a region of the sky that covers $\sim 1\,\mathrm{deg}^2$ at $z = 0$, increasing to $\sim 5.5\,\mathrm{deg}^2$ at $z = 1.4$ (Blyth, 2014). Even with the high sensitivity of MeerKAT (rms noise of $1.6\,\mu$Jy per channel at a spectral resolution of $7\,\mathrm{km \cdot s^{-1}}$), LADUMA does not anticipate directly detecting all galaxies in the survey volume. There are expected to be more

than 10 000 galaxies in the LADUMA survey volume, however it is anticipated that only ~ 3500 of the galaxies will have 5σ HI detections (Blyth et al., 2016). Figure 1.8 shows a simulation of the distribution of the HI mass of the galaxies as a function of redshift in the LADUMA survey volume. This means that there will be thousands of galaxies for which there is no direct HI detection. The LADUMA survey aims to study galaxy evolution by investigating:

- Cosmic Neutral Hydrogen Density ($\Omega_{\rm HI}$): one can obtain an estimate of the total HI density at various epochs and gain insight into the role of HI in galaxy evolution over cosmic time
- The HI Mass Function (HIMF): the HIMF represents the volume density of galaxies as a function of HI mass. The second moment (the HI atoms per unit volume or $\Omega_{\rm HI}$) is an important parameter in the understanding of the evolution of cosmic star formation rate density
- Galaxy HI content in different environments: the HI content of different galaxy populations as a function of redshift will provide insight into how galaxies evolve

At higher redshifts ($z \gtrsim 0.6$) direct detections of the HI emission from galaxies will be scarce in the LADUMA dataset. Thus, the LADUMA team plan to use HI stacking to analyse the high redshift (Holwerda et al., 2011). Section 1.4 discusses some of the HI stacking analyses that have been conducted in the last two decades; these studies have been limited to low redshifts ($z \lesssim 0.25$) due to the constraints of the available data. Due to the large redshift range spanned by LADUMA, HI stacking analyses (such as the analysis in this work – see Chapter 4) that have been confined to the local universe can be extended to higher redshift regions. The comparison of well studied low redshift HI relations (e.g. HIMF, $\Omega_{\rm HI}$, gas scaling relations) to the counterparts at the high redshifts probed by LADUMA may provide more clues as whether there has been any evolution in the gas content of galaxies over the last ~ 7 Gyrs.

The published stacking studies in the last fifteen years have highlighted how powerful the stacking technique can be in extracting information from spectra that contain no obvious HI emission. The variety of the studies mentioned above have proven that stacking is a scientifically useful technique, however there needs to be more consistency in the stacking method if this technique is going to be a reliable tool for high redshift studies in the future. Large survey teams such as the LADUMA Large Survey Project with MeerKAT have already identified stacking as the tool of choice for high redshift ($z > 0.5$) studies which hastens the need to have a consistent or reference method to which all stacking results can be compared.

1.6 Aims

There are two aims to this thesis. The first aim is to design, develop and test an HI line stacking software package. This software package is to be created using the Python programming language so that it is free to use, easy to modify, and accessible to most operating systems. The LADUMA team has identified that there is a need to have an HI stacking package that can serve as the primary stacking tool so that the stacking analyses performed using LADUMA data may be consistent and the systematics easily comparable. The software package whose design and creation is detailed in this work is intended

to fill the stacking package needs of the LADUMA team.

The second aim of this thesis is to use the stacking software to analyse the detected and non-detected HI spectra for an isolated sample of galaxies from the Nançay Interstellar Baryon Legacy Survey (NIBLES; van Driel et al. (2016)). The NIBLES Survey is an optically-selected targeted HI survey of $\sim$ 2500 galaxies in the nearby universe ($z < 0.04$). The NIBLES targets were selected to uniformly fill bins of stellar mass in the range $6 < \log \mathrm{M}_\star\ (\mathrm{M}_\odot) < 12$. In our study of the gas scaling relations, we are able to probe the gas scaling relations down to a stellar mass more than an order of magnitude lower than the mass range studied by Brown et al. (2015).

The structure of the dissertation is thus: in Chapter 2, the outline of the stacking methodology as well as the simulated HI data that was used during the testing phase is presented. Chapter 3 outlines the design specifications and strategy that went into creating the stacking software package which has been named HI Stacking Software (HISS). HISS is used to stack the NIBLES spectra in Chapter 4. Finally, Chapter 5 wraps up the discussion of the NIBLES results and presents the next steps for HISS.

Chapter 2

Stacking Methodology

This chapter discusses the stacking method which is implemented in the HI Stacking Software. Throughout this chapter there are graphic examples of how each step of the stacking method is implemented. The spectra used in the examples are simulated HI spectra based on evaluated galaxy properties from the S^3-SAX catalogue (Obreschkow et al., 2009a,b,c). Section 2.1 outlines the creation of the simulated data and Section 2.2 contains the details of the stacking method.

2.1 Simulated Data

When developing new data analysis tools, one of the most important steps is to gauge the accuracy and reliability of outputs. For this reason simulated data (or more specifically for this work, simulated HI line emission spectra) are used because the properties of the input data are well known. While simulated spectra may lack some characteristics of real spectra, simulated spectra are useful in allowing the quantification of the reliability and accuracy of any generated results.

The HI galaxy profile parameters are obtained from the S^3-SAX catalogue by Obreschkow et al. (2009a,b,c). The S^3-SAX is a simulated database of the neutral and molecular hydrogen in the universe. The semi-analytic models used to determine baryonic content for the galaxies rely on the Millennium Simulation for information about the cosmic structure. The physical models which assign the HI content to each galaxy are able to realistically reproduce the characteristic HI emission line. The S^3-SAX catalogue is hosted by the University of Oxford and can be found at `s-cubed.physics.ox.ac.uk`.

The simulated spectra are used in the later sections of this chapter for illustrative purposes as well as for testing purposes in Chapter 3. The simulated spectra used in this chapter were based on the evaluated properties extracted from the S^3-SAX catalogue for a LADUMA-like field (0.9 square degrees at $z = 0$ up to 5.4 square degrees at $z = 1.4$ spanning $0 < z < 1.4$). The following parameters are extracted from the S^3-SAX catalogue:

- HI mass ($M_\odot$)
- Integrated flux density ($Jy \cdot km \cdot s^{-1}$)
- Observed redshift

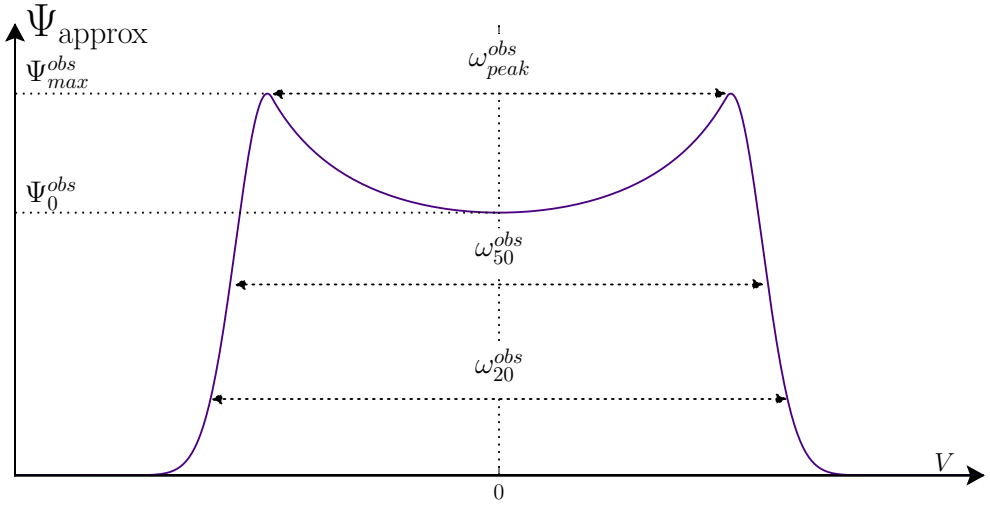


Figure 2.1: *Schematic of the characteristic* HI *galaxy emission profile using the Obreschkow et al. (2009c, Appendix A) formulation which is outlined in Equation 2.1.*

- The 5 spectrum parameters which describe the characteristic HI double-horned profile:
  - Ψ^{obs}_{max}: the peak intensity of the spectrum
  - Ψ^{obs}_{0}: the value of the central depression of the HI double-horned profile
  - ω^{obs}_{peak}: velocity width between the two peaks
  - ω^{obs}_{50}: the velocity width at 50% the peak flux
  - ω^{obs}_{20}: the velocity width at 20% the peak flux

The five spectrum parameters Ψ^{obs}_{max}, Ψ^{obs}_{0}, ω^{obs}_{peak}, ω^{obs}_{50}, and ω^{obs}_{50} are illustrated in Figure 2.1.

The following prescription of how to create HI line profiles from the S^3-SAX catalogue is taken from Obreschkow et al. (2009a, Appendix A). An HI profile such as the one plotted in Figure 2.1 is defined by:

$$\Psi_{\rm approx}(V) = \begin{cases} k_3 \exp\left[-\frac{(|V|-k_1)^2}{k_2}\right] & \text{if } |V| \geq \frac{1}{2}\omega^{\rm obs}_{\rm peak} \\ \frac{k_5}{\sqrt{k_4 - V^2}} & \text{if } |V| < \frac{1}{2}\omega^{\rm obs}_{\rm peak} \end{cases} \tag{2.1}$$

where

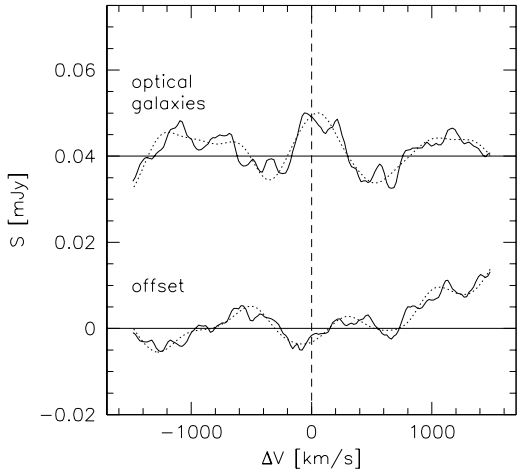
$$k_1 = \frac{\ln(0.5)\left[\omega^{\rm obs\,2}_{\rm peak} - \omega^{\rm obs\,2}_{20}\right] + \ln(0.2)\left[\omega^{\rm obs\,2}_{50} - \omega^{\rm obs\,2}_{\rm peak}\right]}{\ln(0.5)\left[\omega^{\rm obs}_{\rm peak} - \omega^{\rm obs}_{20}\right] + \ln(0.2)\left[\omega^{\rm obs}_{50} - \omega^{\rm obs}_{\rm peak}\right]}$$

$$k_2 = \frac{0.25\left[\omega^{\rm obs\,2}_{\rm peak} - \omega^{\rm obs\,2}_{50}\right] + k_1\left[\omega^{\rm obs}_{50} - \omega^{\rm obs}_{\rm peak}\right]}{\ln(0.5)}$$

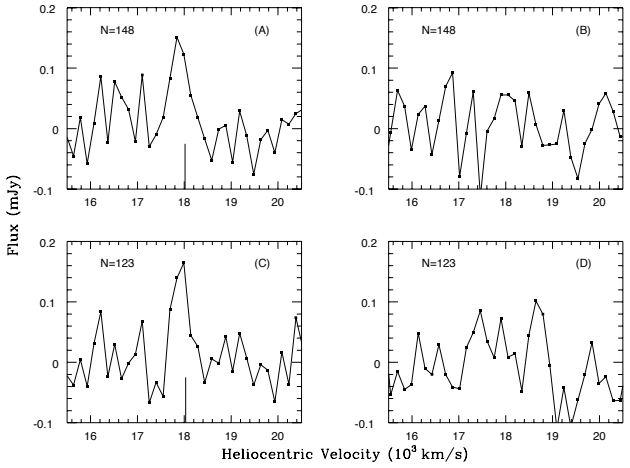
$$k_3 = \Psi^{\rm obs}_{\rm max} \exp\left[\frac{(2k_1 - \omega^{\rm obs}_{\rm peak})^2}{4k_2}\right]$$

$$k_4 = \frac{1}{4}\frac{\omega^{\rm obs\,2}_{\rm peak}\Psi^{\rm obs\,2}_{\rm max}}{\Psi^{\rm obs\,2}_{\rm max} - \Psi^{\rm obs\,2}_{0}}$$

$$k_5 = \Psi^{\rm obs}_{0}\sqrt{k_4}$$



(a) *Detected stacked spectrum of spectra from the central $4' \times 4'$ of Abell 2218. The top panel shows the result of stacking 45 spectra. The bottom panel shows the reference spectrum which is created from stacking spectra that are extracted $40''$ north of every target. Both spectra are smoothed using a Gaussian filter with fwhm of $300 km \cdot s^{-1}$ which is indicated by the dashed lines. (Zwaan et al., 2001, Figure 8.4)*



(b) *Panels A and B correspond to the sample of 148 galaxies taken from the ESO Nearby Abell Cluster Survey. Panels C and D correspond 123 galaxies located outside of the cluster core. All four spectra are smoothed to a velocity resolution $140\,km \cdot s^{-1}$. The stacked spectra for the two samples are shown in Panels A and C, while Panels B and D show the reference spectra which are created by randomizing the redshift values for each sample's spectra. (Chengalur et al., 2001, Figure 4)*

Figure 2.2: *The first stacked spectra from Zwaan et al. (2001) (Fig. 2.2a) and Chengalur et al. (2001) (Fig. 2.2b).*

This prescription produces a spectrum with an integrated flux density of 1. In order to obtain the physically meaningful flux scale, the newly created spectrum must be scaled by the integrated flux density value (given in $Jy \cdot km \cdot s^{-1}$) which is available in the S^3-SAX catalogue.

2.2 The Stacking Method

The idea of co-adding the non-detected HI spectra in studies of the gas content in galaxies was first published by Zwaan et al. (2001) and Chengalur et al. (2001). Both groups were looking at the HI in galaxies in and around clusters. With low detection counts in their samples (Zwaan et al. (2001) had one detection at $z = 0.1766$ and Chengalur et al. (2001) had two detections which had no corresponding optical redshifts), both groups independently co-added the non-detections in an effort to obtain a statistically meaningful detection for their samples. Figure 2.2 show the stacked spectra from the two studies, Figure 2.2a shows the stacked spectrum created by Zwaan et al. (2001) and Figure 2.2b shows the stacked spectrum from Chengalur et al. (2001). Below is an outline of the stacking method implemented by Zwaan et al. (2001) and Chengalur et al. (2001), this method is simple in principle:

1. Use a catalogue of galaxy (RA, Dec, z) positions to create a sample of the associated HI spectra.

2. Convert each spectrum to the galaxy rest-frame and centre each spectrum.

3. Weight the spectra.

4. Co-add the spectra.

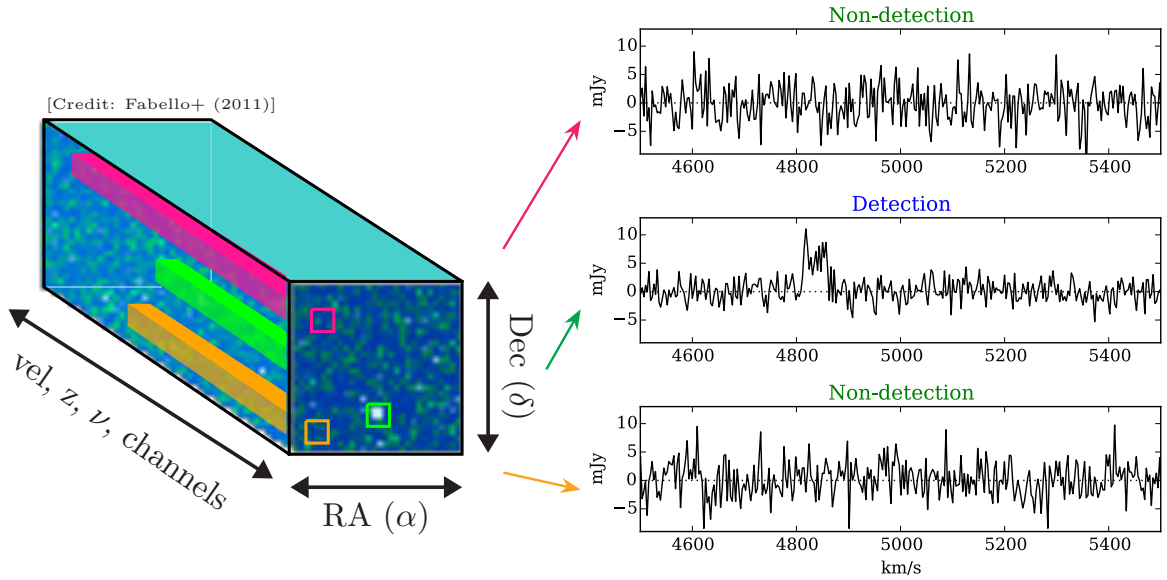


Figure 2.3: *Illustration of how* HI *line spectra are extracted from a radio data cube. The optical RA, Dec and z positions provide the coordinates around which a sub-volume may be centred. The sub-volume is extracted from the data cube and collapsed to create an* HI *spectrum.*

In the sections that follow, each of the above mentioned steps will be discussed in detail.

2.2.1 The Catalogue of HI Spectra

There are two different starting points for stacking, and depend on what the source of the HI spectra are. If the HI spectral data are contained within a radio data cube such as the one in Figure 2.3, a catalogue containing the RA, Dec, and z positions of the galaxies is required so that the HI spectrum associated with each galaxy may be extracted. The RA, Dec, and z positions are used find the centre of the galaxy, a suitable sub-volume containing the galaxy flux is defined around these coordinates. The size of the sub-volume may differ for each galaxy depending on the velocity range spanned by the HI emission as well as its spatial extent. The sub-volume is then collapsed into a spectrum. HI spectra that are obtained from targeted observations are usually already centred on the target galaxy's spatial location, however the spectrum may not be centred spectrally on the target redshift. In this case, the galaxy catalogue need only contain the redshifts associated with each galaxy.

Regardless of how the HI spectra were created, this software package described later in this work requires that the spectra are in a text file type format containing a column for the spectral axis and one for the flux density. The second use of the galaxy catalogue is for the stacking process. The most crucial input to any HI stacking algorithm (aside from the spectra), is the redshift associated with each galaxy. Figure 2.4 shows a selection of HI line emission spectra that will be used to illustrate different points in this chapter. Depending on how the observations were made, the emission from a target source may not necessarily be located at the centre of the spectrum.

The stacking examples used in this chapter make use of 1000 simulated HI line profiles. These profiles have been created using the method outlined in Section 2.1. Figure 2.4

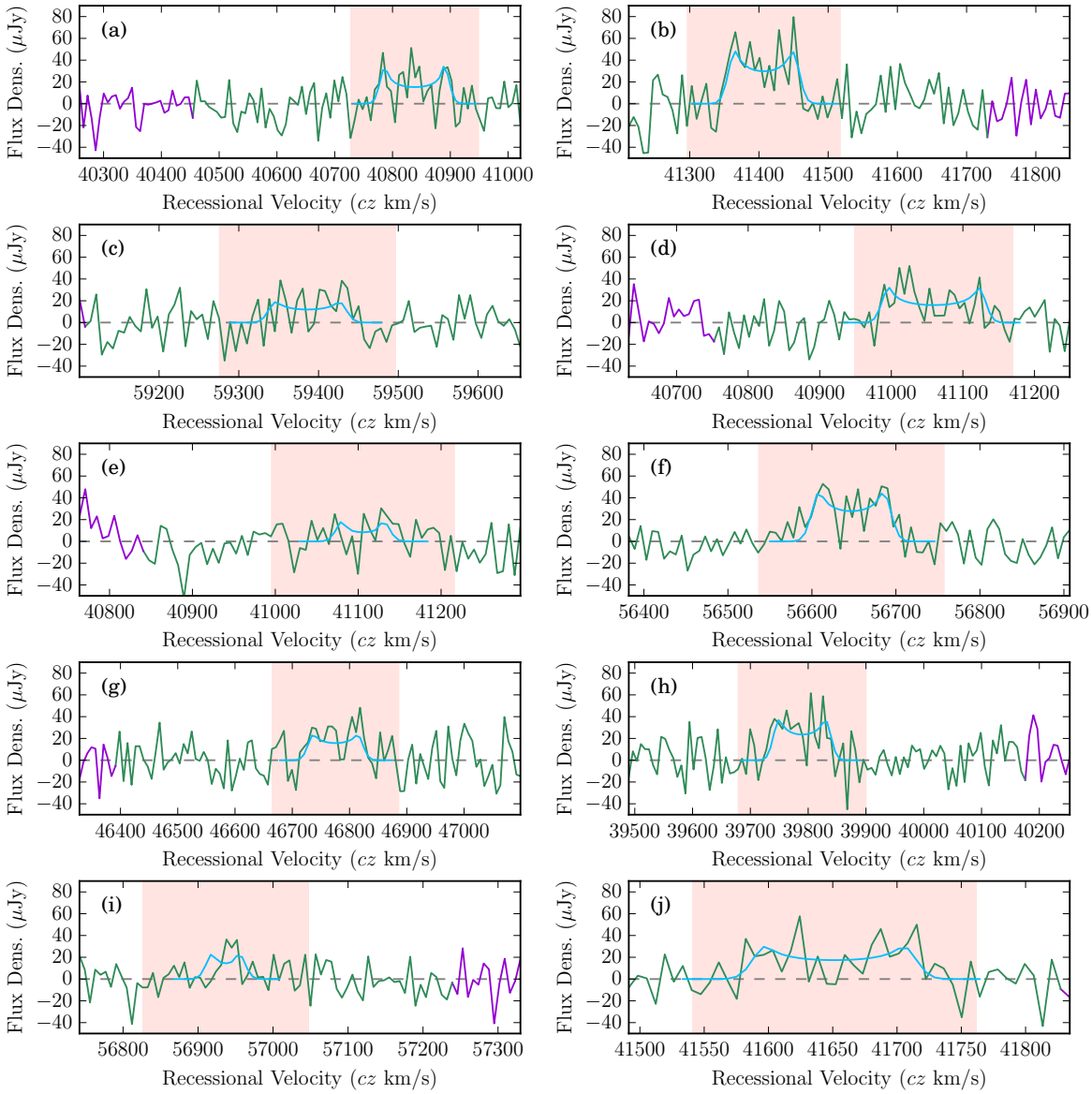


Figure 2.4: *These 10 spectra are a selection of the 1000 simulated spectra used in the stacking example throughout this chapter. The spectra are plotted using their native flux density and spectral axis. The simulated, noiseless spectra are shown in blue. In green and purple are shown the spectra after adding simulated Gaussian noise. The green section of the spectrum shows the part that will be shifted to the centre of the array while the purple section is the part that will be wrapped around and appended to the other end of the spectrum in the shifting process. The pink region shows the section of the profile where the galaxy emission is expected to be located. The width of this region is defined by the user and forms part of the galaxy mask. Panel (j) is an example of a spectrum that will be excluded during the stacking process as there are not enough channels outside of the galaxy window.*

shows a selection of 10 of the HI profiles. Each of the HI spectra in this sample have had 16 μJy Gaussian noise added, so as to simulate the spectra that are anticipated from the LADUMA survey. The galaxy spectra in this illustrative sample (particularly the ones that will be individually displayed) are predominantly clearly detected.

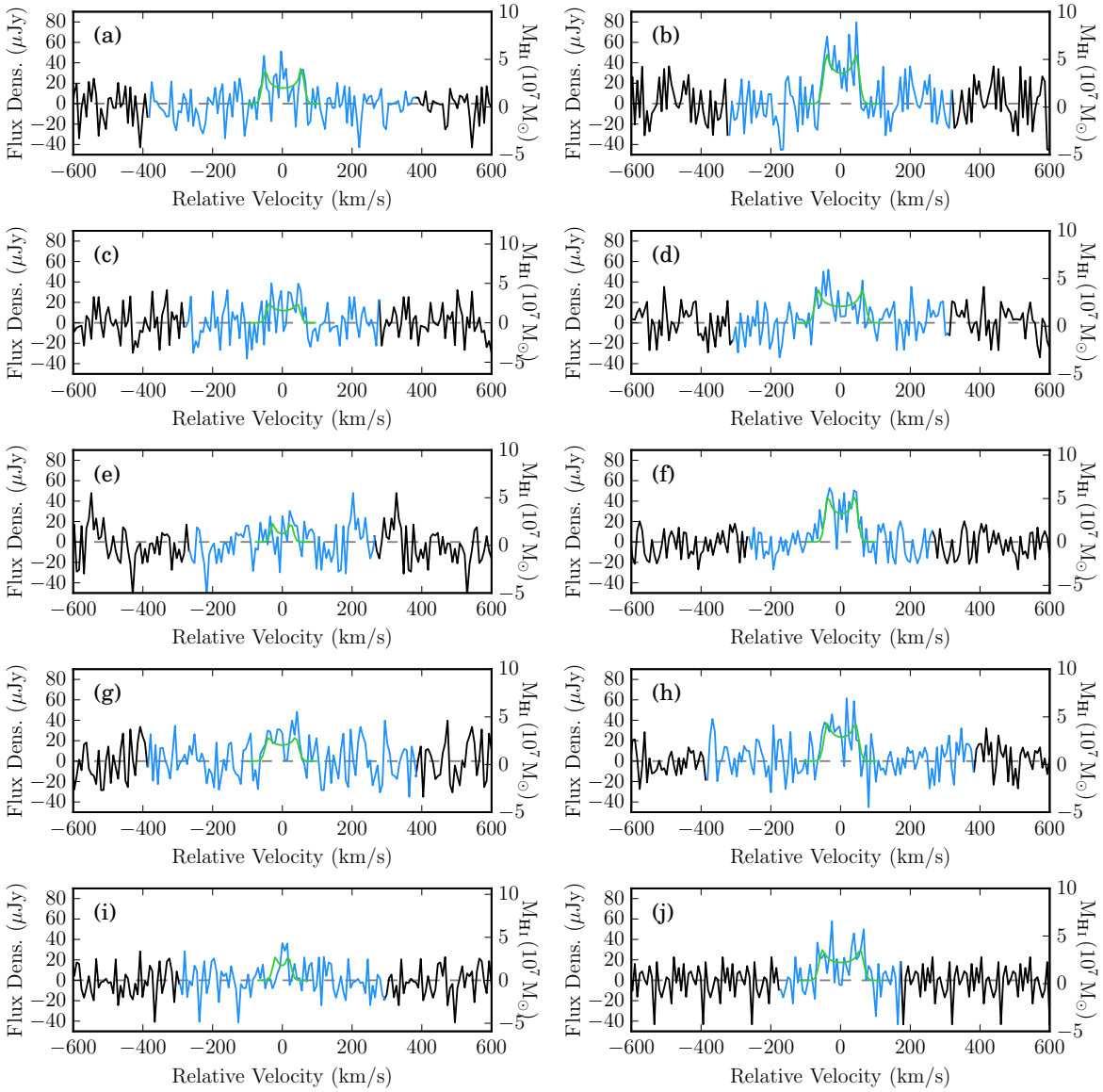


Figure 2.5: *These are the same 10 spectra as those plotted in Figure 2.4. Shown here in blue are the spectra that have been converted to their rest frame and centred. The original noise-free versions of each of the spectra are now shown in green. The black parts of the spectra are the parts of the extended spectrum that have been filled using the flux from outside the galaxy window. The right y-axis shows that each of these spectra have also been converted to mass spectra.*

2.2.2 Aligning the Spectra

As mentioned in Section 2.2.1, depending on how the HI spectrum was created, the centre channel of the spectrum may not correspond to the redshift associated with the target galaxy. The spectra may also all be of different lengths. In order to co-add the spectra to create a meaningful stacked spectrum, the input spectra must first be aligned so that galaxy emission for each spectrum is at the centre of the spectrum (see Figure 2.5 for the now-aligned version of Figure 2.4).

There are two steps to aligning the spectra. The first step is to convert the spectrum from the observed frame to the galaxy rest frame. The spectral axes are converted to the

rest frame by

$$v_{\rm emit} = v_{\rm obs} - cz \qquad \text{or} \qquad \nu_{\rm emit} = \nu_{\rm obs}(1+z) \tag{2.2}$$

where $v_{\rm obs}$ and $\nu_{\rm obs}$ are spectrum's spectral axis in either velocity or frequency respectively. The galaxy's redshift is given by z and the speed of light (c) is given in kilometres per second.

An input spectrum with a velocity spectral axis is assumed to have rest frame channel widths, this means that the conversion from the observed frame to the rest frame only requires a shift from the recessional velocity to $0\,\mathrm{km}\cdot\mathrm{s}^{-1}$ as given by the velocity relation in Equation 2.2. For a spectrum with a frequency spectral axis, the rest frame channel width is larger than the observed frame by $\mathrm{d}\nu_{\rm emit} = \mathrm{d}\nu_{\rm obs}(1+z)$; thus in order to convert the spectrum to the rest frame, both the spectral axis and flux must be appropriately scaled such that the integrated flux is conserved. The frequency axis is converted to the rest frame using the frequency relation in Equation 2.2, and the flux is scaled according to: $S_{\nu,rest} = S_{\nu,obs}/(1+z)$. Plotted in Figure 2.4 are 10 of the sample of 1000 simulated spectra used in this chapter.

The second step is to shift the galaxy emission to the centre of the spectrum. Since the spectra are now in the rest frame the target emission is located at either $0\,\mathrm{km}\cdot\mathrm{s}^{-1}$ or 1420 MHz (depending on the units of the spectral axis), and is easily shifted to the centre of the spectrum. In this step, some authors (Zwaan et al., 2001, etc.) will concatenate the spectra such that every channel has the same number of measurements (i.e. all of the spectra are the same length). Other authors (**?**Lah et al., 2009; Rhee et al., 2013, etc.) do not concatenate the spectra, instead choosing to weight each channel in the stacked spectrum relative to the number of measurements in that channel.

In this work, the flux that would then be shifted out of the now centred spectrum is wrapped around and appended to the other end of the spectrum so that the original spectrum length is maintained; this method also maintains the spectrum noise properties. This is illustrated in Figure 2.4: the green part of the spectrum is the main part that is shifted to the centre and purple highlights the section of the spectrum that is appended to the other side of the spectrum array. The centred spectra are shown in blue in Figure 2.5.

It is clear from each of the panels of input spectra in Figure 2.4 that every spectrum need not be the same length. Each spectrum is thus converted to some appropriate length. In order to avoid issues with small number statistics in the spectrum noise calculations, the length of the stacked spectrum should be at least three times the width of the widest galaxy (i.e. the largest W_{50}). The spectra that are longer than necessary are simply truncated at the edges while keeping the galaxy emission at the centre of the spectrum. The shorter spectra are extended by using the noisy channels from the outer channels to fill up the new empty channels at either edge of the spectrum. The noisy channels are those that are not expected to contain any galaxy emission. In order to determine which channels contain only noise, the galaxy emission is masked. The galaxy mask is centred at the spectral location of the galaxy, and is some pre-determined width (this is usually a conservative estimate on the maximum possible velocity width of galaxies in the sample). In Figure 2.4, for illustrative purposes, the galaxy emission is masked using the pink band, all other channels are considered to contain only noise, and so can be used in the spectrum extension process. There are some caveats associated with this process: if there are too few noisy channels in the spectrum then a ringing phenomenon may appear

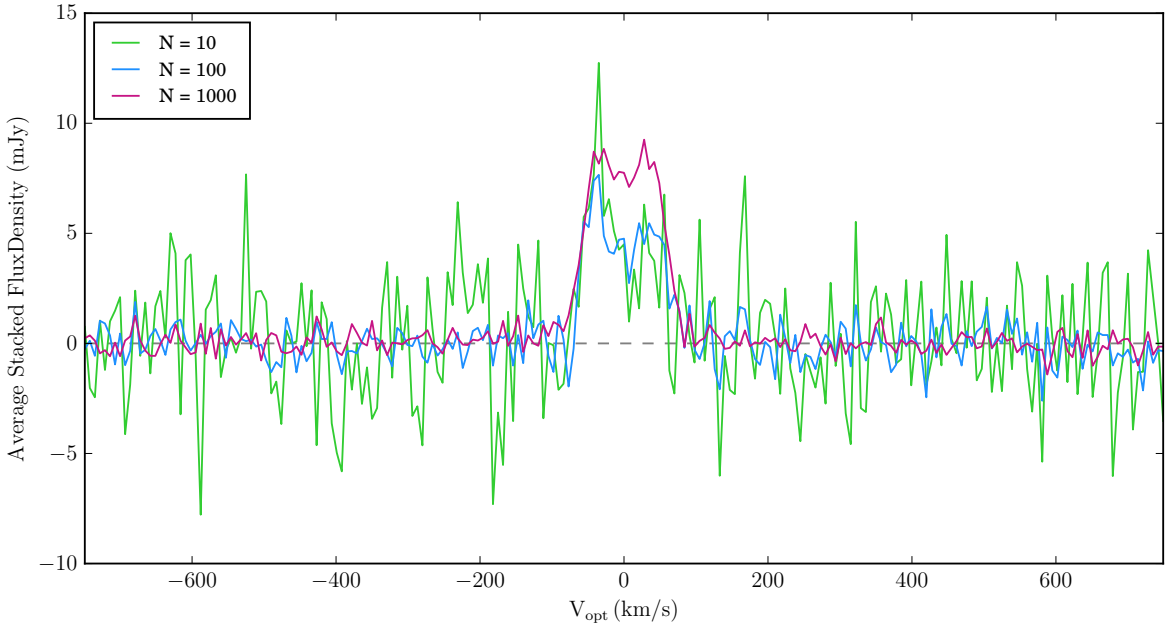


Figure 2.6: *This plot shows how the S/N of the average spectrum increases with increasing number of profiles. The average noise decreases as* $1/\sqrt{N}$.

in the extended spectrum. Figure 2.5 shows the original spectra centred in blue, while the black indicates the noise that has been used to fill up the channels of the extended spectrum. Panel j of Figure 2.5 highlights the ringing phenomenon that can occur with too few noisy channels in the original spectrum.

2.2.3 Weighting the Spectra

Most stacking analyses have used a weighted sum to co-add the spectra. The exact method of weighting spectra differs from study to study, however weighting is typically a variation of

$$w_i = \frac{1}{\sigma_i}, \tag{2.3}$$

where σ_i is the RMS noise in a particular channel (**?**Lah et al., 2009; Rhee et al., 2013, 2016) or σ_i is the RMS noise for that particular spectrum (Chengalur et al., 2001; Fabello et al., 2011a; Delhaize et al., 2013; Brown et al., 2015). Delhaize et al. (2013) explore the effectiveness of using $w_i = (\sigma_i D_{L,i}^2)^{-1}$, but found that while this weighting scheme creates a better S/N in the stacked spectrum, it caused a decrease in the effective volume probed by their survey and increased the cosmic variance.

2.2.4 Co-adding the Spectra

The final step required to produce a stacked spectrum is to co-add the spectra. This is typically done as a weighted average. The benefit of using the average is that Gaussian noise decreases proportional to $1/\sqrt{N}$ (N is the number of profiles included in the stacked spectrum), which is favourable to the creation of a higher S/N spectrum. Figure 2.6 show how the average noise decreases as the number of profiles included in the stack increase.

The weighted average is defined by,

$$S_{\rm stack} = \sum_{i=1}^{N} \frac{w_i S_i}{\sum_{i=1}^{N} w_i} \tag{2.4}$$

where i is the number of the spectrum to be included in the stack and N is the total number of spectra in the sample. Each spectrum S_i has an associated weighting factor w_i.

2.3 Quantities Derived from Stacked Spectra

HI stacking is most commonly used to obtain quantities such as the average HI mass ($\langle \mathrm{M_{HI}} \rangle$) or the average gas fraction ($\langle f_{\rm HI} \rangle$) for a sample of primarily HI non-detected galaxies.

2.3.1 HI Mass

The HI mass of a galaxy can be determined from an HI line profile using the following relation (Wieringa et al., 1992):

$$\frac{\mathrm{M_{HI}}}{\mathrm{M_\odot}} = \frac{2.36 \times 10^5}{1+z} \left(\frac{D_L(z)}{\mathrm{Mpc}} \right)^2 \left(\frac{\int S_v \mathrm{d}v}{\mathrm{Jy \cdot km \cdot s^{-1}}} \right) \tag{2.5}$$

where z is the galaxy redshift and $D_L(z)$ the associated luminosity distance. $\int S_v \mathrm{d}v$ is the galaxy rest frame integrated line flux density for the profile in $\mathrm{Jy \cdot km \cdot s^{-1}}$. This relation between the flux density and the HI mass assumes a spherical HI cloud that is optically thin with a uniform internal velocity distribution.

Early stacking studies of galaxies in clusters (Zwaan et al., 2001; Chengalur et al., 2001; Lah et al., 2009, etc.) would stack the sample of HI spectra in units of flux density. $\langle \mathrm{M_{HI}} \rangle$ was then calculated using Equation 2.5 and $z = \langle z \rangle$ to convert the integrated flux density from the stacked spectrum from $\mathrm{Jy \cdot km \cdot s^{-1}}$ to $\mathrm{M_\odot}$.

Fabello et al. (2011a) showed for their sample of field galaxies that unless the redshift bins were small enough such that

$$D_L^2 \left(\langle z \rangle \right) \simeq \langle D_L(z) \rangle^2 \simeq \langle D_L^2(z) \rangle,$$

where $\langle z \rangle$ is the mean redshift for the sample, it was better to stack the spectra using

$$S_v[\mathrm{Jy \cdot Mpc^2}] = S_i \cdot D_L^2(z_i) \tag{2.6}$$

which is the distance dependent part of Equation 2.5.

The spectra in Figure 2.5 can be stacked using either units of flux density (Jy) or in mass units ($\mathrm{M_\odot}$) as indicated by the right y-axis on each panel. The average stacked spectrum has units of $\mathrm{M_{HI}}$ per channel, thus a measure of the average HI mass ($\langle \mathrm{M_{HI}} \rangle$) can be obtained by integrating the flux over a chosen channel range.

2.3.2 HI Mass to Stellar Mass Ratio (Gas Fraction)

Studies such as Fabello et al. (2011a,b, 2012) and Brown et al. (2015) focus on the properties HI gas fraction as a function various observables. The HI gas fraction (or more accurately, the HI to stellar mass ratio) is simply,

$$f_{\rm HI} = \frac{\rm M_{HI}[M_\odot]}{\rm M_\star[M_\odot]}, \tag{2.7}$$

where $\rm M_\star$ is the stellar mass and $\rm M_{HI}$ is the HI mass. The studies by Fabello et al. and Brown et al. (2015) use a modified version of Equation 2.6 which contains an extra factor of $\rm M_\odot^{-1}$ arising from the stellar mass according to Equation 2.7. The modification accounts for the different contributions of stellar mass to the stacked spectrum:

$$S_v[\rm Jy \cdot Mpc^2 \cdot {M_\odot}^{-1}] = \frac{S_i \cdot D_L^2(z_i)}{\rm M_\star}. \tag{2.8}$$

2.4 Uncertainty in Stacking

The success of stacking studies depends on the quality of the ancillary multi-wavelength data that accompany the HI spectral data. Stacking requires precise and accurate position and redshift data for the galaxies of interest. The uncertainty in the redshift values can be a contributor to the uncertainty on quantities extracted from the stacked spectrum (e.g. Equation 2.5 shows the dependence of the calculated $\rm M_{HI}$ on the redshift due to the $D_L^2(z)$ factor). Redshift uncertainties can also have an affect on the overall shape of the stacked spectrum.

2.4.1 Redshift Uncertainty

Maddox et al. (2013) show using simulated noise-free HI galaxy profiles that the success of stacking relies on accurate redshift measurements, assuming that the optical redshift is essentially equivalent to the HI redshift (Maddox et al. note that this assumption may break down in dense environments or if galaxies are under going mergers). Stacking requires the galaxy redshifts to align and convert each spectrum to its rest frame. This means that if the redshifts are not accurate, the stacked spectrum will be smeared out as the individual profiles do not line up optimally. For studies that have accurate redshift positions for the sample of galaxies, stacking is an immensely powerful tool as it is able to recover HI masses that can be up to an order of magnitude or more lower than the 5σ detection limit (Chengalur et al., 2001).

In their study, Maddox et al. (2013) show that only when the redshift uncertainties become larger than the typical width ($\rm W_{50}$) of a galaxy in the sample, these uncertainties can start to affect the width of the stacked spectrum. The smearing affect due to large uncertainties starts to happen when the redshift uncertainties are around $\sim 150\,\rm km \cdot s^{-1}$. The typical error on spectroscopic redshifts from the Sloan Digital Sky Survey is $\sim 70\,\rm km \cdot s^{-1}$.

In this work, a physically motivated method of determining the error on stacked quantities is explored. Given the dependence of calculation of $\rm M_{HI}$ on z, this method uses a Monte Carlo-like approach to sample the galaxy redshift from a normal distribution given by $N(\mu = z; \sigma = u(z))$. Figure 2.7 shows, using the simulated HI spectra, how the uncertainties are calculated.

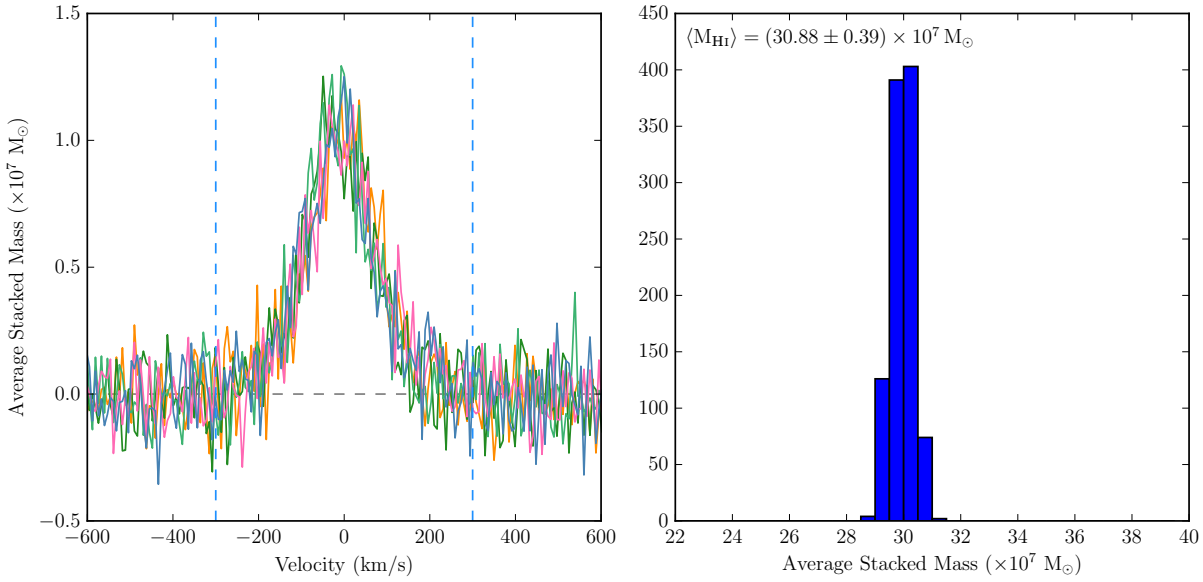


Figure 2.7: *The left panel shows five iterations of the stacked spectrum. Each of the stacked spectra have been created by varying the redshifts of the individual spectra. The errorbars on the stacked spectrum are determined from the range of values in each channel after 1000 iterations. The right panel shows the $\langle M_{\rm HI} \rangle$ calculated from each iteration of the stacked spectrum by integrating the spectrum between the two vertical dashed lines.*

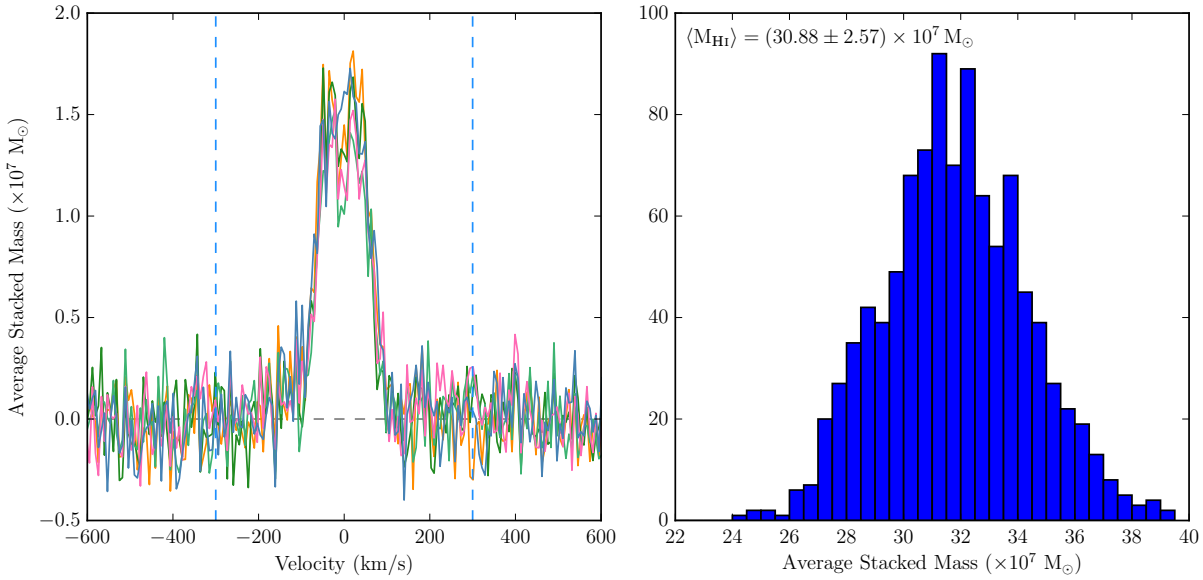


Figure 2.8: *The left panel shows five iterations of the stacked spectrum. The errorbars on the stacked spectrum are determined from the range of values in each channel after 1000 iterations. Each of the stacked spectra have been created using a sub-sample of 75% of the full sample. The right panel shows the $\langle M_{\rm HI} \rangle$ calculated from each iteration of the stacked spectrum by integrating the spectrum between the two vertical dashed lines.*

2.4.2 Error Calculation

There have been many ways to estimate the error on stacking results discussed in the literature. The few groups that explicitly state how they arrived at the error bars on stacked quantities (Chengalur et al., 2001; Fabello et al., 2011a, 2012; Rhee et al., 2013; Delhaize et al., 2013; Brown et al., 2015) each use a different method of determining the

uncertainty of their result. Re-sampling statistics are the most popular.

Fabello et al. (2011a) and Rhee et al. (2013) make use of the Jackknife method, while Brown et al. (2015) use a modified version called Delete-a-Group Jackknife (DAGJK) method. The Jackknife method is used to determine the variance and bias of a statistic by randomly sampling $N-1$ observations from a sample (where N is the size of the total population) to calculate the statistic of interest. The DAGJK method is similar to the usual Jackknife method, however instead of sampling $N-1$ observations one samples $n-k$ observations where k is representative of some percentage of the total number of observations (Brown et al. use $k = 20\%$ of N). The Jackknife methods provide a measure of the spread of the interesting quantity while assuming that the errors on the input values are negligible. These methods are useful for samples that potentially contain a diversity of spectra, or if the stacked spectrum is being dominated by a few individual spectra.

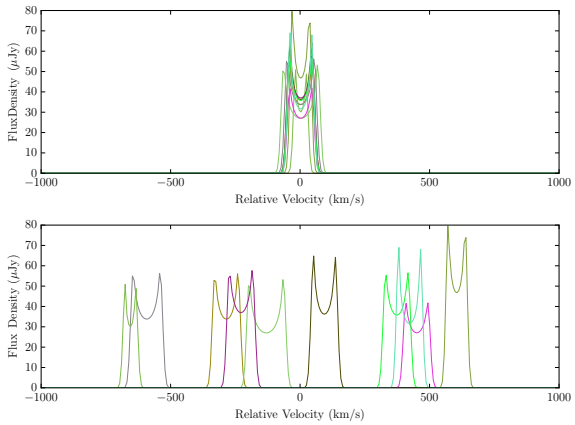
Figure 2.8 shows how the DAGJK method may be implemented to obtain a measure of the uncertainty on the average HI mass. In this example, 750 spectra (25% of the sample is removed for each iteration) are stacked every iteration. The left panel of Figure 2.8 shows how the stacked spectrum changes when different sub-samples containing 75% of sample are stacked. The average HI mass in this example was calculated from the mean of the histogram of the $\langle M_{HI} \rangle$ in the right panel of Figure 2.8, and the uncertainty from the width of the distribution.

2.5 Stacking Diagnostic Tools

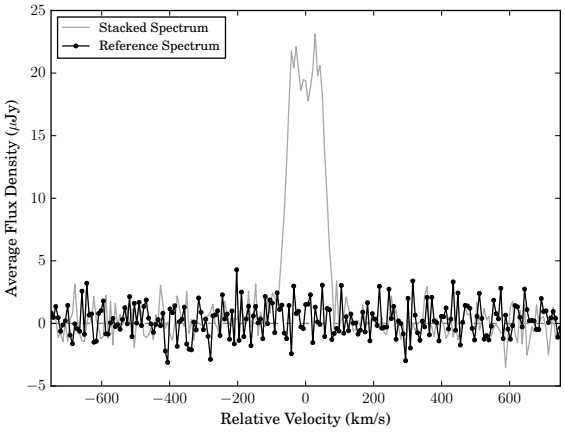
Since stacking typically involves the co-adding of non-detected spectra, identifying potential issues due to spectrum artefacts can be difficult. To aid in identifying problems during the stacking process, or rather to confirm the integrity of the stacked spectrum there are two diagnostic tools that are used in stacking analyses. The first tool is the reference or control spectrum, and the second is the stacked noise plot.

2.5.1 Reference/Control Spectrum

The practice of creating a reference spectrum has been used since the first HI stacking experiments. The stacked spectra in Figure 2.2 from Zwaan et al. (2001) and Chengalur et al. (2001) both include reference spectra. The various stacking studies have used two methods to create the reference spectrum. Some authors, who have had access to the radio data cube from which the HI spectra were extracted, choose to create the reference spectrum by extracting spectra from random locations in the data cube and using the target redshifts to align the spectra (e.g. Zwaan et al. (2001)). Other authors will use the spectra from their sample but randomize the sample redshifts (e.g. Chengalur et al. (2001); Delhaize et al. (2013)). By randomising the sample redshifts, both the reference spectrum and the stacked spectrum have the same systematics arising from the shifting process (Chengalur et al., 2001). If the redshifts are randomly assigned to the sample spectra, the resulting average spectrum should take the appearance of a noise spectrum. If this is the case, it can be further confirmation that a detection in a stacked spectrum is legitimate.

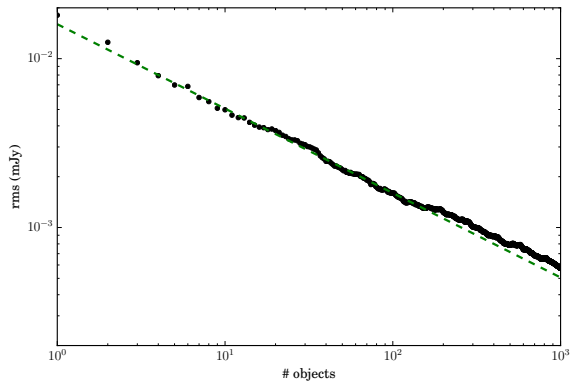


(a) *Top: the noise-free versions of the same 10 spectra that have been used previously. These spectra have been aligned using the correct redshifts. Bottom: the 10 spectra have been shifted using the redshift from another galaxy.*

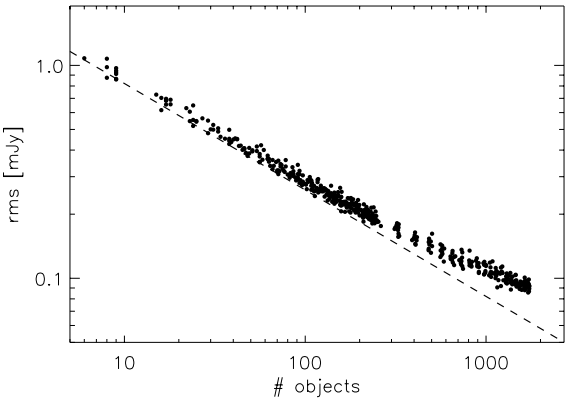


(b) *The reference spectrum for the 1000 noisy simulated spectra is plotted in black. The stacked spectrum for this sample is plotted in gray for comparison.*

Figure 2.9: *The reference/control spectrum*



(a) *Stacked noise for stacking experiment using 1000 simulated spectra. The noise added to each of the simulated spectra was Gaussian with a standard deviation of 16 μJy. The average noise in this stacking experiment decreases by 16 μJy/$\sqrt{N}$ (shown by the dashed green line) as expected.*



(b) *The average stacked noise as a function of the number of stacked profile. The dashed line shows the expected decreasing trend of purely Gaussian noise ($\sigma/\sqrt{N}$). It is clear from the plot that the non-Gaussian components of the noise start to dominate around $N \sim 300$, thus causing a change in the decrease of the average stacked noise. (Fabello et al., 2011a, Figure 5.)*

Figure 2.10: *Stacked noise for both real and simulated data*

Figure 2.9a shows how when the spectra are converted to the rest frame with the wrong redshift they end up being completely misaligned. Each of the spectra in Figure 2.9a have been shifted using the redshift belonging to one of the other galaxies. The result of co-adding the misaligned spectra using Equation 2.4 is shown in Figure 2.9. The lack of any significant emission at $0\,\mathrm{km \cdot s^{-1}}$ supports the integrity of the detection of the corresponding stacked spectrum that has been created using the correct redshifts.

2.5.2 Stacking Noise Analysis

The noise in an HI spectrum is expected to be Gaussian in nature which means that when the noise is co-added and averaged during the stacking process, it is expected to

decrease as $\sigma/\sqrt{N}$ where N is the number of stacked profiles and σ the average noise of the input spectra. This ideal relationship between the stacked noise and number of profiles is shown in Figure 2.10a, where pure Gaussian noise has been added to each of the simulated spectra included in the stack. The dashed line clearly shows the $\sigma/\sqrt{N}$ relationship where, in this case, $\sigma = 16\,\mu$Jy.

In reality the noise of the input spectra is not purely Gaussian as there can be systematic contributions from standing waves or poor baselines. Fabello et al. (2011a) found in their sample of HI spectra from the ALFALFA survey that around $N \sim 300$ the non-Gaussian noise starts to dominate. The noise will continue to decrease once the non-Gaussian noise dominates, but it does not follow the same $1/\sqrt{N}$ trend. This phenomenon is evident in the stacked noise plot from (Fabello et al., 2011a, Figure 5) which is shown in Figure 2.10b. The black dashed line shows the theoretical $\sigma/\sqrt{N}$ decline.

Chapter 3

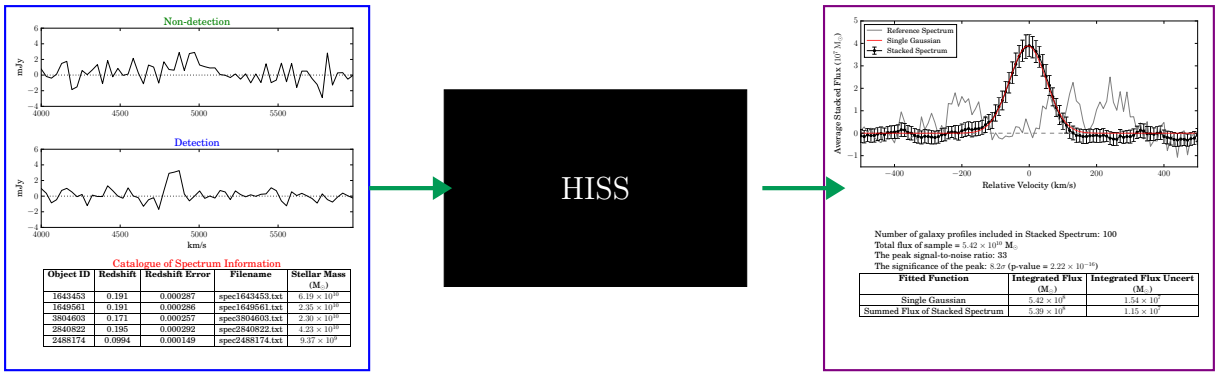
Software Design and Architecture

Since the upcoming HI surveys on the SKA Pathfinder Telescopes will be using stacking to analyse thousands of galaxy HI spectra to obtain average properties from many different galaxy sub-samples, the LADUMA team thought it would be beneficial to the community to develop a software package that can stack spectra in a reliable and consistent manner.

3.1 Design Motivation and Goals

The LADUMA Stacking Working Group (LSWG) require a stacking tool that is able to reliably handle the volumes of data that they anticipate receiving once the survey starts. This tool would serve as the reference for stacking studies using LADUMA data and possibly for the wider astronomy community. In order for results of various stacking studies to be reliably compared, consistent analysis methods should be used. With the goal of creating a stacking software package, the SWG created a list of objectives that such a package needs to achieve:

- be freely available and accessible to most operating systems
- to be open source
- easy to modify
- stack hundreds or thousands of galaxy spectra in an efficient and reliable manner
- be able to stack in different units (Jy, $M_\odot$, etc.) as well as handle different units for the input spectra (Hz or $km \cdot s^{-1}$ and Jy or mJy etc.)
- provide a summary of the values that can be calculated from a stacked spectrum, and save the calculations to file
- take as input to the stacking process a catalogue file that contains the redshift positions and other pertinent information (such as weighting factors and stellar masses) for the galaxy sample
- keep a record of the galaxy profiles included in the stack while checking each spectrum for data quality and discarding those spectra flagged as poor quality



Object ID	Redshift	Redshift Error	Filename	Stellar Mass ($M_\odot$)
1643453	0.191	0.000287	spec1643453.txt	6.19×10^{10}
1649561	0.191	0.000286	spec1649561.txt	2.35×10^{10}
3804603	0.171	0.000257	spec3804603.txt	2.30×10^{10}
2840822	0.195	0.000292	spec2840822.txt	4.23×10^{10}
2488174	0.0994	0.000149	spec2488174.txt	9.37×10^{9}

Figure 3.1: *The LADUMA Stacking Working Group wants a software package that can take in a selection of* HI *galaxy spectra along with a catalogue containing the galaxies' optical redshifts to produce a stacked spectrum with error-bars and a basic analysis.*

A cartoon of the software package is shown in Figure 3.1. The software needs to take in an optical galaxy catalogue and a selection of galaxy spectra, and produce a stacked spectrum from which average properties for the sample may be extracted.

The Python programming language is one of the most accessible programming languages as it can be used regardless of the operating system and also one of the most commonly used within the astronomy community. It was therefore chosen as the language to use in writing the HI Stacking Software (HISS). The Python modules AstroPy, NumPy, SciPy, and Matplotlib have been optimised in C for data input, manipulation, output and display which means that these modules can be used for HISS. AstroPy, in particular, has been designed for use by astronomers and includes useful methods such as unit handling, conversion and a comprehensive library of astronomical constants. HISS will be released to the astronomy community to use for data analysis and if desired, to make additions or changes as is tradition with open-source software packages. Taking a modular design approach adds a level of accessibility to future users allowing them to swap modules (that are more suited to their analysis end goals) in and out of the pipeline or modify modules without compromising the overall stacking process. Python is an object-oriented language which makes it ideally suited to the creation of a modular Stacking Package.

3.2 Stacking Package Design

A top-down design approach was used in creating HISS. This approach starts with what is known about the input and desired output before tackling the details of the black box connecting the input to the output. This method was suited to designing HISS as the type of input and the desired results have been established by the LSWG (see Section 3.1).

Based on the requirements for a stacking package set by the LSWG, there are four main processes that need to be incorporated into the design of the HISS. The processes are:

1. Input user information and catalogue (e.g. Section 2.2.1)
2. Stack spectra (e.g. Section 2.2.2, 2.2.3, and 2.2.4)

3. Analysis of stacked spectrum (e.g. Section 2.3, and 2.4)

4. Display results

The four processes mentioned in the list above dictate the flow of the software – how the package takes the provided HI spectra to produce a single stacked spectrum which contains information about the average HI properties for the galaxies within the sample. Python-based programs are a collection of modules which contain the classes which define the objects that are used throughout the program. Figure 3.2 contains a flow diagram of how the processes required to create a stacked spectrum may be incorporated into different modules.

The science requirements for the software package necessitated the addition of two optional modules: one that could handle creating sub-sets of the input catalogue, and another that could calculate the uncertainty on the quantities extracted from the stacked spectrum. How these two modules fit in to the flow of HISS is shown in Figure 3.2. The first of the new modules is called the Bin Module as it will bin the input galaxy catalogue according to some parameter, and the second new module is called the Uncertainty Module as it tells HISS how to calculate the uncertainties on the stacked quantities. The six modules that make up HISS are represented by the orange rectangles in Figure 3.2.

The flow diagram only gives a high level overview of how the software runs, and where user interaction is required in the process of creating the final product. Thus, in order to more clearly present the architecture of HISS, it is necessary to use a class diagram. The class diagram for HISS is shown in Figure 3.3. Each of the six modules contains at least one class which defines an object which can carry out the task assigned to that module, the modules are represented by the green rectangles with the top right corner folded. In Figure 3.3, the classes defined in each of the modules are represented by the blue split rectangles. The top segment of the class block gives the name of the class, the middle segment lists the most important class attributes (or properties), and the bottom segment lists the important methods (or functions).

Figure 3.3 highlights the how each of the modules depend of the each other. All of the classes excluding the `inputCatalogue` from the Input Module require an instance of the `inputCatalogue` as an input parameter. Upon initialization, the `anaData` class in the Analysis Module also requires an instance of the `objStack` class which is defined in the Stack Module. It is the `objStack` class which holds the stacked spectrum. The `dispData` and `uncertAnalysis` classes inherit the public functions from `anaData` class, inheritance is a useful tool in programming as it makes it possible to define a set of functions once that can then be used by other classes. The function of the Bin Module is to create a subset of the Input Module, so the `binData` class inherits the attributes from the `inputCatalogue`.

The modules work together under the directions from the pipeline module (which is represented by the yellow block in Figure 3.3). The pipeline module controls the overall running of the stacking package. When called, the pipeline will check the system on which it is about to run for the required software dependencies. If all the required dependencies are present, the `inputCatalogue` is initialised. Next it steps through the input catalogue file, initialising an instance of the `specObj` for each spectrum listed in the catalogue file. A `stackObj` is initialised to manage the co-adding of all the `specObj` objects. Once all

the entries in the input catalogue have been processed, the pipeline initialises the `anaData` to analyse the stacked spectrum.

HISS has been designed for a number of different types of users and system configurations. It offers different options for the first-time user and for more experienced users. This package has been designed to allow the user to choose different configurations (e.g. the plotting methods, by default, use the built in LaTeX distribution to typeset the labels, this mode can be turned off by the user). Running the uncertainty calculations of large datasets can be computationally expensive if run serially, therefore HISS offers the option of running the uncertainty calculations in parallel (HISS has been designed to offer both parallel and serial modes). These different options can be accessed using command-line arguments.
The following command is used to call HISS:

```
python pipeline.py [-f <config_filename>] [-s] [-m] [-l] [-d] [-p]
```

The `[-f]` option is for users who are experienced with their dataset and would rather not make use of the interactive nature of the software. HISS by default will display at each different step in the Analysis Module all plots to screen; to *suppress* this functionality when working on another computer via *ssh* protocol, use the `[-s]` command-line argument. The command-line argument `[-m]` enables *multiprocessing* for the uncertainty calculations, and `[-l]` will disable the use of LaTeX for the plots. Arguments `[-d]` and `[-p]` enable certain diagnostic features which are discussed in Section 3.2.3.

Sections 3.2.1 – 3.2.6 detail the design and layout of each of the modules that make up HISS. Each module contains one or more object classes that contain the base to which the different attributes (such as the spectrum, reference spectrum, noise spectrum in the case of the stacking object) are assigned. The different classes have different methods, some of which are private and are only to be used by that particular class and others are public and are inherited by other classes.

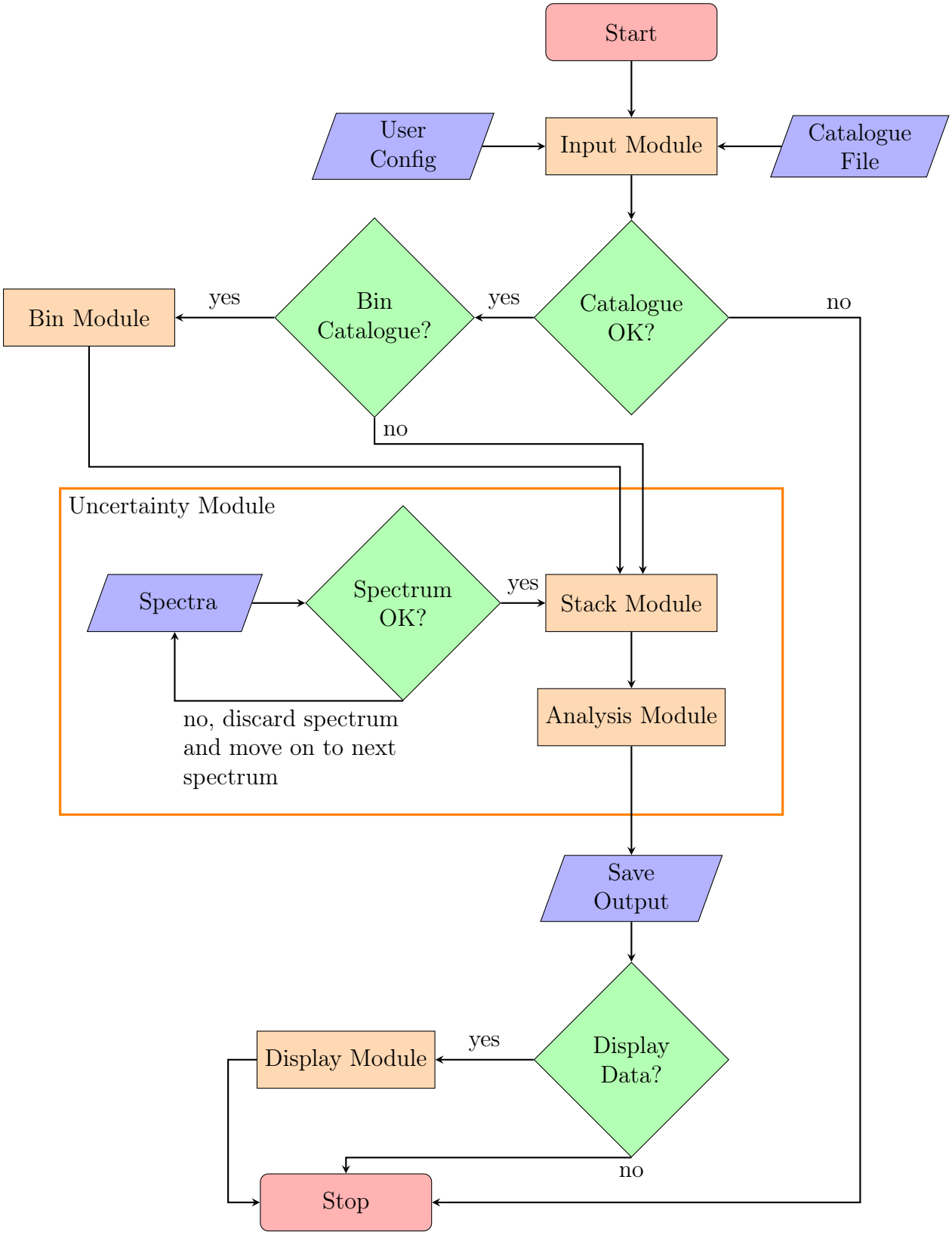


Figure 3.2: *This flow diagram shows how the individual spectra and user information are taken by HISS to produce a stacked spectrum from which average galaxy properties (such as $M_{\mathrm{H\,I}}$ and $f_{\mathrm{H\,I}}$) may be extracted. The orange rectangles show the six modules of the package, the blue parallelograms show the points of input or output, and the green diamonds show the checks that are critical to the running of the software.*

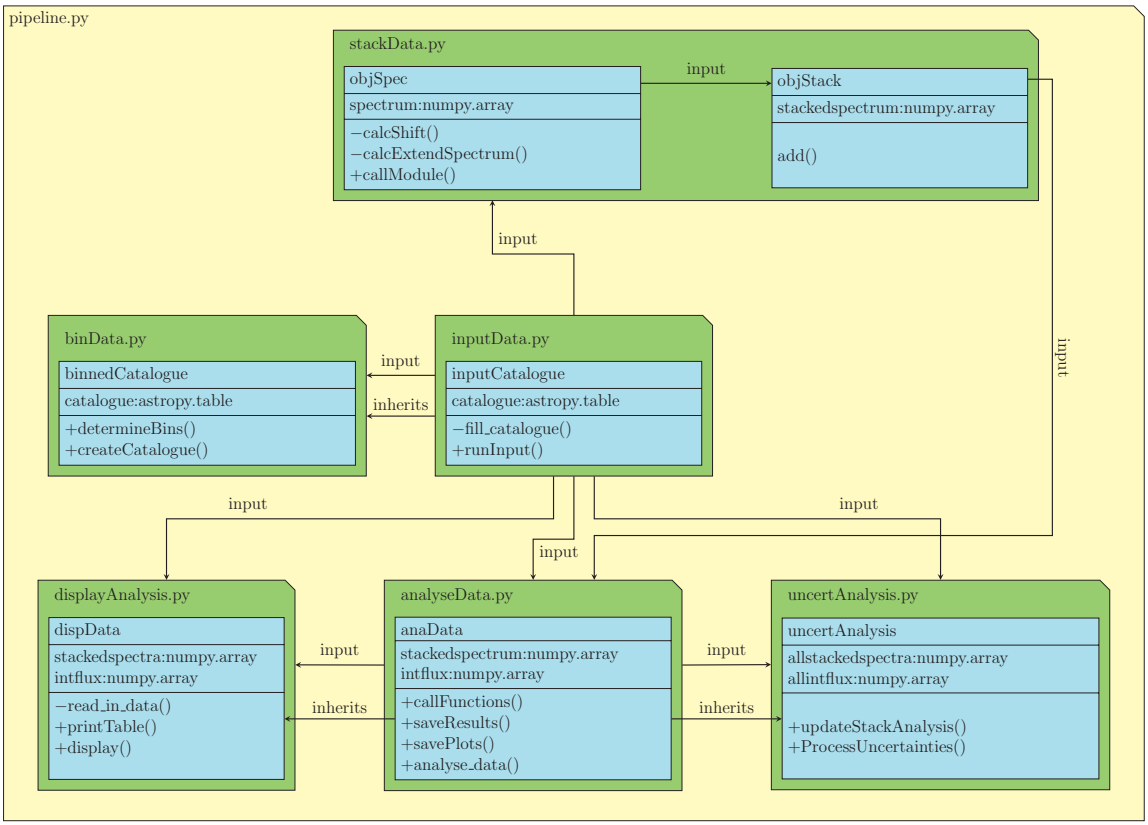


Figure 3.3: *The class diagram for HISS. The green rectangles with folded corners represent the different modules. The classes defined in each of modules are represented by the blue split rectangles. The top segment of the class block gives the name of the class, the middle segment lists the most important class attributes, and the bottom segment lists the important methods (functions) defined in the class (the "+" symbol represents the public methods, while the "−" represents the private methods).*

```
galaxyid,zapparent,ra,decl,himass,filename
6000036000696,0.255735,0.787977,-0.380374,227000000.0,spec_6000036000696.txt
3001932000010,0.277361,0.340264,0.898596,153000000.0,spec_3001932000010.txt
1002590000008,0.251487,-2.45585,-0.931149,63600000.0,spec_1002590000008.txt
2000497000049,0.31738,-1.99373,1.76517,248000000.0,spec_2000497000049.txt
2002476000011,0.301701,-0.487782,1.68676,404000000.0,spec_2002476000011.txt
2003119000010,0.300059,0.933986,-1.22683,387000000.0,spec_2003119000010.txt
5000047004034,0.316553,-1.29382,-1.56919,947000000.0,spec_5000047004034.txt
3001945000010,0.276697,-0.011291,1.27994,386000000.0,spec_3001945000010.txt
3003405000010,0.293927,-0.187541,-1.8823,236000000.0,spec_3003405000010.txt
1000061000269,0.342923,0.493996,1.47245,610000000.0,spec_1000061000269.txt
3000042006397,0.332297,-1.22729,1.08679,729000000.0,spec_3000042006397.txt
7002514000010,0.283451,2.05904,1.34245,135000000.0,spec_7002514000010.txt
```

Figure 3.4: *An example of the optical galaxy catalogue file. The file must be in CSV format as shown in this screen shot.*

3.2.1 Input Module

The Input Module (IM) contains one class: the `inputCatalogue`. The purpose of the `inputCatalogue` is to gather and store information from the user which will be needed during the stacking process. This is the object that does most of the interfacing with the user.

The most important input is the optical galaxy catalogue file. The catalogue file is a text file in CSV[1] formate that contains at least,

- Object ID
- Spectrum filename
- Redshift
- Redshift uncertainty
- Stellar Mass (optional)
- Other data (optional)

for each spectrum that the user would like to include in the stack. The optical galaxy catalogue file may contain any number of columns, however the columns mentioned above are the columns required to run HISS. An example of an optical galaxy catalogue file is shown in Figure 3.4. Other columns containing numerical information such as stellar mass or optical colour may be used to refine the catalogue into a number of different sub-samples (this process is discussed in Section 3.2.2 and applied in Section 4.1). The optical galaxy catalogue file is read in using the AstroPy ASCII module which loads the text file into an AstroPy Table.

The required user input can be entered using a configuration file which uses JSON[2] format to assign values to the required properties. The user can opt to use a configuration file in favour of entering values on the command-line when prompted by HISS. If the latter option is chosed, the entered data are written to a JSON file at the end of the stacking process. The properties for which the user is required to enter values are :

[1] comma separated values

[2] JSON stands for JavaScript Object Notation, a JSON formatted file has the extension `.json`

```
{
    "#": "json does not have comments, so the '#' entries are comment placeholders",
    "catfile": "/path/to/catalogue.txt",
    "catcolnum": [0,5,1,"",4,""],       "#": "column numbers for [Object ID, Filename, Redshift,
        Redshift Uncertainty, Stellar Mass, Other Data], Other Data is other data which is used to bin
        the sample - if one of the columns is not needed leave the column entry as an empty string",
    "z_max": 0.04,      "#": "maximum redshift value",
    "z_min": 0.002,         "#": "minimum redshift value",
    "fluxunit": 1,      "#": "1-Jy, 2-mJy, 3-uJy, 4-Jy/beam, 5-mJy/beam, 6-uJy/beam",
    "stackunit": 2,         "#": "1-Jy, 2-Msun",
    "spectralunit": 5,      "#": "1-Hz, 2-kHz, 3-MHz, 4-m/s, 5-km/s",
    "veltype": "optical",   "#": "[optical/radio] - Only change this if you know that the spectra use
        radio velocity",
    "chanwidth": 18.0,      "#": "in same units as spectralunit",
    "specloc": "/path/to/spectra/",     "#": "Location of spectra",
    "rowstart": 0,      "#": "first row of data in spectrum file (numbering starts at 1)",
    "rowdelim": "None",     "#": "spectra file data delimiter - if the delimiter is whitespace, use none
        ",
    "speccol": [0,1],       "#": "column numbers for [freq/vel, flux]",
    "cosmo": "y",       "#": "Change default cosmology [y/n]",
    "H0": 70,       "#": "Hubble constant (H0) in units of km/Mpc/s (default is 70 km/Mpc/s)",
    "Om0": 0.3,         "#": "Omega matter (Om0) (default is 0.3)",
    "changeoutloc": "y",        "#": "Change output location [y/n] default is ./output/",
    "outloc": "/Users/juliahealy/Documents/Masters/NIBLES/Stacking/PHSICC/Sample2",
    "chanlen": "y",         "#": "Change default length of stacked spectrum [y/n]",
    "galwid": 300,      "#": "default galaxy width is 450km/s, this option must use same units as
        spectralunit",
    "speclen": 1500,        "#": "default stacked spectrum length is 2000km/s, this option must use
        same units as 'spectralunit'",
    "weight": "y",      "#": "Use a weighting scheme [y/n]",
    "wopt": 1,      "#": "1-w=1.[default], 2-w=1/rms, 3-w=1/rms^2, 4-w=dl^2/rms^2",
    "entcat": "y",      "#": "stack entire catalogue [y/n]",
    "catstack": 50,         "#": "number of spectra to stack if entcat=y",
    "uncert": "n",      "#": "Perform uncertainty analysis, [y/n]",
    "binyn": "n",       "#": "Bin the catalogue sample, [y/n]",
    "bininfo": ["Stellar Mass", 1.0, 7.0, 12.0],        "#": "[Column name of data to bin, Bin width,
        Start of Bin range, End of Bin range]"
}
```

(a) *The configuration file allows the user to bypass the manually entering of the user data. Shown here is an excerpt of the configuration file.*

```
What units do you want your final stacked profile to have?
       1. Jansky (Jy)
       2. Solar Masses (Msun)
       3. Gas Fraction (Msun/Msun)
Please enter the number of the applicable units: 2
Should I calculate uncertainties? n
Initialising the catalogue file.
Enter location and filename of catalogue file: examplecatalogue.txt

The following columns are available:
0: galaxyid
1: zapparent
2: ra
3: decl
4: himass
5: filename
Please enter the column number of Object ID (if this column isn't necessary, please leave blank): 0
Please enter the column number of Filename (if this column isn't necessary, please leave blank): 5
Please enter the column number of Redshift (if this column isn't necessary, please leave blank): 1
Please enter the column number of Redshift Uncertainty (if this column isn't necessary, please leave blank):

Please enter the column number of Stellar Mass (if this column isn't necessary, please leave blank):

Please enter the column number of Other Data (if this column isn't necessary, please leave blank):


Please enter the maximum redshift value of your sample: 0.35
Please enter the minimum redshift value of your sample: 0.25

What are the flux units of your spectra?
       1. Jansky (Jy)
       2. Milli Jansky (mJy)
       3. Micro Jansky (uJy)
       4. Jansky/beam (Jy/beam)
       5. Milli Jansky/beam (mJy/beam)
       6. Micro Jansky/beam (uJy/beam)
Please enter the number of the applicable units: 1
```

(b) *A screen shot of the interactive questions asked by the input module should the user choose manually enter the required information.*

Figure 3.5: *Figures (a) and (b) show the two different ways the user can interface with the input module.*

- Location of the optical galaxy catalogue file.
- Column numbers for Object ID, Spectrum filename, Redshift, Redshift Uncertainty, Stellar Mass, Other Data (in this order).
- Minimum and maximum redshift values – this is used to make sure that all spectra included in the stack are in the desired redshift range.
- Flux density units of the input spectra (e.g. Jy or Jy/beam) as well as the spectral units (e.g $\mathrm{km \cdot s^{-1}}$ or MHz).

- Flux units in which to stack: flux density (Jy), HI mass ($M_\odot$), or gas fraction (f_{HI}). The units are stored as AstroPy Units.
- Column numbers and row start number for the spectrum flux and spectral axis data stored in each of the spectrum files (this must be the same for every spectrum).
- Location to which the stacked spectrum and other calculated data should be saved.
- Preferred H_0 and Ω_{matter} which is input to AstroPy's Lambda CDM cosmology (with $k = 0$) and used to determine the Luminosity Distance (D_L) to each galaxy.
- Maximum galaxy velocity width.
- Channel width.

An example of the interactive and configuration file methods of input are shown in Figure 3.5.

The maximum velocity width is used to determine how many channels are expected to contain HI emission; this then defines the size of the galaxy window. The galaxy window may be changed later on the in Analysis, however it is important that the initial size is conservative so that no HI emission is assumed to be noise. The minimum and maximum redshift values, as well as the channel width are used by the Stacking Module for the spectrum checks.

The `inputCatalogue` object created by initialising this module is passed into each of the following modules allowing them to use information entered by the user. This object contains the information that may be needed by all of the modules, and so each of the modules can update certain attributes in the `inputCatalogue`.

3.2.2 Bin Catalogue Module

Often, it is more scientifically interesting to break a large sample of galaxies into smaller sub-samples based on different parameters such as colour, stellar mass, etc. and stack the sub-samples. The Bin Module allows the user to input the large catalogue file into HISS and specify which column in the catalogue they want to use to sub-divide the data. The module uses the *group by* function from the AstroPy Table package to group the full catalogue into smaller sub-catalogues based on the user specified bin width and bin range. The Stacking Package will iterate through each of the sub-catalogues which eliminates the need for the user to create multiple catalogue files.

This module contains a class that is a child of the `inputCatalogue` class. If the user chooses to sub-divide the data, this class creates a copy of the input object with a modified catalogue table. The class keeps track of the original catalogue table, the sub-tables and which one has been stacked.

3.2.3 Stacking Module

The Stacking Module is the heart of the HISS. This module is divided into two classes: the first handles the individual spectra and prepares the spectra for stacking and the second

class controls how the individual spectra are added to create the stacked spectrum and associated information.

Spectrum Class

To initialise an object of the Spectrum Class (SC), an instance of the `inputCatalogue` (input) is required. Stored within the inputCatalogue is a catalogue table containing all the information from the catalogue file which the user was required to provide during the initialisation of the input object. Stored within the catalogue table are the object ID, filename and redshift of every galaxy profile to be included in the stacked spectrum. This class uses a series of methods to prepare the individual spectra so that they can be included in the stack.

The first method called by the class is that which reads the spectrum into memory. This function assumes that all the spectra to be read in have the same units as those previously specified by the user. The spectrum is then put through a series of quality checks:

- The spectrum's channel width is checked such that it is within 5% of the entered channel width. A warning is raised if the spectrum channel width is $5\% - 10\%$ different; however, if the channel width is more than 10% different the spectrum is discarded.
- The input object contains the minimum and maximum redshift values for the sample of galaxies. This check makes sure that the redshift associated with the spectrum falls within these bounds.
- If the user has decided to stack the spectra in units of gas fraction, the catalogue will be checked for an associated stellar mass value.
- The final check is for spectrum length. As mentioned in Section 3.2.1, a galaxy mask is created based on the expected maximum width of the galaxies within the sample. Every spectrum is checked such that it has more channels than the length of mask. An example of a spectrum that would fail such a check is shown in the panel (j) of Figure 2.4 and Figure 2.5.

For the spectra that do not pass one of the quality checks, a record of their object ID and reason for being excluded from the stack is added to a table that is written to file at the end of the stacking process. If the spectrum passes all the quality checks, it moves on to the next step which is to convert the spectrum to the galaxy rest frame using Equation 2.2. The spectrum flux is then shifted such that the galaxy redshift is aligned with the centre of the spectrum.

The next method to be applied to the spectrum is the unit conversion method. This method is called if the user has specified that they want to stack in HI mass units or in gas fraction ($f_{\rm HI}$) units. If neither of those two units is chosen, HISS will check that the spectrum is in units of flux density (Jy). For the conversion to units of HI mass and $f_{\rm HI}$, the spectrum is first converted to the HI mass using Equation 2.5; the conversion from $\rm M_{HI}$ to $f_{\rm HI}$ requires the spectrum to be scaled by the galaxy's stellar mass (Equation 2.7). The conversion from flux density to HI mass (Equation 2.5) requires that the spectrum

channel width be given in units of velocity ($\mathrm{km \cdot s^{-1}}$). If the spectrum spectral axis is in units of frequency, the channel width is converted from frequency to velocity by

$$\mathrm{dv} = \frac{\mathrm{d}\nu \cdot c(1+z)}{\nu_{\mathrm{emit}}}, \tag{3.1}$$

where dv is the channel width in $\mathrm{km \cdot s^{-1}}$, $\mathrm{d}\nu$ is the channel width in MHz, c is the speed of light in $\mathrm{km \cdot s^{-1}}$, and ν_{emit} is the rest-frame frequency of the emission line (for the 21 cm emission line, $\nu_{\mathrm{emit}} = 1420.405751\,\mathrm{MHz}$), and z is the redshift of the galaxy.

As discussed in Section 2.2.2, the input spectra need not be the same length, however they need to be converted to the same length before the spectra can be added together. The new spectrum length is specified by the user in the input phase. The new spectrum is called the extended spectrum. Converting the spectrum from its original length to the length of the extended spectrum is a two step process. The first step is to create a noise spectrum using all the spectrum channels that are not protected by the galaxy mask (in Figure 2.4, these are all the channels not in the pink region which denotes the galaxy mask). The shifted original spectrum is placed into the extended spectrum such that the centre of the two arrays are aligned. In some cases the original spectrum will be longer than the extended spectrum and so the edges will be truncated, but in other cases the original spectrum will the shorter of the two. If the original spectrum is shorter than the extended spectrum, the empty edge channels of the extended spectrum are filled using the values from the noise spectrum. In Figure 2.5, the original shifted spectra are shown in blue, however it is clear that all of the plotted spectra are shorter than the extended spectra. The black part of the extended spectra show the noise that is used to fill up the empty channels. Once again, the bottom right panel of Figure 2.5 shows why short spectra are not included in the stack – there are too few channels in the noise spectrum and so when extending the spectrum a kind of ringing effect appears.

The last piece of information calculated by the SC is the weighting factor (w_i). By default the weighting factor is 1, but the user may choose an alternate option such as

$$w_i = \frac{1}{\sigma^2_{\mathrm{rms}}} \tag{3.2}$$

$$\text{or } w_i = \frac{1}{\sigma_{\mathrm{rms}}} \tag{3.3}$$

$$\text{or } w_i = \frac{D^2_{L(z)}}{\sigma^2_{\mathrm{rms}}} \tag{3.4}$$

where σ_{rms} is the RMS of the channels in a spectrum that contain no galaxy emission (i.e. the noise channels) and $D_{L(z)}$ is the luminosity distance to the galaxy.

There are two options to allow the user to watch the build-up the stacked spectrum, this is a useful tool if any particular spectrum needs to be removed from the stack. The two progress options are:

1. Progress window: this displays Figure 3.6 to the screen. Figure 3.6 has four panels, the top panel on the left shows the original spectra as they are read in to memory, the middle left panel shows the spectrum centred and extended, the bottom left panel shows how the noise of the average spectrum is changing as a function of the

number of stacked profiles, and finally the right panel shows how the total spectrum is growing. The progress window can be activated using the `[-d]` when calling the HISS, this feature cannot be used in conjunction with `[-s]`.

2. Save the progress window: this option is the alternative to the Progress window in that it saves the progress window to a folder in the output location instead of displaying to screen. This option is called using `[-p]` and can be used in conjunction with `[-s]` or `[-d]`.

Stacked Spectrum Class

The Stacked Spectrum Class (SSC) tells HISS how to carry out the stacking algorithm which is defined by Equation 2.4. To carry out this process, the Stacked Spectrum Class requires four attributes from an instance of the Spectrum Class:

- extended spectrum
- noise spectrum
- reference spectrum
- weighting factor

The SSC applies Equation 2.4 to the extended, noise and reference spectra. As each spectrum is added to the stacked profile, the RMS of the stacked average noise spectrum is calculated; this is an important diagnostic tool as the spectrum noise is dominated by the instrument noise which is expected to be Gaussian in nature, so the noise of the average stacked spectrum is expected to decrease as $\frac{1}{\sqrt{N}}$ (N is the number of spectra included in the stack). The expected average stacked noise decrease is suggestive that the stacking is working properly, but problems may still occur.

As HISS steps through the supplied input catalogue, the SSC only performs the numerator of Equation 2.4 while separately keeping track of the denominator. Figure 2.6 shows how the total stacked spectrum (or the weighted sum of the sample of spectra) grows as a function of the number of profiles. Only once HISS has run through the entire input catalogue does it divide the weighted sum by the sum of the weights.

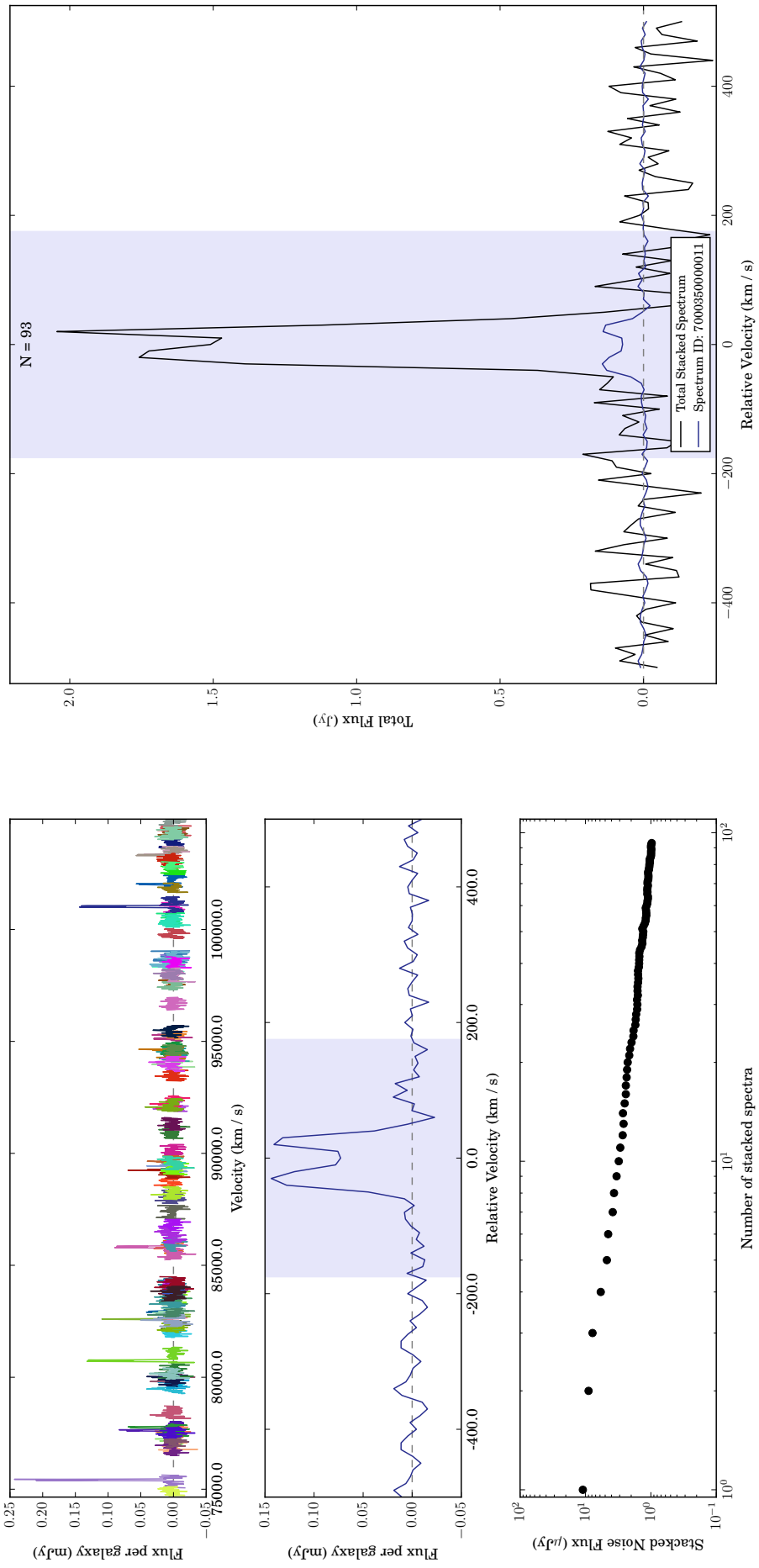


Figure 3.6: *The top left panel shows where the incoming spectrum lies based on its recessional velocity. The middle left panel shows the now aligned spectrum in the units it will be stacked. The Bottom left panel shows how the average noise changes with the addition of each new spectrum. The large panel on the right shows the total stacked spectrum with the incoming spectrum on the same axes.*

3.2.4 Analysis Module

The Analysis Module contains one class which will save all the calculated data to file and, should the user wish, will fit a number of functions that can parametrise the stacked spectrum. The class contained within this module requires an input of a stacked spectrum type object and an input catalogue type object. This class serves as the parent object for the classes defined in the Display Module (Section 3.2.6) and the Uncertainty Module (Section 3.2.5).

The methods in this class are divided into three different types:

- Analysis methods – these include the fitting of various functions and the calculations of different statistics.
- Spectrum manipulation methods – these methods allow the stacked spectrum to be re-sampled or smoothed.
- Output methods – saving the calculated data, stacked spectrum and associated plots to file.

Upon initialisation of an instance of the analysis class, an initial statistical analysis is performed and the results are displayed either to a window or to the command line if running in suppress mode. The user is also asked if the integration window size should be changed from what it was set as during the input phase. The next steps are then displayed to the user:

1. Smooth the spectrum.
2. Fit a number of other functions to the spectrum.
3. Rebin the spectrum.
4. Do not fit anything to the spectrum, but continue with the analysis.
5. Exit this programme without doing anything else. (This option will not save anything)

Analysis Methods

Spectrum Statistics

The collection of statistics that are calculated by this module all characterise the significance of the stacked spectrum in some way. The statistical analysis of the stacked spectrum includes three different methods of calculating the signal-to-noise (S/N):

- peak S/N: $\mathrm{F_{HI}}/\sigma_{\mathrm{rms}}$
- integrated S/N: $\sum_i^{N_{\mathrm{chan}}} S_i \cdot \mathrm{dv}/\sigma \mathrm{dv}\sqrt{N_{\mathrm{chan}}}$ where S_i is the flux density associated with channel i, σ is the rms noise, dv is the channel width, and N_{chan} is the number of channels integrated over.

- ALFALFA S/N (only calculated if the spectral axis is in velocity units (Haynes et al., 2011)): $\left(\left(\frac{\mathrm{F_{HI}}}{W_{50}[\mathrm{km{\cdot}s^{-1}}]}\right)\sqrt{2RW_{50}}\right)/\sigma_{\mathrm{rms}}$ where the peak flux in the spectrum is given by $\mathrm{F_{HI}}$, W_{20} and W_{50} are the widths of the spectrum at 20% and 50% the peak flux value and R is the velocity resolution of the spectrum (this is also called the channel width)

The next statistic calculated is the p-value and the associated significance level. This is done by fitting a single Gaussian to the spectrum and calculating the χ^2 value using

$$\chi^2 = \sum\left(\frac{\mathrm{data}-\mathrm{model}}{\sigma}\right)^2 \tag{3.5}$$

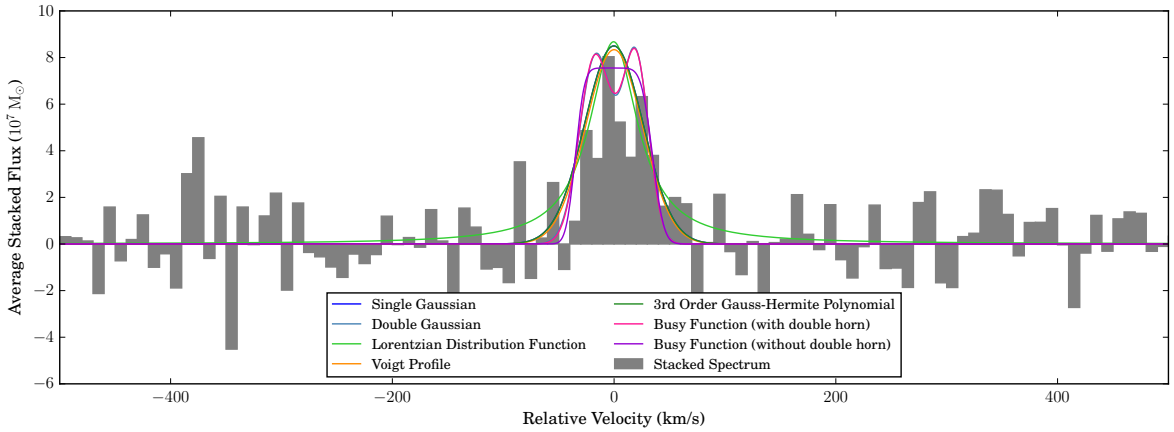
where the model is the fitted single Gaussian and σ are the fit weights (which are initially σ_{rms}). The single Gaussian is used in favour of other models as it is the simplest model that can describe the general shape of an HI detection; in the case of a very clear detection which cannot be characterised by a single Gaussian, the p-value statistic provides no extra information. The null hypothesis is a horizontal line at zero. Using χ^2_{stack} and χ^2_{null}, the log-likelihood ratio is calculated from which the p-value is determined using the area of a χ^2-distribution for 3 degrees of freedom. The p-value is the probability that the null hypothesis is true, so the significance of the stack is calculated using the complement of the p-value and the area of a Gaussian. The theory behind the calculation of the p-value is detailed in Appendix A.

Characterising the shape of the stacked spectrum

Stacking has been used to determine the HI Tully-Fisher relation of non-detected galaxies (Meyer et al., 2016); for this type of analysis, there needs to be a robust method of determining the HI line width (W_{50}). Low S/N stacked spectra can have a lot of noise spikes and dips that are unhelpful in determining the integrated flux of the spectrum. Thus, if one is wanting to perform a Tully-Fisher analysis or obtain robust estimates of the HI flux, it is useful to be able to fit an appropriate function to the stacked spectrum. This class offers a selection of 6 different functions that have been used in previous HI studies to characterise HI line profiles:

1. Single Gaussian
2. Double Gaussian
3. Lorentzian Distribution
4. Voigt Profile
5. 3^{rd} Order Gauss-Hermite Polynomial
6. Busy Function (Westmeier et al., 2014) (separate options are available for the inclusion or exclusion of the double-horn)

Figure 3.7 shows the 7 functions fitted to the stack of the 1000 profiles. Also shown is a table containing the integrated flux and a measure of the goodness of fit. The user may select any number of the functions, and HISS checks that the user is satisfied with



Fitted Function	**Integrated Flux ($M_\odot$)**	χ^2	χ^2_{red}
Single Gaussian	5.62×10^8	204	2.08
Double Gaussian	5.35×10^8	120	1.26
Lorentzian Distribution Function	7.01×10^8	328	3.35
Voigt Profile	5.37×10^8	207	2.14
3rd Order Gauss-Hermite Polynomial	5.62×10^8	204	2.1
Busy Function (with double horn)	5.33×10^8	120	1.27
Busy Function (without double horn)	5.27×10^8	132	1.38

Figure 3.7: *The top panel shows the 7 different functions (in various colours) fitted to the stacked spectrum which is represented by the grey bars. The table in the bottom panel shows the calculated integrated flux as well as two measures of the goodness of fit. A variation of this plot forms Diagnostic Plot 3 which allows the user to inspect the quality of the fit to the data before committing a selection of functions.*

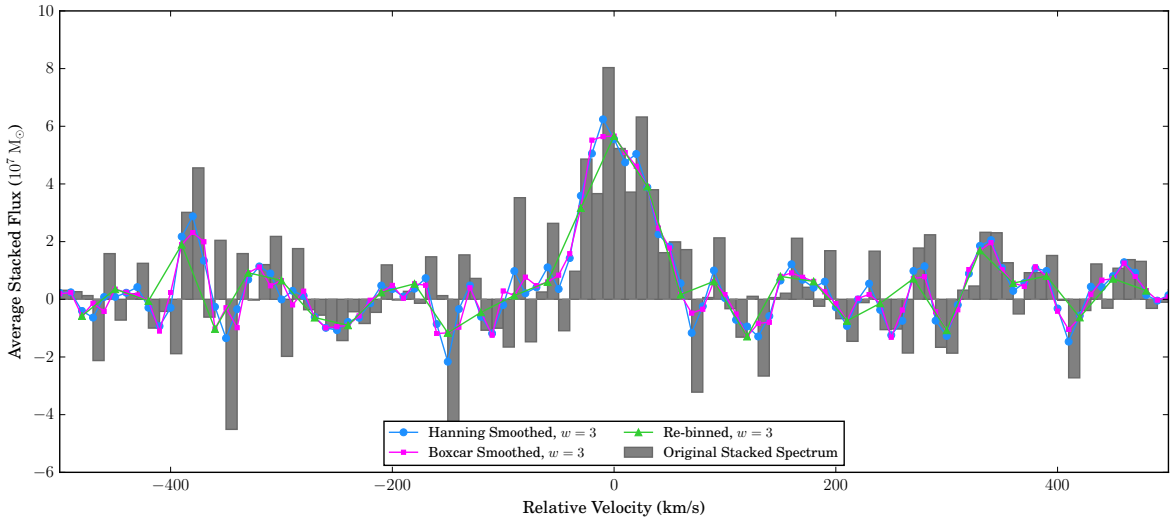
the selected functions before continuing. A version of Figure 3.7 is saved in PNG[3] format in the output location as Diagnostic Plot 3. The diagnostic plot is saved to file if HISS is run in suppress mode so that the user may inspect the function fit before continuing on with the analysis.

Spectrum Manipulation Methods

The final stacked spectrum depends on the input S/N of the individual spectra and the the number of spectra co-added, and therefore, the resulting final stacked spectrum is not guaranteed to show a detection. Thus, it may be necessary to manipulate the spectrum such that the signal appears more clearly above the noise. The Analysis Class offers the user the choice of re-binning the data (lowering the data resolution, this method has been used by **?**Rhee et al. (2016)), or to smooth the spectrum using either a boxcar algorithm or a Hanning Window.

All three manipulation routines allow the user to select either the window or the number of old bins to new bins. HISS will not continue to the next step until the user is happy with the new version of the stacked spectrum. Diagnostic Plot 2 (which is identical to Diagnostic Plot 1, Figure B.2, but displays the updated spectrum) along with a recalculation of the initial statistics is produced every time there is a change to the spectrum.

[3]Portable Network Graphics



Spectrum Type	σ_{noise} (M$_\odot$)	Peak S/N	Significance (σ)	p-value
Original Spectrum	1.51×10^7	**5.33**	> 8.2	1.59×10^{-17}
Hanning Smoothed	8.65×10^6	**7.21**	> 8.2	1.50×10^{-51}
Boxcar Smoothed	8.04×10^6	**7.04**	> 8.2	3.48×10^{-59}
Re-Binned ($\Delta v_{\text{new}} = 3\Delta v_{\text{original}}$)	8.63×10^6	**6.56**	> 8.2	1.32×10^{-16}

Figure 3.8: *Smoothing or re-gridding the stacked spectrum*

Output Methods

This section of the Analysis Class is the most critical for the HISS, this is where all the data and plots are saved to file. There are a possible nine files that are saved to the output location:

1. Stacked Catalogue: this text file contains a list of all the spectra included in the stack. Along with the data provided by the user from the input catalogue, this table also contains the integrated flux for each spectrum.

2. Output Data: a FITS table file containing the spectrum data (stacked spectrum, spectral axis and reference spectrum), the fitted parameters of any fitted functions, the stacked noise, and if the spectrum has been smoothed then the original version of the spectrum data is also saved.

3. Stacked Spectrum Plot: a PDF file of the stacked spectrum plotted with the fitted functions and the reference spectrum (a version of this is shown in the top left panel of Figure 3.9).

4. Stacked Noise Plot: a PDF file of the stacked noise which has the expected $\sigma/\sqrt{N}$ line over-plotted (top right panel of Figure 3.9).

5. Integrated Flux Data File: a table of the calculated integrated flux from the different functions as well as the flux integrated within the galaxy window. This file is saved in Encapsulated Comma Separated Value (`ecsv`) format using the AstroPy ASCII module which allows the units of the flux to be saved to file. Other columns in this file include the goodness of fit values from the function fits. A version of this table is shown in the bottom panel of Figure 3.9.

6. Diagnostic Plot 1: this is the first plot of the stacked spectrum that is displayed and saved upon initialisation of the Analysis Module.

7. Diagnostic Plot 2: a PNG file that is only produced if the user decides to rebin or smooth the spectrum.

8. Diagnostic Plot 3: a PNG file that is created when the user selects a function to fit to the stacked spectrum.

9. Diagnostic Plot 4: this plot is only produced if the uncertainties have been calculated. The plot contains a series of histograms showing the spread of the integrated fluxes. Diagnostic Plot 4 is saved to file in PNG format.

3.2.5 Uncertainty Module

The Uncertainty Module allows for two types of uncertainty calculations: a statistical error analysis and the redshift error analysis. The user specifies the type of uncertainty calculation (the two methods cannot be run simultaneously). The Uncertainty Module facilitates the storing of the final analysis options chosen by the user and uses those options to repeat the stacking process 1000 times. Figure 3.2 indicates the Uncertainty Module as a wrapper around the Stack and Analysis modules.

If the user chooses to engage the redshift uncertainty calculations, another loop is activated around the Stack and Analysis Modules which repeats the two processes 1000 times, each time changing the uncertainty associated with each spectrum in the input catalogue by an amount that is selected from a normal distribution with a standard deviation equivalent to the particular spectrum's redshift uncertainty. Thus, the redshift for each spectrum becomes $z = z + \mathrm{d}z$, where $\mathrm{d}z$ has been sampled from a normal distribution defined by $N(\mu = z; \sigma = u(z))$.

The result of repeating the stack and analysis process means that there are 1000 slightly different versions of the stacked spectrum and various integrated flux values. Each of the 1000 versions of the analysis object (which each contain a selection of integrated fluxes and the stacked spectrum data) are stored by the Uncertainty Class. Upon completion, the Uncertainty Class processes the stored data: the error bars on the spectrum data are calculated from the minimum and maximum values per channel of the 1000 stored spectra, the data points are given by the median values. There error-bars on the integrated fluxes are determined from the difference between the median value (which serves as the quoted value) and the 25^{th} and 75^{th} percentile.

The statistical error analysis implemented by this module is the *Delete-A-Group Jackknife* error method. In the same way as mentioned above, another loop is activated around the Stack and Analysis Modules repeating the process 1000 times, however each iteration discards a percentage (as chosen by the user) of the total catalogue without replacement. The 1000 analysed stacked spectra are stored by the Uncertainty Module. The flux values for the final stacked spectrum are taken as the mean value of the 1000 versions in each

channel. The error bars on the flux in each channel are calculated according to

$$\sigma(s) = \sqrt{\frac{R-1}{R}\sum_{n=0}^{R}(s_n - \bar{s})^2}, \tag{3.6}$$

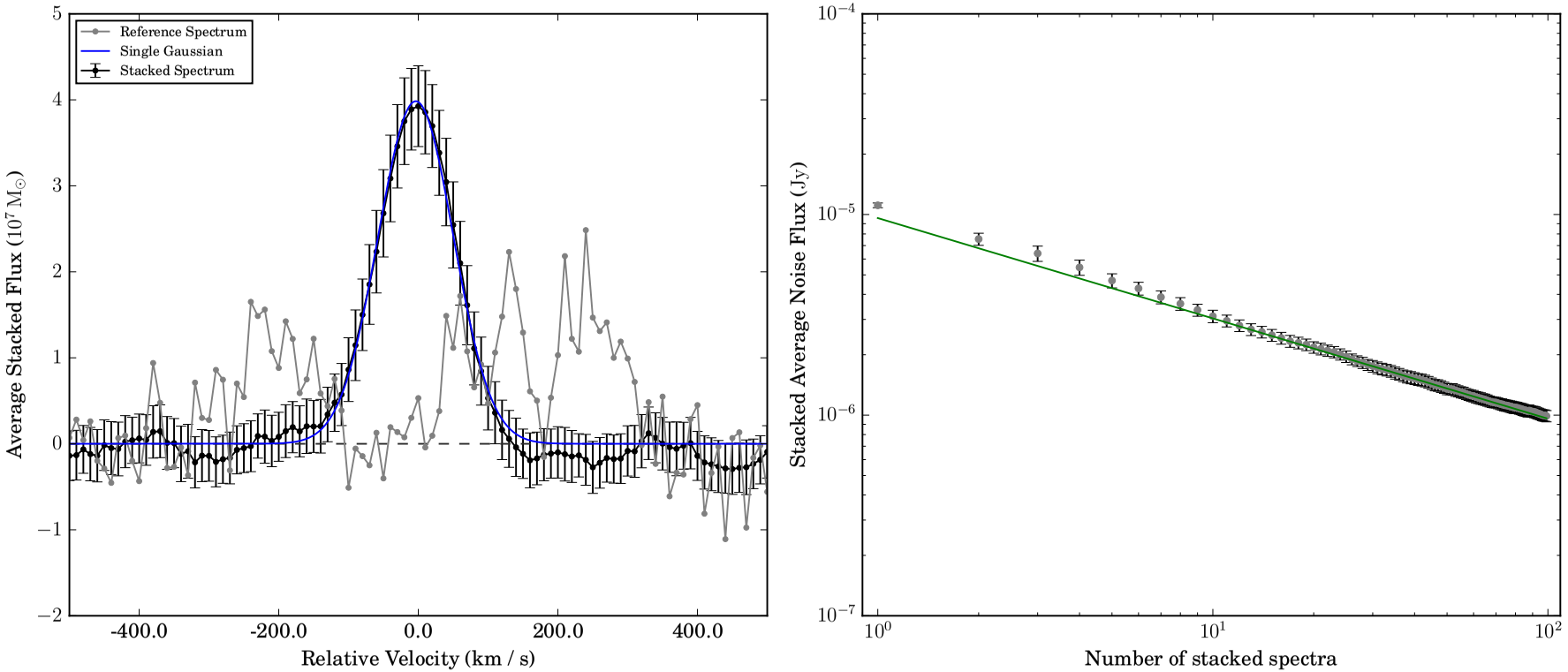
where R is the number of repeated estimates (which would be 4 if only 75% of the population was used for each iteration), s_n is the flux for the particular channel of the nth subset, and $\bar{s}$ is the mean flux value for the particular channel (Kott, 2001; Brown et al., 2015).

The class then calls the Output Methods from the Analysis Class to save the calculated data to file. Figure 3.9 shows the stacked spectrum and the stacked noise plots with the calculated uncertainties. The uncertainties on the integrated fluxes are also quoted in the table in the bottom panel of Figure 3.9.

3.2.6 Display Module

The Display Module is the final main module. This is the module that will display the final calculated quantities to the user. There are two modes to this module – a GUI mode and terminal-only mode which are activated when calling HISS. HISS will by default start in GUI mode, but can be changed to terminal-only mode using the command line parameter, `[-s]`. In the terminal-only mode, only the calculated integrated fluxes are displayed in the terminal; while in the GUI mode, the plots and the table of integrated fluxes are displayed in a graphics window (the graphics window for the sample of spectra used throughout this chapter is shown in Figure 3.9). This is not an essential module as all the data (as detailed in Section 3.2.4) is saved to disk, however it is a useful module in that it allows the user to view the results in one place.

The display class inherits the analysis class to re-use many of the methods required to display the results of the stacking process to screen. The few methods that are unique to this class are the ones which read in the data files that were saved to disk during the analysis.



Number of galaxy profiles included in Stacked Spectrum: 99
Total flux of sample = 5.09×10^{10} $M_\odot$
The peak signal-to-noise ratio: 34
The significance of the peak: 8.2σ (p-value = 1.64×10^{-141})

Fitted Function	**Integrated Flux** ($M_\odot$)	**Uncertainty Integrated Flux** ($M_\odot$)	**Fit RMS**	**Fit ChiSquare**
Single Gaussian	5.38×10^8	5.38×10^8	0.438	19.3
Summed Flux of Stacked Spectrum	5.40×10^8	1.13×10^7	0	0

Figure 3.9: *Screen shot of displayed output in full display mode. The top left panel shows the average stacked spectrum. In the top right panel is the stacked noise as a function of the number of profiles. The bottom panel contains the table of calculated quantities (integrated flux), and measures of the quality of the function fits.*

3.3 Software Testing

An essential step in the design of any software is the testing phase. HISS was designed in blocks. Each block had to pass a unit test before it was incorporated into the relevant module of HISS. Every method detailed in the earlier sections of this chapter have been tested using simulated HI spectra. The catalogue of simulated galaxies contained the HI masses for each galaxy so that there was a point of comparison for the non-graphical tests i.e. when testing that the conversion from flux density to HI mass, the HI mass from the catalogue could be compared with the calculated HI mass from the spectrum.

3.3.1 Package Test

Since each module has been thoroughly tested using simulated data, the test below uses data from the HIPASS survey (Barnes et al., 2001). The publicly available spectra for HIPASS can be obtained with either the flux as function of the frequency, or as a function of the velocity. The HIPASS catalogues also contain the calculated HI mass for each galaxy. This test is to make sure that the mass calculations produce the expected numbers regardless of the units of the input spectra's spectral axis, as well as to confirm that the functions used to characterise the final stacked spectrum can correctly fit the final spectrum.

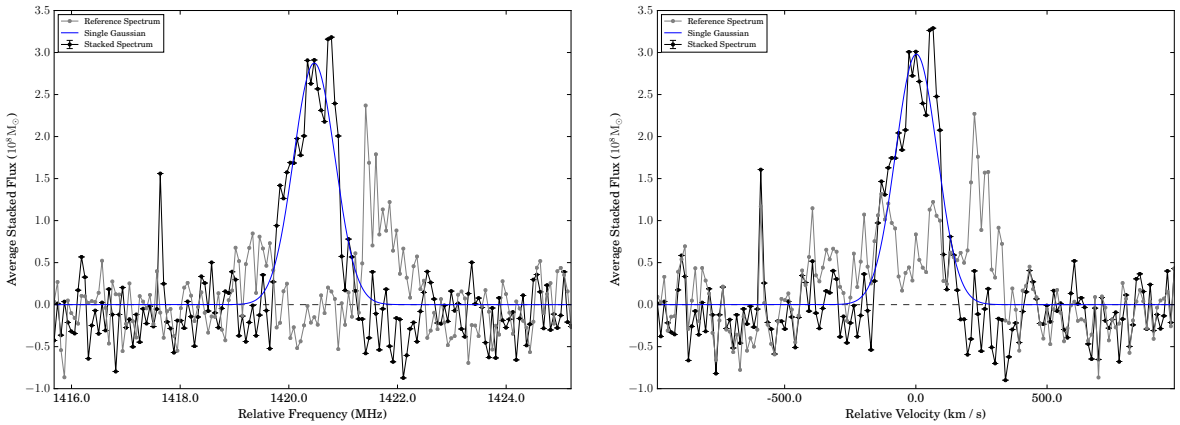
For this test, a selection of 5 HIPASS spectra were obtained for both spectral axis types. The five spectra are plotted in Figure B.1. The catalogue data for the 5 spectra are shown in Table 3.1. Based on the catalogue values, the sample as an average HI mass of $\langle \mathrm{M_{HI}} \rangle = (4.5 \pm 2.0) \times 10^9\,\mathrm{M_\odot}$.

Table 3.1: *HICAT data for the 5 HIPASS spectra (Meyer et al., 2004)*

HIPASS Name	Recessional Velocity ($\mathrm{km \cdot s^{-1}}$)	Redshift	$\mathrm{M_{HI}}$ ($\mathrm{M_\odot}$)
HIPASSJ0146-89	2445	0.00815	2.545×10^9
HIPASSJ2311-89	2524	0.00841	7.513×10^9
HIPASSJ0852-88	4981	0.01661	5.736×10^9
HIPASSJ1427-87	2250	0.00750	4.672×10^9
HIPASSJ1532-87	2286	0.00762	1.989×10^9

The stacked spectra for the two spectral axis types are shown in Figure 3.10. The HI masses derived from these spectra are listed in Table 3.2. It is clear from the agreement between the average catalogue mass and the values in Table 3.2 that HISS performs as designed. The fitted functions are able to recover the same average HI mass as integrating the spectrum and the catalogue value.

A similar test was performed using 10 simulated noiseless HI profiles from Elson et al. (2016), based on the Obreschkow et al. (2009c) catalogue, at an average redshift of $z = 0.71$. The average HI mass for the sample as given by the catalogue is $\mathrm{M_{HI}} = 2.6 \times 10^9\,\mathrm{M_\odot}$. The 10 spectra were stacked with both frequency and velocity spectral axes. Using an integration window of $\Delta v = 600\,\mathrm{km \cdot s^{-1}}$, the average HI mass calculated from the two stacked profiles is consistent with the average HI mass obtained from the catalogue.



(a) *Stacked spectrum for 5 HIPASS spectra with a frequency spectral axis.*

(b) *Stacked spectrum for 5 HIPASS spectra with a velocity spectral axis.*

Figure 3.10: *Stacked spectra for the 5 HIPASS spectra. The black line is the stacked spectrum, the reference spectrum is plotted in grey, and the single Gaussian used to characterise the shape of the spectrum is plotted in blue.*

Table 3.2: HI *masses derived from the stacked HIPASS spectra.*

	Velocity Spectral Axis $\langle M_{HI} \rangle$ ($M_\odot$) ($\Delta v = 450\,km \cdot s^{-1}$)	Frequency Spectral Axis $\langle M_{HI} \rangle$ ($M_\odot$) ($\Delta \nu = 2.85\,MHz$)
Summed Flux	$(4.5 \pm 1.3) \times 10^9$	$(4.8 \pm 1.2) \times 10^9$
Single Gaussian	$(5.1 \pm 1.2) \times 10^9$	$(5.0 \pm 1.2) \times 10^9$

3.3.2 Testing Software Data Handling Limitations

HISS has been tested using both simulated HI spectra from the S^3-SAX catalogue (see Section 2.1) and data from the single dish HI survey, HIPASS. The number of simulated HI spectra used during the testing process have mirrored the number of spectra used in published stacking studies (this ranges from $\sim$ 100 spectra to $\sim$ 2000 spectra). The format of the HIPASS data used in Section 3.3.1 is representative of the format of the majority of HI spectra currently available.

HISS has also been tested on three different operating systems. The system on which HISS was created, is Mac OS X El Capitan operating with 8Gb of RAM and an Intel Core i5 processor. HISS has also been tested on a server running Ubuntu 16.04 with 16Gb of RAM. The final system used to test HISS operates Ubuntu 14.04 with 4Gb of RAM.

Chapter 4

Stacking with CRUMBS and NIBLES

In this chapter, HISS is used to implement the stacking technique to analyse HI spectra for an isolated sample of galaxies observed as part of the Nançay Interstellar Baryon Legacy Extragalactic Survey (van Driel et al., 2016). Stacking the HI spectra of the sample of galaxies for which there was no direct HI detection can provide a measure of the average undetected HI mass. By stacking both non-detected and detected HI spectra together, we explore the average gas properties as a function of stellar mass and colour.

4.1 The Nançay Interstellar Baryon Legacy Extragalactic Survey

The Nançay Interstellar Baryon Legacy Extragalactic Survey (NIBLES) (van Driel et al., 2016) is a targeted HI survey of galaxies selected by stellar mass from the Sloan Digital Sky Survey (SDSS) (York et al., 2000). The HI data were obtained using the 100 m Nançay Radio Telescope (NRT) (Figure 4.1) in Nançay, France, and provided for this work by the Principal Investigator W. van Driel.

As mentioned in Section 1.3, both the HIMF and the optical luminosity function are required to link what is observed in the local universe to galaxy formation models. These two galaxy population tracers are usually studied independently. One of the main science goals for NIBLES is to study the two relations in conjunction thus learning about how they may be related (Butcher et al. in prep.). NIBLES provides a unique opportunity to study the baryonic content in galaxies with a wide range of stellar mass ($10^6 < M_\star[M_\odot] < 10^{12}$) galaxies. The range of ancillary data available for the survey (Section 4.1.5) supports the study of the HI content as a function of various global properties. Other science goals include determining the relationships and variability between the dynamical, stellar and HI masses of galaxies. NIBLES also aims to determine the gas content and gas fraction as a function of stellar mass, morphology, and average stellar density for the galaxy sample (van Driel et al., 2009).

In this thesis, HISS is used stack HI spectra from NIBLES to study the gas properties of different galaxy sub-samples.

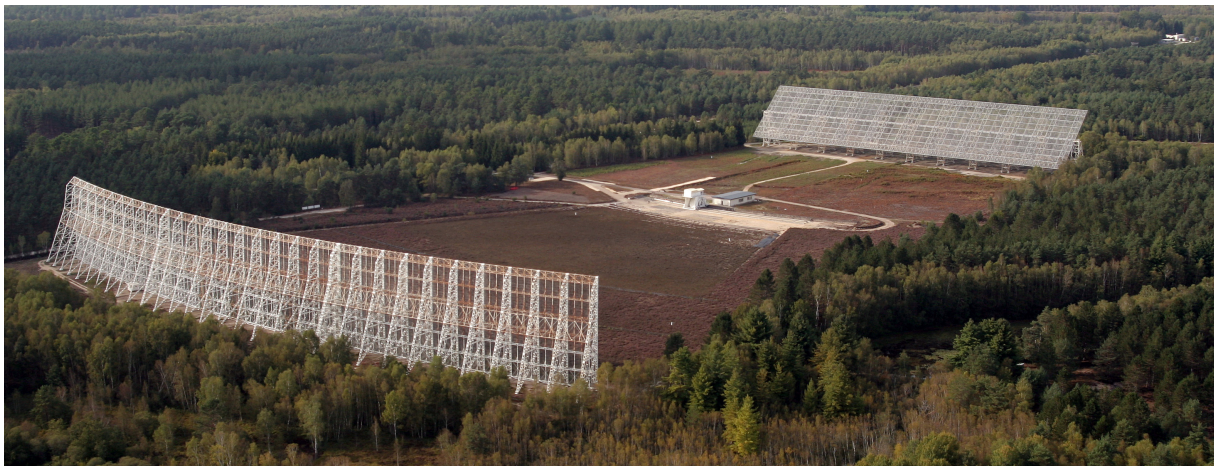


Figure 4.1: *The Nançay 100 m Radio Telescope located outside the town of Nançay in the centre of France. The telescope is comprised of the primary reflector (right), a tiltable 200×40 m reflector. The secondary reflector (left) is a 300×35 m curved reflector that is fixed 460 m from the primary. The instruments are located on a carriage that curves between the two reflectors. (van Driel et al., 2009, Figure 1)*

4.1.1 Galaxy Selection

The NIBLES sample of 3000 galaxies was selected from SDSS Data Release 5 (DR5). The selected galaxies were chosen to span a uniform stellar mass range. The particular selection criteria (van Driel et al., 2016) were as follows:

1. The target must have both SDSS DR5 spectroscopy and photometry.
2. NIBLES targets are limited to the Local Volume: $900 < cz < 12000\,\mathrm{km \cdot s^{-1}}$. Beyond $cz = 12000\mathrm{km \cdot s^{-1}}$, the NRT can only detect HI rich galaxies which typically have low optical surface densities making them more difficult to observe optically. The SDSS database has issues with correctly determining the photometry of galaxies with large angular sizes such as those typically located at $cz \leq 900\mathrm{km \cdot s^{-1}}$.
3. In order to study the HI in galaxies as a function of stellar mass, ~ 150 galaxies were selected per 0.5 mag bin for $-16.5\,\mathrm{mag} \leq M_z \leq -24\,\mathrm{mag}$ ($H_0 = 70\,\mathrm{km \cdot s^{-1}}$). M_z is used as a proxy for stellar mass.
4. In order to maintain a morphologically diverse sample, no colour selection was made.

In addition to the criteria listed above, van Driel et al. (2016), when selecting the target galaxies tried to avoid galaxies in and around the Virgo cluster as the dense environment is known to have noticeable effects on the HI properties of the galaxies in the region. The distances to galaxies in the Virgo region also have large uncertainties. The ALFALFA $\alpha.40$ footprint was also avoided (see van Driel et al. (2016) for details).

4.1.2 HI Data

The Nançay Radio Telescope is a 100 m single dish radio telescope which is located just outside the town of Nançay, France. The telescope (Figure 4.1) is a transit instrument with a collecting area of $6900\,\mathrm{m^2}$, and consists of two large reflectors. The primary reflector is 200 m long in the E-W direction, 40 m high and can rotate on its axis depending

on the observed declination. The secondary reflector is part of a sphere that has a radius of 560 m, this reflector is 300 m long and 35 m high. The instruments are located on a carriage which runs between the two reflectors which are separated in the N-S direction by 460 m. (Joseph, 2008; van Driel et al., 2016, and references therein.)

At 21 cm, the right ascension half-power beam width is 3.5′. Due to the shape of the telescope, the declination half-power beam width increases from 22′ at low declinations to 33′ at higher declinations. The change in the N-S beam width can be approximated by the polynomial

$$y = (4.02 \times 10^{-9})x^5 - (6.90 \times 10^{-8})x^4 + (3.63 \times 10^{-6})x^2 + (5.45 \times 10^{-3})x + 21.95 \quad (4.1)$$

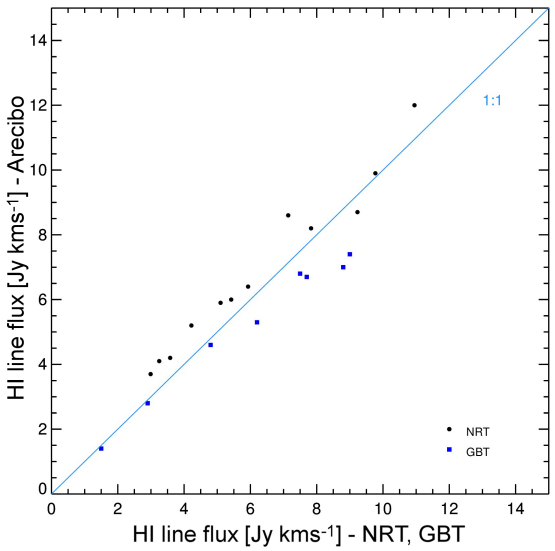
where y is the beam width in arcminutes and x is the declination in degrees (Matthews & van Driel, 2000, Fig. 1.).

The spectra were collected from January 2007 to December 2010. Each target was observed for roughly 40 minutes using successive 40 s ON source and 40 s OFF source pointings. The OFF source pointing was chosen to be 20′ E of each target. The data reduction process is discussed in detail by van Driel et al. (2016). Each spectrum has a velocity resolution of 18 km · s^{-1} and a typical noise level of 3 mJy. Spectra were obtained for 2840 of the 3000 selected targets of which 2600 were usable. 240 of the observed spectra were excluded from the final sample due to a number of different reasons, which include detections in the OFF source pointing and incorrect source classification in SDSS (i.e. source is not a galaxy) (van Driel et al., 2016, Sec. 4.8).

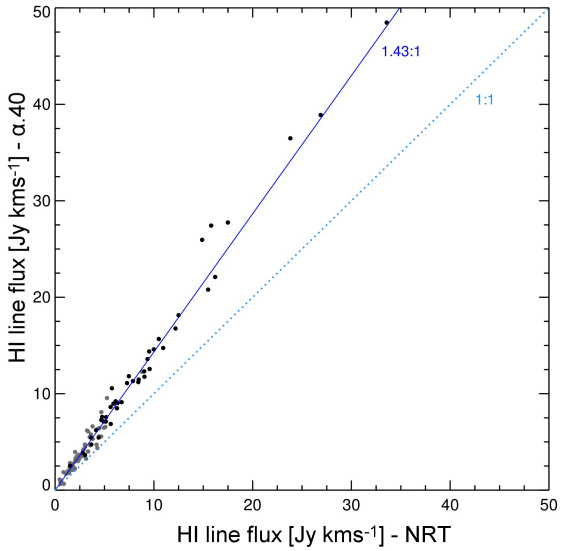
4.1.3 Comparison between NIBLES and other HI Surveys

In order to check the flux calibrations for the Nançay Radio Telescope (NRT), van Driel et al. (2016) observed 12 sources selected from the New Reference Galaxy Standards for HI Emission Observations by O'Neil (2004). The catalogue by O'Neil (2004) contains the data of various HI emission calibration sources obtained using the single-horn instruments on the Arecibo Telescope and the Green Bank Telescope (GBT). The NRT line fluxes were found to be consistent with the O'Neil (2004) line fluxes from both Arecibo and GBT. The flux ratio for Arecibo (using the single-horn instrument) to NRT is Arecibo/NRT $= 1.13 \pm 0.12$, and GBT/NRT $= 0.93 \pm 0.24$. Figure 4.2a shows the comparison of NRT, Arecibo, and GBT line fluxes.

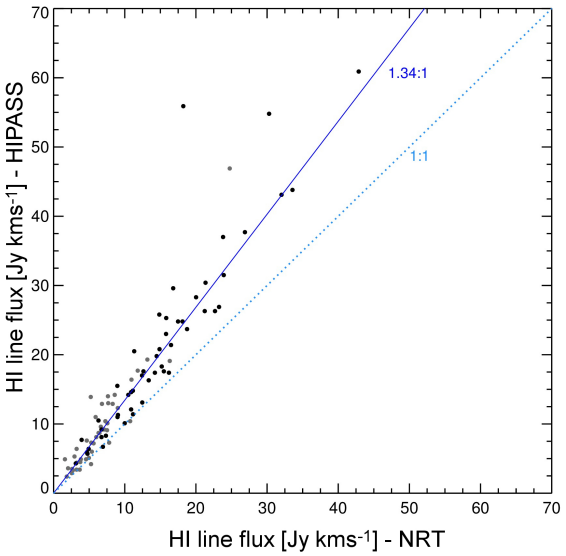
Van Driel et al. (2016) also compared the line fluxes of sources common to NIBLES, ALFALFA $\alpha.40$, and HIPASS. Sources from the three surveys were matched by position. Van Driel et al. (2016) also required that the difference of the central velocity and W_{50} of the matched sources were less than 20 km · s^{-1}. There were 82 matched sources with ALFALFA $\alpha.40$ (Figure 4.2b) and 51 with HIPASS (Figure 4.2c). In Figures 4.2b and 4.2c the black lines indicate the weighted mean relationship between the line fluxes from the two surveys. It is clear from Figures 4.2b and 4.2c that the line fluxes from ALFALFA and HIPASS are significantly higher than those from NIBLES. The flux ratio for HIPASS/NRT $= 1.34 \pm 0.28$ and for ALFALFA/NRT $= 1.45 \pm 0.17$. Van Driel et al. (2016) noted that the ALFALFA fluxes are higher than the line fluxes published by O'Neil (2004) for the Arecibo Telescope. One reason given for the difference is the difference in instrument and subsequently the method by which line fluxes are calculated



(a) Line flux density comparison for the Nançay Radio Telescope and the Green Bank Telescope with the Arecibo Telescope single feed receiver.



(b) Line flux density comparison for the Nançay Radio Telescope with ALFALFA survey conducted using the Arecibo Telescope in multi-beam mode.



(c) Line flux density comparison for the Nançay Radio Telescope with HIPASS survey conducted using the Parkes Telescope in multi-beam mode.

Figure 4.2: *NIBLES line flux densities compared to literature values (van Driel et al., 2009, Fig. 8)*

from the observations. Van Driel et al. (2016) acknowledged that determining the line fluxes from multi-beam observations (like ALFALFA and HIPASS) is a complex process; variations in choices of how the data is gridded can significantly affect the reconstructed line fluxes.

In this work, the NIBLES fluxes are quoted. In the event of comparison between NIBLES measurements and ALFALFA measurements, the ALFALFA measurements are scaled to the NIBLES flux values using the ALFALFA/NRT flux ratio.

4.1.4 Accounting for known possible contamination sources

Studies have shown that nearby galaxies can significantly contribute to the total emission in a target HI galaxy spectrum (Jones et al., 2015; Elson et al., 2016). For example, Elson et al. (2016) showed for a stacking experiment using simulated Parkes data at 15′ resolution of $0.04 < z < 0.13$, the average contaminant HI mass per galaxy spectrum was $\sim 1.4 \times 10^{10}\,\mathrm{M}_{\odot}$. Thus, before finalising the sample of spectra for stacking, it is necessary to remove spectra that are known to have nearby galaxies that could contribute significant contaminant emission to the spectrum. In other works (Fabello et al., 2012, etc.), a secondary source is considered to contaminate the target spectrum if it lies spatially within the half-power beam width and $\pm 300\,\mathrm{km \cdot s^{-1}}$ spectrally of the target. In this work, galaxies near the target source are considered as possible contaminators if

- the secondary galaxy is located spatially within 1.2× N-S half-power beam width (given by Equation 4.1) and 2× E-W half-power beam width, and
- the redshift of the secondary galaxy is within $\pm 300\,\mathrm{km \cdot s^{-1}}$ of the target redshift.

The above criteria were used to search for sources in both the SDSS Spectroscopic Database and the NASA/IPAC Extragalactic Database (NED). A number of the target galaxies were found to have possible contaminators with very small angular separation distances. Upon inspection of the optical images for these sources, it was discovered that the possible contaminators were in fact part of the galaxy. The SDSS pipeline has a known issue of deblending large sources, particularly those with knots of star formation. In order not to over classify the contaminant sources, a new criterion is introduced to the search: the contaminant source must lie outside the edge of the target source. The edge of each target galaxy is defined by twice the Petrosian radius, which according to Stoughton et al. (2002) is a good estimate of the size of the galaxy. The criteria used to search for possible contaminants are stricter than what was used in Fabello et al. (2012, etc.), but due to the size of the beam it was found that the stricter criteria were necessary.

Figure 4.3 is an illustration of the method used to find any secondary contaminating sources. On the left of the figure is an optical image that is 8′ wide and 1.2 × beam height (the beam height is obtained using Equation 4.1 and is dependent on the declination) centred on the coordinates of the target galaxy. Overlaid in red is the half-power size of the beam; the yellow circle has a radius of $2r_{\mathrm{P}}$ which indicates the size of the galaxy. The magenta open circular marker indicates the target source while the blue, green, and orange open circular markers indicate secondary targets in the survey volume ($900 < cz < 12000\,\mathrm{km \cdot s^{-1}}$). Each of the targets highlighted in the large optical image are listed in the table to the right. The galaxies that are in the survey volume, but whose recessional velocities are not in the spectrum velocity range, are indicated by the orange circles. The green circles highlight galaxies that are within the velocity range covered by the spectrum – these sources are also marked on the spectrum by the green dashed vertical lines. The blue-highlighted sources are the ones that are considered contamination sources. These sources have a redshift that is $\pm 300\,\mathrm{km \cdot s^{-1}}$ of the target redshift (highlighted by the dashed blue line on the spectrum). The small optical image to the left of the galaxy's HI spectrum is a $3.7' \times 3.7'$ close-up centred on the target source.

As previously mentioned only 2600 of the 2840 observed HI spectra are usable. In this analysis, 2497 of the 2600 HI spectra were used; 2 spectra were excluded due to different

velocity resolutions ($10\,\mathrm{km \cdot s^{-1}}$ instead of $18\,\mathrm{km \cdot s^{-1}}$), and the other 101 spectra not included in this study were eliminated due to unreliable photometry or unreliable redshift flags in the SDSS data. Secondary sources that could contribute contaminant emission to a stacked spectrum were found for 761 spectra.

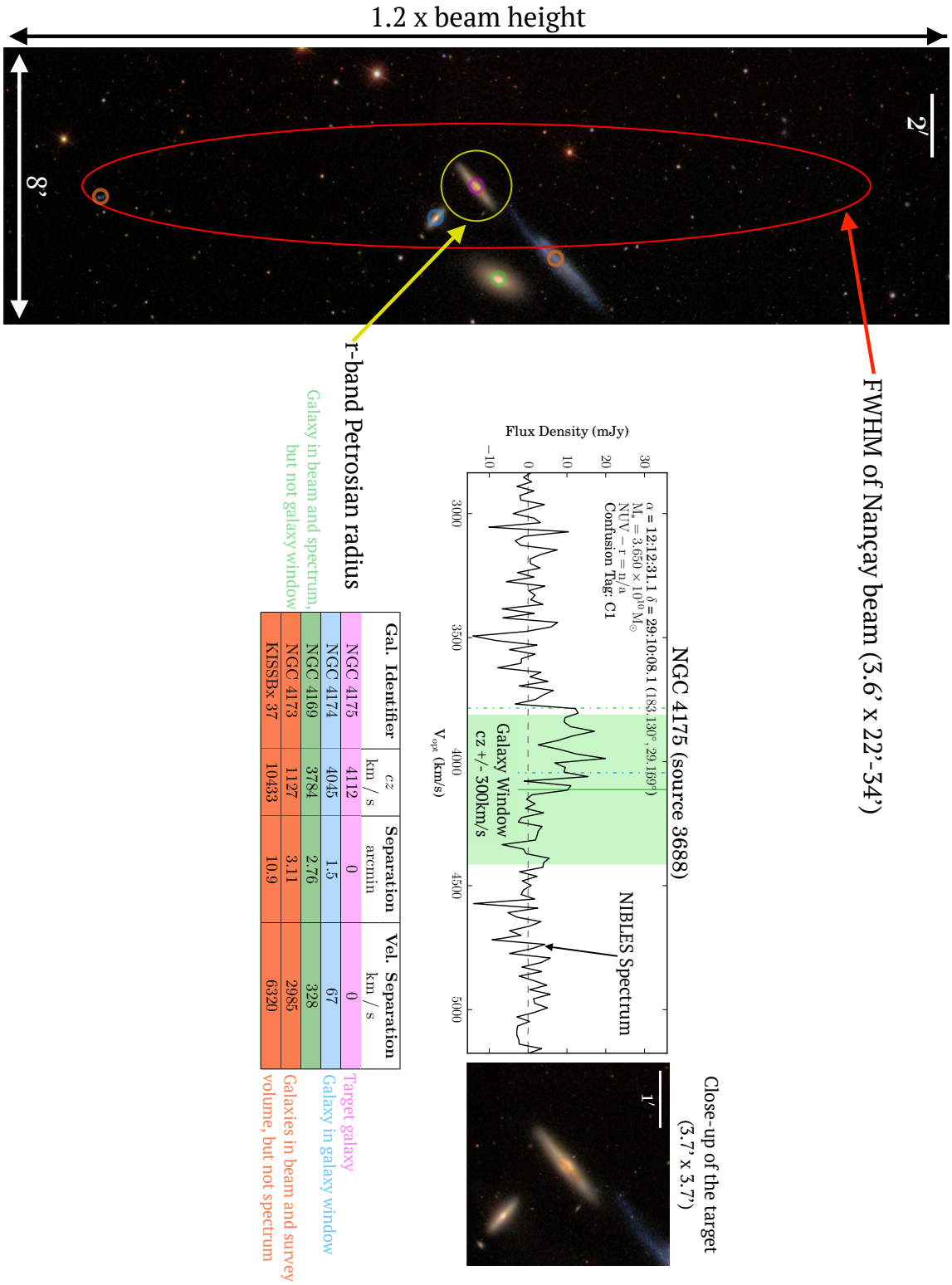


Figure 4.3: *The large optical image on the left is an cutout of the NRT beam centred on the target position. The yellow circle marks the edge of the galaxy according to the Petrosian radius. The target source is marked by the magenta circle. The blue, green, and orange circles mark galaxies that are spatially close to the target galaxy. Each of the marked galaxies are listed in the table which is ordered by spatial and velocity separation. The spectrum shows the* HI *spectrum for the target galaxy – the so-called galaxy window is highlighted by the green shaded region. Next to the spectrum is a* $3.5' \times 3.5'$ *optical close-up of the target galaxy.*

4.1.5 Ancillary Data

As mentioned in Section 4.1.1, the galaxies in NIBLES were selected from SDSS Data Release (DR) 5. In this work, data from the more recent SDSS DR12 (Alam et al., 2015) was used (more on this in the next section). SDSS has imaged roughly $8000\,\mathrm{deg}^2$ of the sky in five photometric bands (u, g, r, i, z). During the intervening years there have been more SDSS data releases. SDSS DR8-12 came with an updated photometry pipeline. The optical photometry used by NIBLES makes use of the DR9 photometry (except for a few cases where the original DR5 data are used, the exceptions are detailed in van Driel et al. (2016, Sec. 4)).

The SDSS photometric pipeline fits a number of models to each source in order to determine its magnitude. In this work, the model magnitudes are used. The model magnitudes are determined using the model (de Vaucouleurs profile or exponential profile) which better fits the r-band flux. The optical magnitudes are corrected for Galactic extinction using the SDSS-provided reddening corrections which are calculated following Schlegel et al. (1998) at the position of each source (Stoughton et al., 2002). In this work, the Galactic extinction corrected model magnitudes are used to define the galaxy colours.

Where available, the stellar mass and star formation rate for each galaxy are obtained from the "value-added" JHU/MPA catalogues (Brinchmann et al., 2004) which have been matched to the SDSS DR12 spectroscopic database; thus the stellar mass values used in this work have been obtained from the catalogue stored in the SDSS database by matching SDSS object IDs. The catalogue provides a number of different measures of the stellar mass and star formation rate, however for the purposes of the NIBLES analyses, the median values are used (van Driel et al., 2016).

Funded by NASA, the Galaxy Evolution Explorer (GALEX) satellite has provided a comprehensive view of the ultra-violet universe (Bianchi & GALEX Team, 2000). The GALEX survey area overlaps with SDSS making it the ideal complementary source of near-UV (NUV) and far-UV (FUV) data for SDSS-selected targets. The NUV data used in this analysis are obtained from GALEX GR6/7 which are publicly available on CASjobs[1] where GR6/7 has been cross-matched with SDSS DR12. In this work, the NUV data have been corrected for Galactic extinction using $A_{\mathrm{NUV}} = 8.2 \times E(B-V)$ (Wyder et al., 2007).

SDSS Photometry Issues

A portion of this work is to study the HI properties of galaxies in the NIBLES sample for which there was no direct HI detection. The HI detected sample provides a useful sample to understand how the HI properties of the non-detected sample relate to the optical properties. While investigating the distribution of the HI mass ($\mathrm{M_{HI}}$) with respect to the stellar mass ($\mathrm{M_\star}$), a sample of low stellar mass red galaxies were discovered. The stellar masses from the Brinchmann et al. (2004) catalogue are calculated using the SDSS photometry from DR7. The data points in left panel of Figure 4.4 are coloured by the optical colour classification (i.e. the blue points are classified as blue galaxies, and the red points are red galaxies).

[1] http://galex.stsci.edu/casjobs

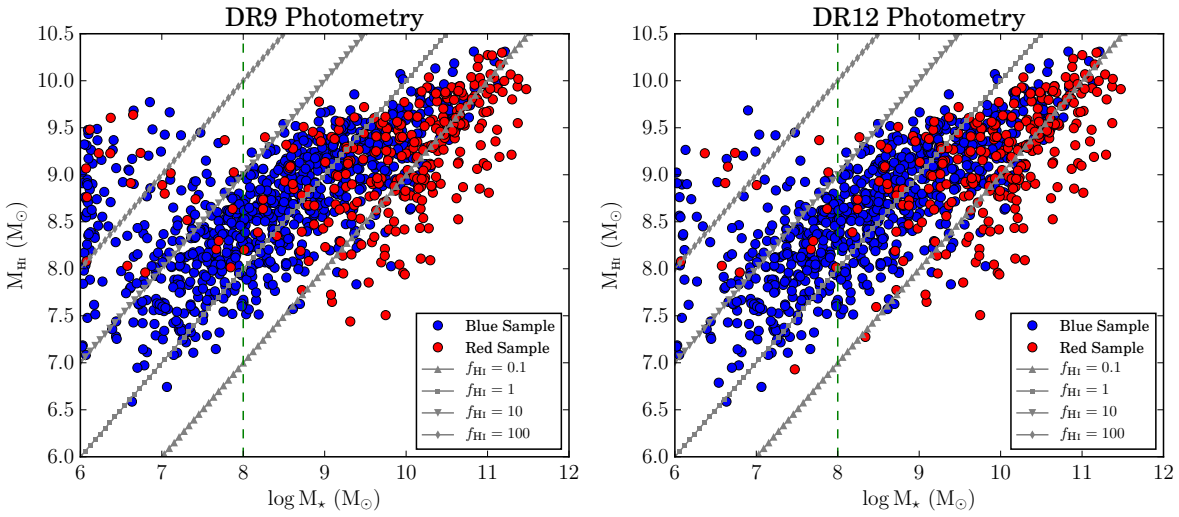


Figure 4.4: *$M_\star$ vs $M_{\rm HI}$ for the NIBLES detections. The grey markers denote the lines of constant $M_{\rm HI}$ to $M_\star$ fractions (gas fraction). The blue and red markers indicate which sample the source belongs to based on its $u-r$ colour and the Baldry et al. (2004) colour divider.*

Galaxies with $M_\star < 10^9\,M_\odot$ are typically considered dwarf galaxies and are usually blue in colour. Upon closer inspection of the optical images of the galaxies (Figure 4.5) it was discovered that there may be problems with the photometry. For example, image 1942 (top left) in Figure 4.5 has u-r= 2.51 which is very red, however $NUV - r = 1.40$ which is definitively blue. Upon checking the SDSS pipeline photometry flags, it was discovered that many of the sources with seemingly problematic photometry indeed had various flags indicating that there were issues with the photometry.

Updating the photometry to SDSS DR12 reduced the number of low mass red galaxies, evidenced by the fewer red data points with $M_\star < 10^8\,M_\odot$ in the right panel of Figure 4.4. However, the revised photometry does not remove all of the issues. Figure 4.6 shows that there are some red galaxies with dwarf-like stellar masses which do not seem reasonable. For example image 3241 (2nd row, 3rd column) in Figure 4.6 is approximately 1 arcmin in diameter with a dust lane which usually indicates late-type morphology, both of these details are indicative of a higher stellar mass than what is quoted. Some masses are potentially overestimates – image 5184 (bottom left) is of a dwarf-like low surface brightness galaxy that is barely distinguishable in the $3.5' \times 3.5'$ image, thus the quoted stellar mass for this galaxy seems like an over estimate. The stellar masses were based on older SDSS photometry (DR7), which means they are still subject to the photometric errors of the earlier data pipelines. The problematic stellar masses are particularly pronounced at the lower stellar masses, so for the purposes of this study, only galaxies with stellar masses of $M_\star > 10^8\,M_\odot$ are used.

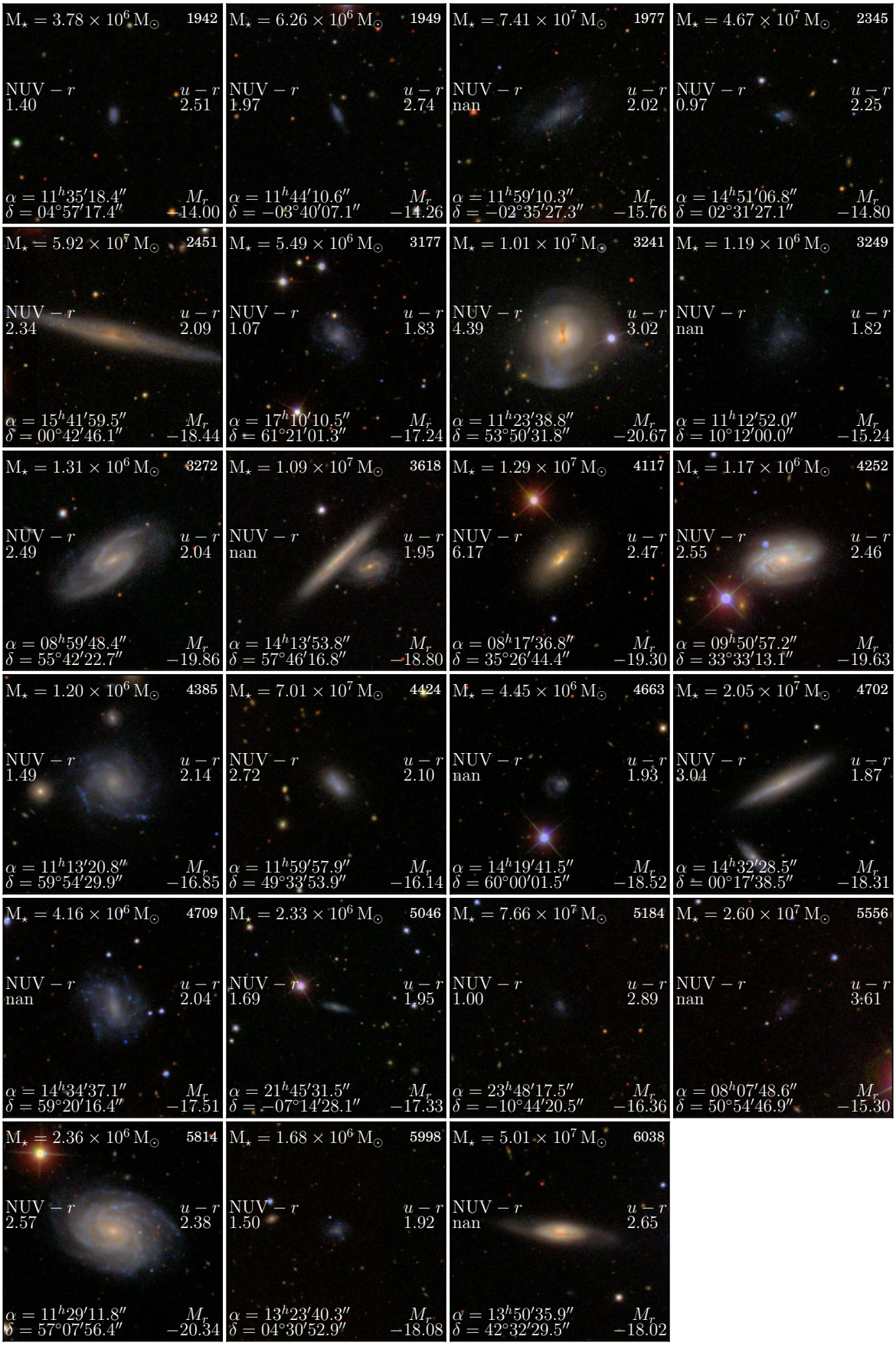


Figure 4.5: *Galaxies that have been classified as red by the Baldry et al. (2004) criteria using SDSS DR9 photometry. These galaxies also have stellar mass of* $M_\star < 10^8\ M_\odot$.

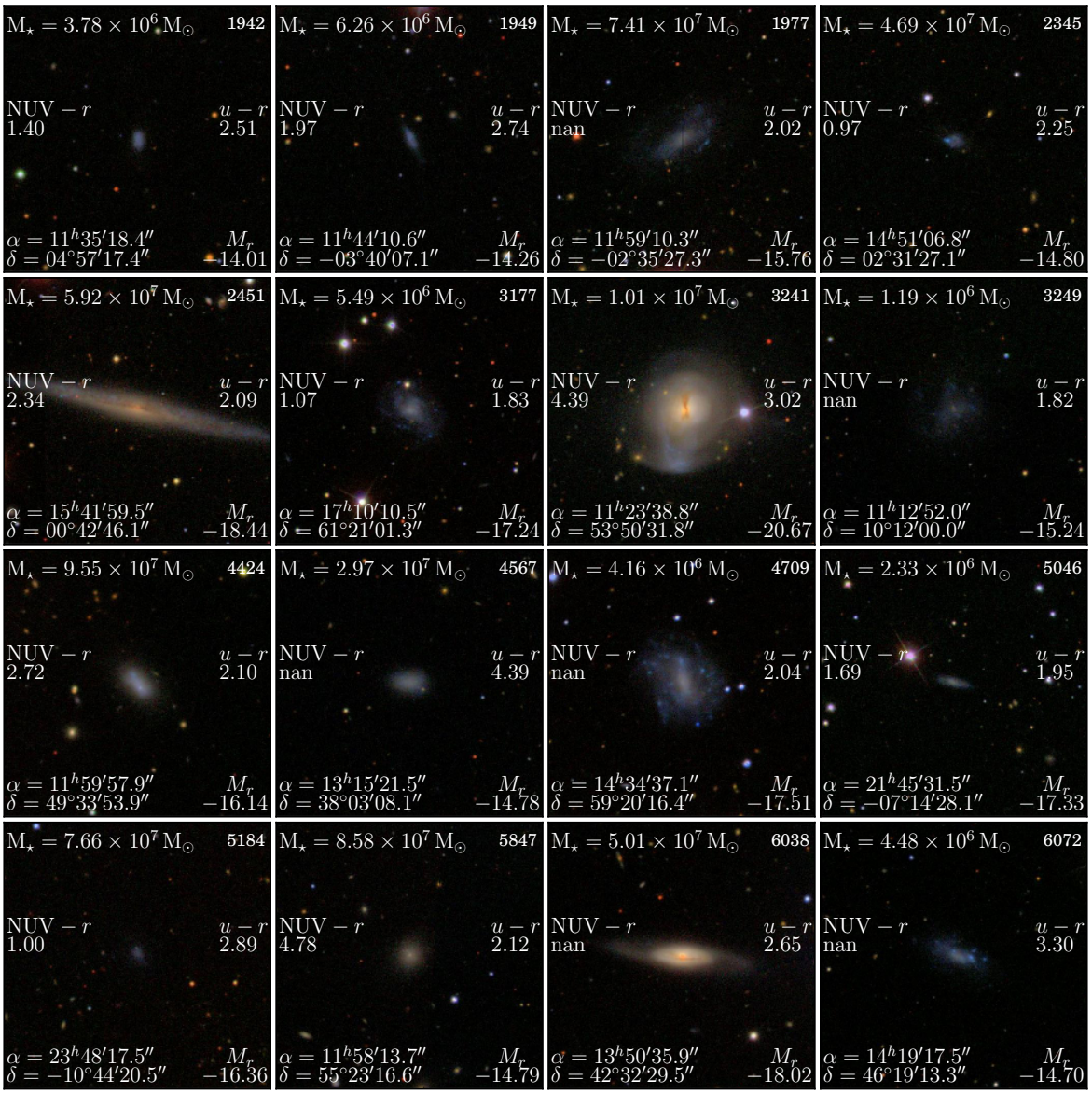


Figure 4.6: *Galaxies that have been classified as red by the Baldry et al. (2004) criteria using SDSS DR12 photometry. These galaxies also have stellar mass of* $M_\star < 10^8\ M_\odot$.

4.2 Characterizing Radio-Undetected Masses in Baryonic Surveys (CRUMBS)

This study focusses on characterising the HI mass of the NIBLES galaxies that lie below the NRT sensitivity limit. This is the first application of the HI Stacking Software, developed in this work, to non-simulated data. There are 1736 HI spectra used in the stacking analyses that follow, these spectra have been cleared of potential nearby contaminant emission (see Section 4.1.4).

4.2.1 Defining a sample of non-detected HI spectra

The first step in this study was to quantitatively define a sample of non-detected HI spectra. In order to create a sample of non-detected spectra, the HI detected spectra needed to be removed from the sample. Classifying a spectrum as detected can be tricky as galaxies can span as few as 1 or 2 channels ($\sim 40\,\mathrm{km\cdot s^{-1}}$), but as many as 30 channels ($\sim 600\,\mathrm{km\cdot s^{-1}}$). An edge-on very massive and distant galaxy rotating $600\,\mathrm{km\cdot s^{-1}}$ will span up to 30 channels, but may not have a very high S/N. Thus using the typical methods of defining a detection can be inadequate. It was, therefore important to define an objective, reproducible method of classifying the HI spectra as detected or not.

The detection classifications by van Driel et al. (2016) were made by eye by three independent judges. These classifications are sufficient for Butcher et al. (in prep.), Lehnert et al. (2016), Lehnert et al. (in prep.) etc., but for this work an objective, quantitative way to separate the detections from the non-detections is needed. The method that is described below is one that searches through each spectrum looking for an HI detection. Since it is the spectra that contain no HI detections that are of interest in this study, the spectra that are found to contain an HI detection are removed leaving only the spectra which contain no HI detections. This detection method produced very similar detection classifications to van Driel et al. (2016).

Every spectrum was smoothed to a velocity resolution of $54\,\mathrm{km\cdot s^{-1}}$ using the box car smoothing algorithm with a window size of 3 channels. Box-car smoothing with a window size of 3 channels averages every change with the channel on either side. The result of smoothing decreases the noise fluctuations and enhances the signal. The effect of the smoothing is clear in Figure 4.7, the noisy unsmoothed spectrum is plotted in light blue and the smoothed spectrum in grey. Next a measure of the spectrum's noise needed to be determined; this was done using the interquartile range of the smoothed spectrum (the difference between the 50^{th} and 25^{th} flux percentiles). The interquartile range (IQR) is a more robust measure of the noise level as it is not sensitive to outliers and allows for the entire spectrum to be used in determining the noise level. The 1 IQR is indicated by the light purple band in Figure 4.7. A window of $600\,\mathrm{km\cdot s^{-1}}$ centred on the target redshift was defined as the galaxy window – if there is any detectable emission in the spectrum, it should be contained to the galaxy window. A spectrum was classified as detected if (within the galaxy window) any one of the following criteria were satisfied:

- more than 1 channel with flux above 5 IQR
- more than 2 consecutive channels above 4 IQR

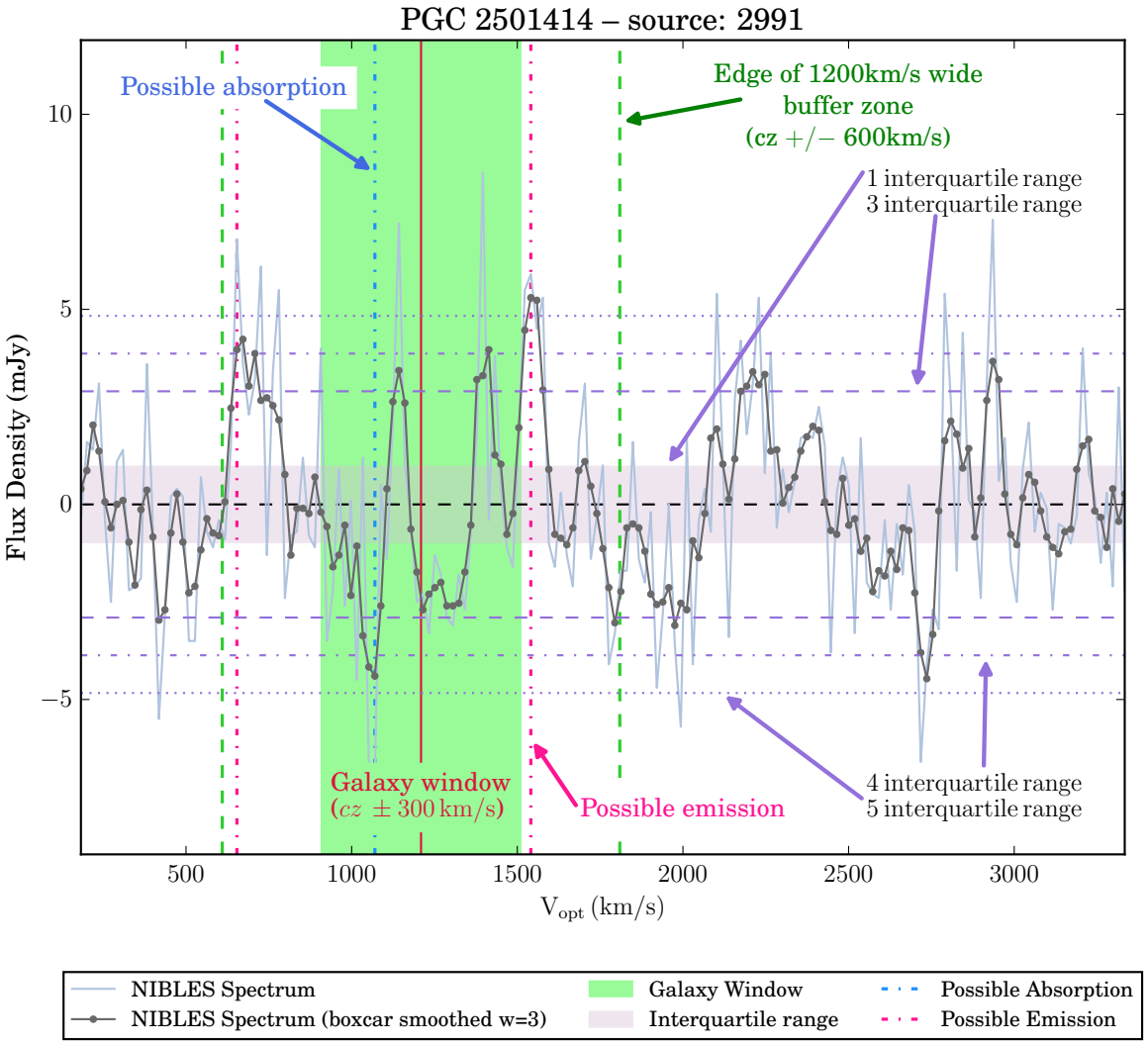


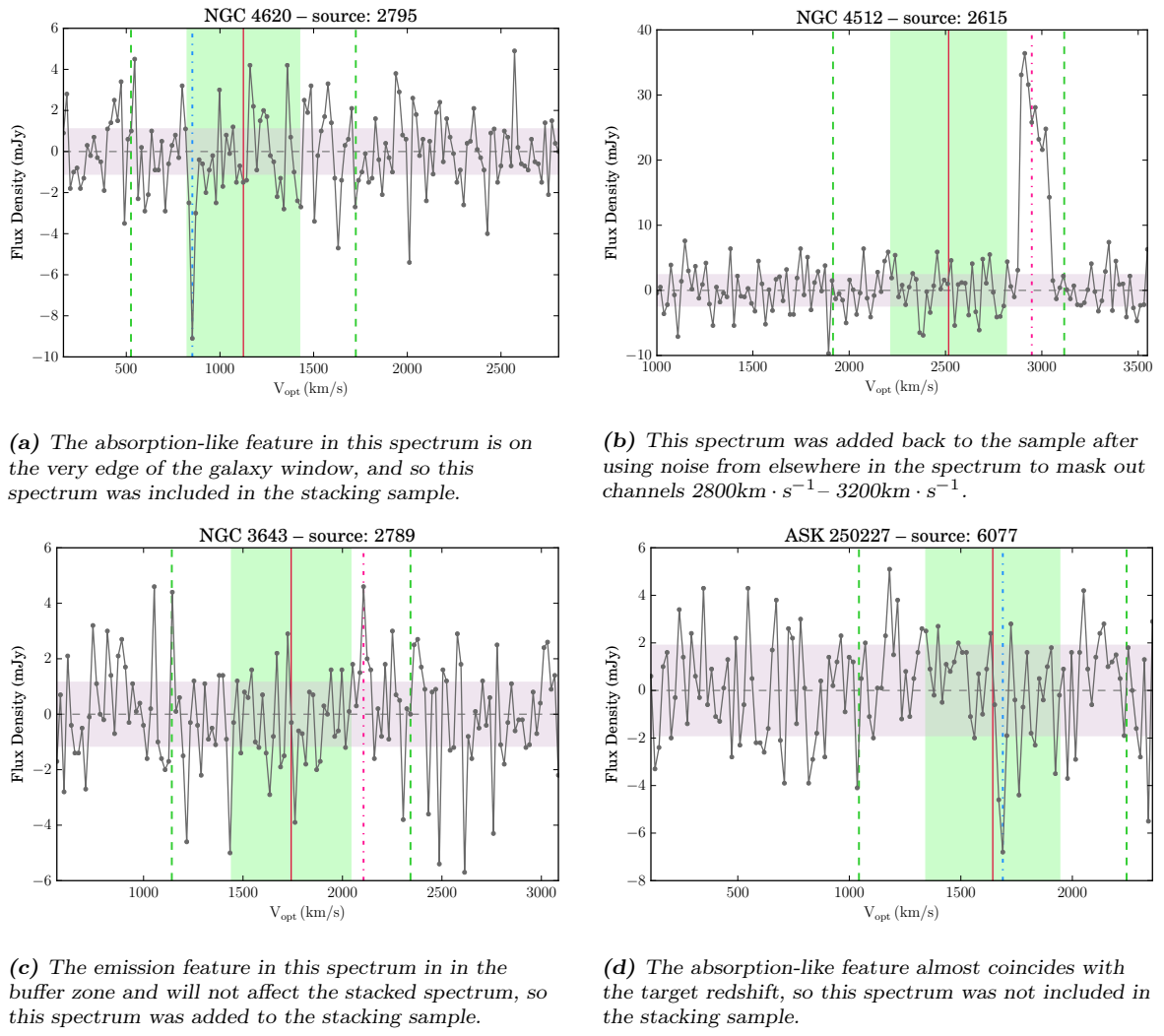
Figure 4.7: *Graphic example of how the detection algorithm flags different features in a spectrum. The spectrum used in this example is classed as a non-detection.*

- more than 4 consecutive channels above 3 IQR

Table 4.1: *The three columns show the detection classifications by van Driel et al. (2016) and the two rows are the detection classifications from this work.*

	Detected	Marginally Detected	Not Detected
Detected	1192	89	50
Not Detected	6	24	377

Using this method, 407 out of 1738 spectra were not classified as detections (i.e they were non-detections). The exact breakdown of the detection status for the sample with respect to the van Driel et al. (2016) classifications is shown in Table 4.1. Each of the 407 non-detections was then checked by eye. It was noted that a number of them contained possible absorption within the galaxy window or possible emission features at the edge of the galaxy window. If these possible absorption/emission features were legitimate ad-



(a) *The absorption-like feature in this spectrum is on the very edge of the galaxy window, and so this spectrum was included in the stacking sample.*

(b) *This spectrum was added back to the sample after using noise from elsewhere in the spectrum to mask out channels 2800km · s^{-1}– 3200km · s^{-1}.*

(c) *The emission feature in this spectrum in in the buffer zone and will not affect the stacked spectrum, so this spectrum was added to the stacking sample.*

(d) *The absorption-like feature almost coincides with the target redshift, so this spectrum was not included in the stacking sample.*

Figure 4.8: *A collection of spectrum that were flagged as containing possible absorption/emission features.*

ditions to the spectrum they could be due to contamination from other galaxies or RFI[2] sources, thus it would be prudent to exclude them from the stacking sample.

The detection method was re-run looking for absorption features (using the same criteria as for emission, but negative) and emission features in a wider window of 1200 km · s^{-1} (the buffer zone) centred on the target redshift. An example of the results from this search is shown in Figure 4.8. Any spectra that contained either possible absorption or emission features was flagged. Of the 407 non-detected spectra, 69 spectra were flagged containing absorption or emission features. Any spectra that contained an absorption feature on the very edge of the galaxy window (highlighted by the vertical green band in Figure 4.7) and did not affect the emission in the galaxy window were added back into the stacking sample (Figure 4.8a). The spectra that contained emission features that did not affect the galaxy window or dominate the spectrum flux were added to the stacking sample (Fig-

[2]Radio Frequency Interference

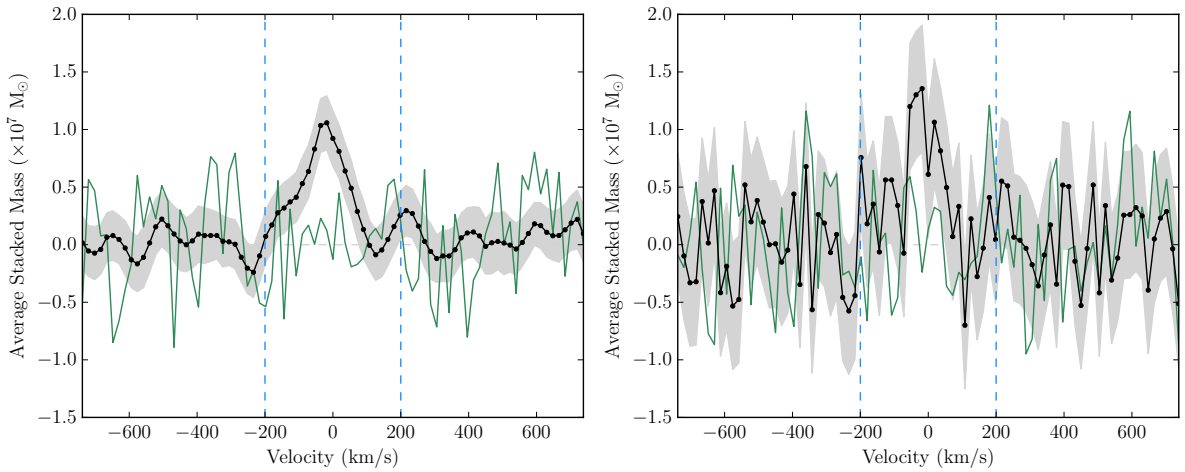


Figure 4.9: *The stacked spectrum arising from stacking 361 non-detections. The left panel shows the stacked spectrum produced when calculating the uncertainty due to redshift errors, while the right panel shows the stacked spectrum produced using the DAGJK method. The black line indicates the stacked spectrum, the green represents the reference spectrum and the vertical dashed blue lines denote the integration width. The grey bands in the two plots indicate the uncertainties in each channel.*

ure 4.8c), while those which contained emission features in the buffer zone that dominated the spectrum flux were masked using noise from elsewhere in the spectrum (Figure 4.8b). Examples of the flagged spectra are plotted in Figure 4.8. Out of the 69 flagged spectra, 50 spectra were added back to the stacking sample. The 19 of the remaining flagged spectra were eliminated due to either absorption-like or emission-like features in the galaxy window.

The final sample contains 1716 spectra, of which 1330 are classified as HI detections and 386 spectra are not classified as HI detections (this is the non-detected sample). As mentioned in Section 4.1.5, this work makes use of SDSS DR12 photometry rather than the SDSS DR9 Photometry. Once the stacking samples had been finalised, the SDSS object IDs were used to obtain the DR12 photometry for each source. Of the 1716 sources in the stacking sample, 243 did not have an ID match in the DR12 photometric database. The sample is now comprised of 1112 detections and 361 non-detections.

4.2.2 Stacking the non-detections

The final sample of non-detected spectra contains the HI line profiles for 361 galaxies. The catalogue data and spectra for the non-detected spectra can be found in Section C.1. The stacked spectrum for the non-detected sample is shown in Figure 4.9. The following statistics for the stacked spectrum are obtained from HISS (see Section 3.2.4). The spectrum has an integrated S/N of 170 and the significance of the detection is 7σ (p-value $= 1.3\times10^{-12}$). The mean noise level for the NIBLES survey is 2.1 mJy (van Driel et al., 2016). At the NIBLES mean distance and redshift ($\langle D_L\rangle = 43$ Mpc and $\langle z\rangle = 0.01$) the 5σ upper limit of the HI mass not detected by NIBLES is $\mathrm{M_{HI}} = 3.85\times10^8\,\mathrm{M_\odot}$ (assuming a width of $400\,\mathrm{km\cdot s^{-1}}$). The average undetected HI mass determined from the stacked spectrum is

$$\langle \mathrm{M_{HI}}\rangle = (9.72 \pm 0.72(\mathrm{sys}) \pm 0.24(\mathrm{stat})) \times 10^7\,\mathrm{M_\odot}.$$

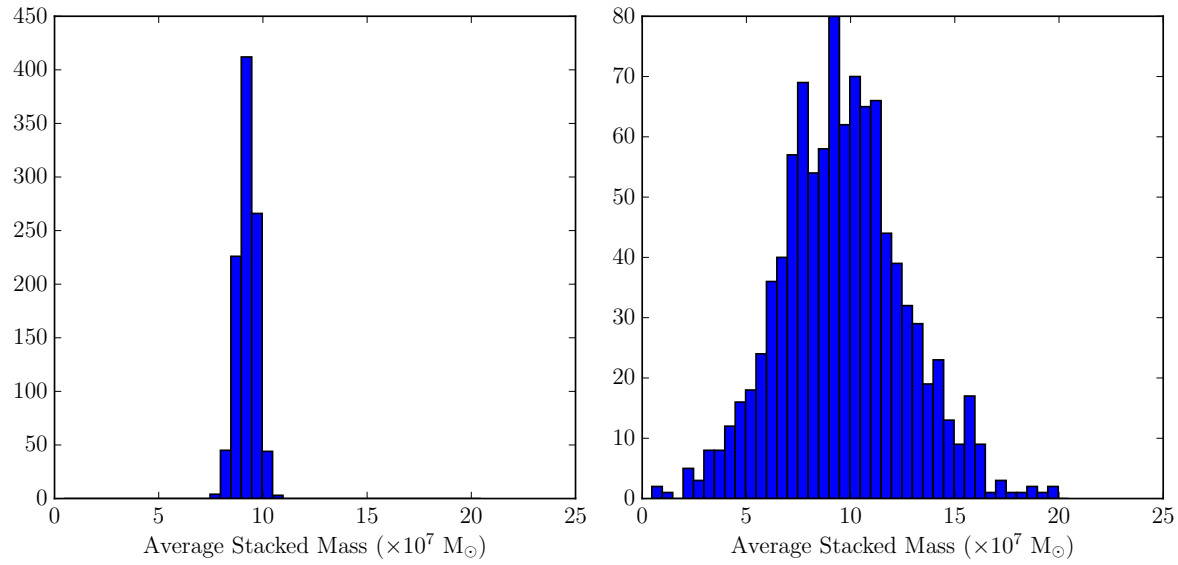


Figure 4.10: *The left panel is the distribution of $\langle M_{\rm HI} \rangle$ arising due to redshift uncertainty. The right panel shows the $\langle M_{\rm HI} \rangle$ obtained after 1000 iterations of the DAGJK method.*

Estimating the uncertainties

There are two uncertainties quoted in this analysis – the systematic uncertainty which arises from the redshift uncertainty, and the statistical uncertainty due to the variation within the sample under study. In calculating the uncertainties, HISS repeats the stacking experiment 1000 times, perturbing the stack on each iteration (for redshift uncertainties, this means that the individual redshifts are shifted thin their uncertainty; for the statistical uncertainties, the Delete-A-Group Jackknife (DAGJK) method is used, where 25% of the catalogue is discarded for each iteration).

For each of the 1000 stacking iterations, a stacked spectrum is produced from which an average HI mass (or gas fraction) is determined by integrating the stacked spectrum within the chosen width. For this spectrum, the range over which the spectrum is integrated is $\Delta v = 400\,\mathrm{km \cdot s^{-1}}$ (as indicated by the dashed vertical lines in Figure 4.9). The average mass (or gas fraction) is calculated from each iteration, which upon completion of the iterations, is a collection of 1000 measurements. From this distribution the average mass and the uncertainty of the average mass are extracted. The mean of the distributions shown in Figure 4.10 are the quoted average HI mass and the standard deviation is the quoted uncertainty.

The average stacked spectrum is plotted in black with a grey band representing the systematic error bars in the left panel of Figure 4.9, while the right panel shows the average stacked spectrum with the grey band representing the statistical error bars. The green line indicates the reference spectrum which was created by randomly assigning a redshift to each spectrum and stacking (as discussed in Section 2.5.1).

The average HI mass of the sample provides an estimate of how much gas is present in the galaxy, however it is the average gas fraction which provides a measure of the sample's gas-richness. As mentioned in Section 4.1.5, the stellar masses below $10^8\,\mathrm{M}_\odot$ are unreliable; thus for the sections that follow, only galaxies with stellar masses of $\mathrm{M}_\star \geq 10^8\,\mathrm{M}_\odot$

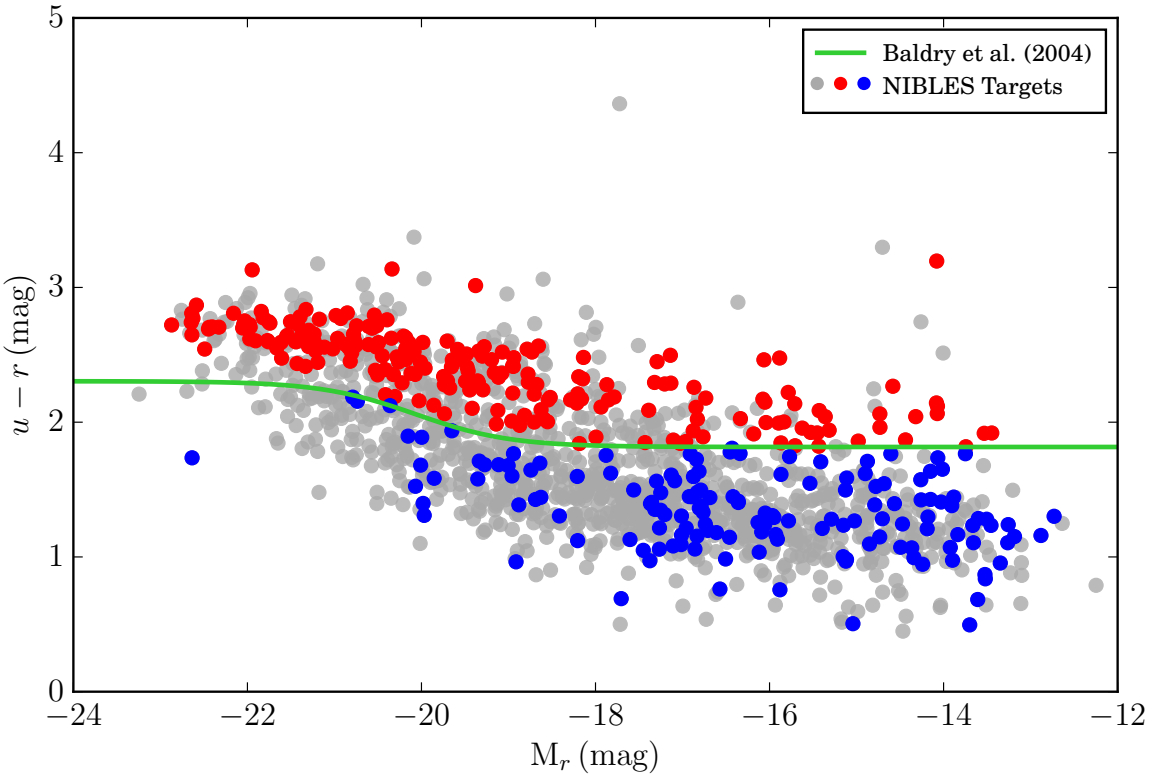


Figure 4.11: *Optical colour-magnitude diagram for the NIBLES galaxies. The* HI *non-detected galaxies are plotted in either red or blue based on their position above or below the Baldry et al. (2004) colour divider which is indicated by the green line. The grey data points represent the* HI *detected galaxies.*

are used in the stacking experiments.

The stacked spectra in this work are created using the HI Stacking Software. The average HI mass measurements of the different sub-samples are made from stacking the spectra in units of HI mass per channel; similarly, the average gas fractions are calculated from the stacked spectrum created by stacking in gas fraction units per channel (Equation 2.7).

Stacking by colour and by stellar mass

One of the selection criteria of the NIBLES sample was no *a priori* galaxy colour selection, this means that the galaxies making up the sample are diverse in morphological type. Galaxies of different morphological properties have been shown to have different HI gas properties. Optical colour is often used as a proxy for morphological type (e.g. Baldry et al. (2004)). In this work, the SDSS $u-r$ colour is used as a proxy for morphological type. The galaxy colour is determined by the difference of the Galactic extinction corrected model magnitudes u and r, such that the colour is

$$C_{ur} = u_{model} - r_{model}$$

The non-detections are separated into blue (mostly star-forming) and red (mostly non-star forming) galaxies using the Baldry et al. (2004) optimal colour divider (Equation 1.1); galaxies with $u - r$ colour less than the colour divider are classified as blue,

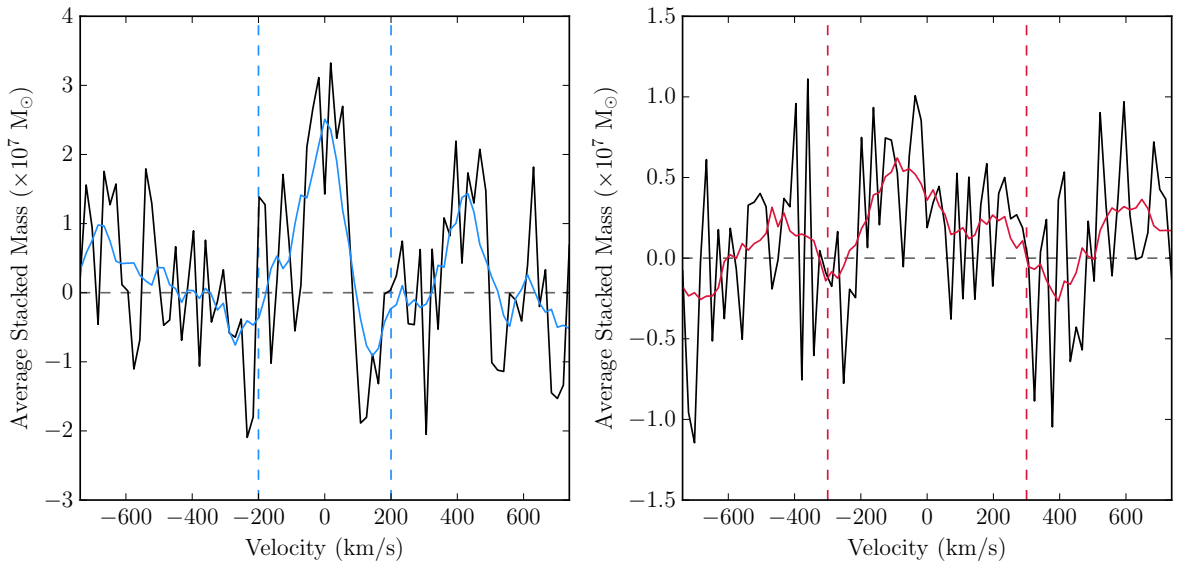


Figure 4.12: *The stacked spectra for the blue and red* HI *non-detected sample. The black spectra in each panel represent the unsmoothed spectra for each sample, while the coloured spectra represent the average spectrum after a boxcar smoothing algorithm has been applied. The dashed vertical lines represent the area over which the spectra are integrated to determine* $\langle M_{\rm HI} \rangle$.

and those with $u - r$ colour greater than the divider are classified as red. The distribution of the red and blue galaxies is shown on the colour-magnitude diagram in Figure 4.11.

Table 4.2: *Full width at half maximum of* $M_{\rm HI}$ *stacked spectra using on the NIBLES detections. These values are used to determine the number of channels to use when smoothing the stacked spectra arising from the non-detected sample.*

Stack Description	Red Sample ($\mathrm{km \cdot s^{-1}}$)		Blue Sample ($\mathrm{km \cdot s^{-1}}$)	
Full ND Sample ($\mathrm{M_{HI}}$)	317 (17 chan)		183 (11 chan)	
Full ND Sample ($f_{\rm HI}$)	180 (9 chan)		124 (7 chan)	
$\log \mathrm{M_\star} = (8, 9.5)\,\mathrm{M_\odot}$ ($\mathrm{M_{HI}}$)	182 (11 chan)	334 (19 chan)	144 (9 chan)	213 (11 chan)
& $\log \mathrm{M_\star} = (9.5, 12)\,\mathrm{M_\odot}$ ($f_{\rm HI}$)	159 (9 chan)	272 (15 chan)	121 (7 chan)	194 (11 chan)

Having separated the non-detections into blue and red sub-samples, there are 73 blue galaxies and 191 red galaxies. The stacked mass spectra are shown in Figure 4.12. Both the blue and red stacked spectra have low signal-to-noise (S/N) ratios. Techniques such as smoothing (e.g. Fabello et al. (2011a)) and re-binning (e.g. **?**) have been used to improve the S/N ratios of stacked spectra. In this work, the smoothing technique is applied to stacked spectra with low S/N using a boxcar kernel. The velocity width properties of the galaxies that make up each of the non-detected sub-samples are unknown, so similarly defined sub-samples of the detected HI spectra can provide a measure of the optimal kernel size. The smoothing kernel sizes are determined by the number of channels spanned by the width of the stacked spectra created from stacking sub-samples of the HI detections. The width of stacked spectra also provide a guide to an appropriate integration width to be used for the non-detected sample, especially when there is no clear signal in the stacked spectrum. The widths of HI detected sample's stacked spectra are listed in

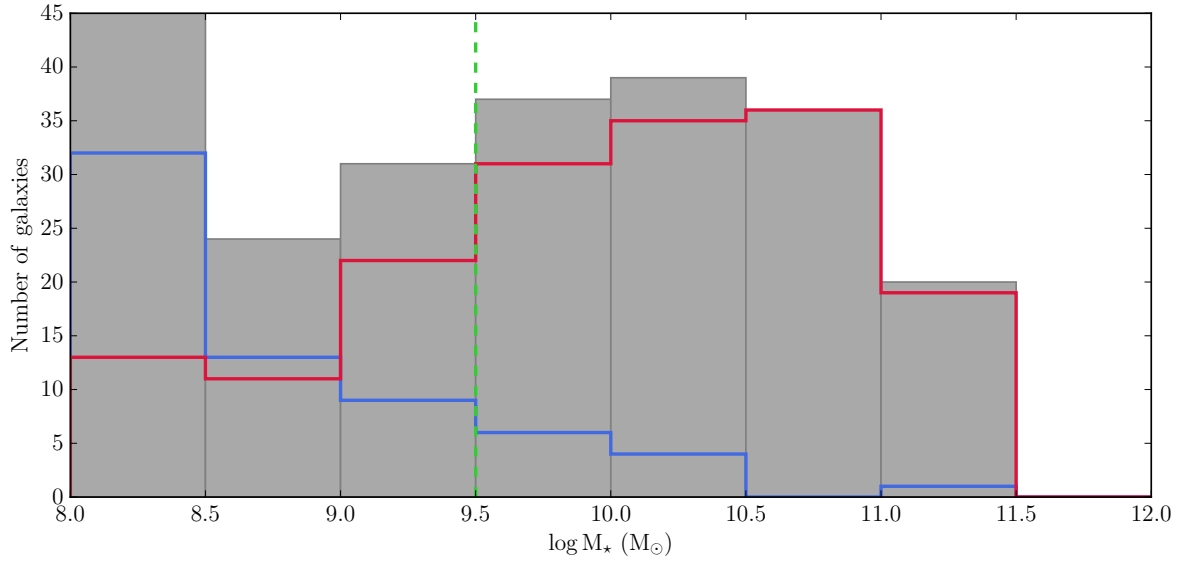


Figure 4.13: *The stellar mass distribution for all the non-detections. The grey bars indicate the number of galaxies per bin, while the blue and red lines indicate the number of blue and red galaxies per mass bin. Each bin is* 0.5 *dex.*

Table 4.2 along with the smoothing kernel sizes, and the stacked spectra are plotted in Appendix D.2.

After stacking each of the colour sub-samples, the blue and red stacked spectra were smoothed using kernel sizes of 11 and 17 channels respectively for the M_{HI} spectra, and 7 and 9 channels for the f_{HI} spectra. The stacked spectra are plotted in Appendix D.1.1. The average M_{HI} and f_{HI} extracted from the stacked spectra are listed in Table 4.3. The systematic uncertainties arise from the uncertainty of the galaxy's redshift, and the statistical uncertainty arises due to the inherent diversity of the galaxy sample.

Table 4.3: *Average properties of non-detected sample obtained from the stacked spectra. Results are from stacking a total of 264* HI *spectra.*

Sample	N	$\langle z \rangle$	$\langle M_\star \rangle$ ($M_\odot$)	$\langle M_{HI} \rangle$ ($M_\odot$)	$\langle f_{HI} \rangle$
All non-detections	264	0.01245	1.1×10^9	$1.059^{\pm0.015(sys)}_{\pm0.429(stat)} \times 10^8$	$0.0469^{\pm0.0012(sys)}_{\pm0.0434(stat)}$
Blue galaxies	73	0.01500	6.2×10^8	$1.384^{\pm0.038(sys)}_{\pm1.027(stat)} \times 10^8$	$0.1183^{\pm0.0059(sys)}_{\pm0.1433(stat)}$
Red galaxies	191	0.01469	9.9×10^9	$8.301^{\pm0.065(sys)}_{\pm3.585(stat)} \times 10^7$	$0.017190^{\pm0.000094(sys)}_{\pm0.010046(stat)}$

The HI-to-stellar mass fraction (gas fraction or f_{HI}) is a useful tool to probe the gas richness of a galaxy or sample of galaxies. Kannappan et al. (2013) proposed that in addition the "bimodality scale" in galaxy evolution, there is also a "gas-richness scale" which traces the change from gas-rich to gas-poor. This "gas-richness scale" which occurs at $M_\star \sim 10^{9.5}\,M_\odot$. Due to the unique selection of the NIBLES galaxies, there are HI spectra spanning both of these scales allowing the investigation of how the gas properties change for the different morphological types across both scales.

In order to investigate the gas properties of galaxies spanning the "gas-richness scale", the non-detected sample is split into two bins of stellar mass: $10^8 M_\odot < M_\star < 10^{9.5} M_\odot$ and $10^{9.5} M_\odot < M_\star < 10^{12} M_\odot$. Each stellar mass bin is also separated into blue and red galaxies. The six sub-samples were stacked in units of both M_{HI} and f_{HI} in order to obtain measurements of both the average gas mass and average gas fraction for the samples. The results from these stacks are summarised in Table 4.4. The stacked spectra from which each of the Table 4.4 values are calculated are plotted in Appendix D.1.2.

The stellar mass distribution for the non-detected sample is shown in Figure 4.13; the grey bars indicate the full non-detected sample, while the blue and red line represent the number of blue and red galaxies in each bin. The vertical dashed green line indicates the separation into low stellar mass and high stellar mass.

Table 4.4: *$\langle M_{HI} \rangle$ and $\langle f_{HI} \rangle$ from stacking the non-detections in two stellar mass bins. Each stellar mass bin has also been separated into blue and red sub-samples. Quoted with each of the quantities are the systematic and statistical errors.*

	N	$\langle M_{HI} \rangle$ ($M_\odot$)	$\langle M_\star \rangle$ ($M_\odot$)	$\langle f_{HI} \rangle$
$10^8 M_\odot < M_\star < 10^{9.5} M_\odot$				
All Non-detections	112	$5.17^{\pm 0.50(sys)}_{\pm 3.62(stat)} \times 10^7$	8.379×10^8	$0.1512^{\pm 0.0029(sys)}_{\pm 0.0756(stat)}$
Blue Sample	61	$7.79^{\pm 0.14(sys)}_{\pm 6.53(stat)} \times 10^7$	6.176×10^8	$0.2426^{\pm 0.0043(sys)}_{\pm 0.1403(stat)}$
Red Sample	51	$2.14^{\pm 0.16(sys)}_{\pm 1.47(stat)} \times 10^7$	1.097×10^9	$0.0405^{\pm 0.0015(sys)}_{\pm 0.0293(stat)}$
$10^{9.5} M_\odot < M_\star < 10^{12} M_\odot$				
All Non-detections	152	$1.806^{\pm 0.056(sys)}_{\pm 0.600(stat)} \times 10^8$	4.394×10^{10}	$0.007330^{\pm 0.000075(sys)}_{\pm 0.003602(stat)}$
Blue Sample	12	$5.697^{\pm 0.035(sys)}_{\pm 4.006(stat)} \times 10^8$	2.313×10^{10}	$0.02488^{\pm 0.00073(sys)}_{\pm 0.01901(stat)}$
Red Sample	140	$1.472^{\pm 0.059(sys)}_{\pm 0.559(stat)} \times 10^8$	4.584×10^{10}	$0.005842^{\pm 0.000054(sys)}_{\pm 0.003663(stat)}$

4.3 Stacking HI detected and non-detected NIBLES galaxies

Stacking only the detections or the non-detections can only provide a limit for that particular sample. In order to obtain a more representative average measurement for the sample using stacking, one needs to stack both detections and non-detections together. In Section 4.2.2 the non-detections were separated by colour and into two stellar mass bins, this was revisited using the full NIBLES stacking sample (non-detections and detections) in this section. The full sample of both Hi detections and non-detections with stellar masses of $M_\star > 10^8\, M_\odot$ contains 856 spectra.

The average HI mass for the full stacking sample (non-detections and detections)

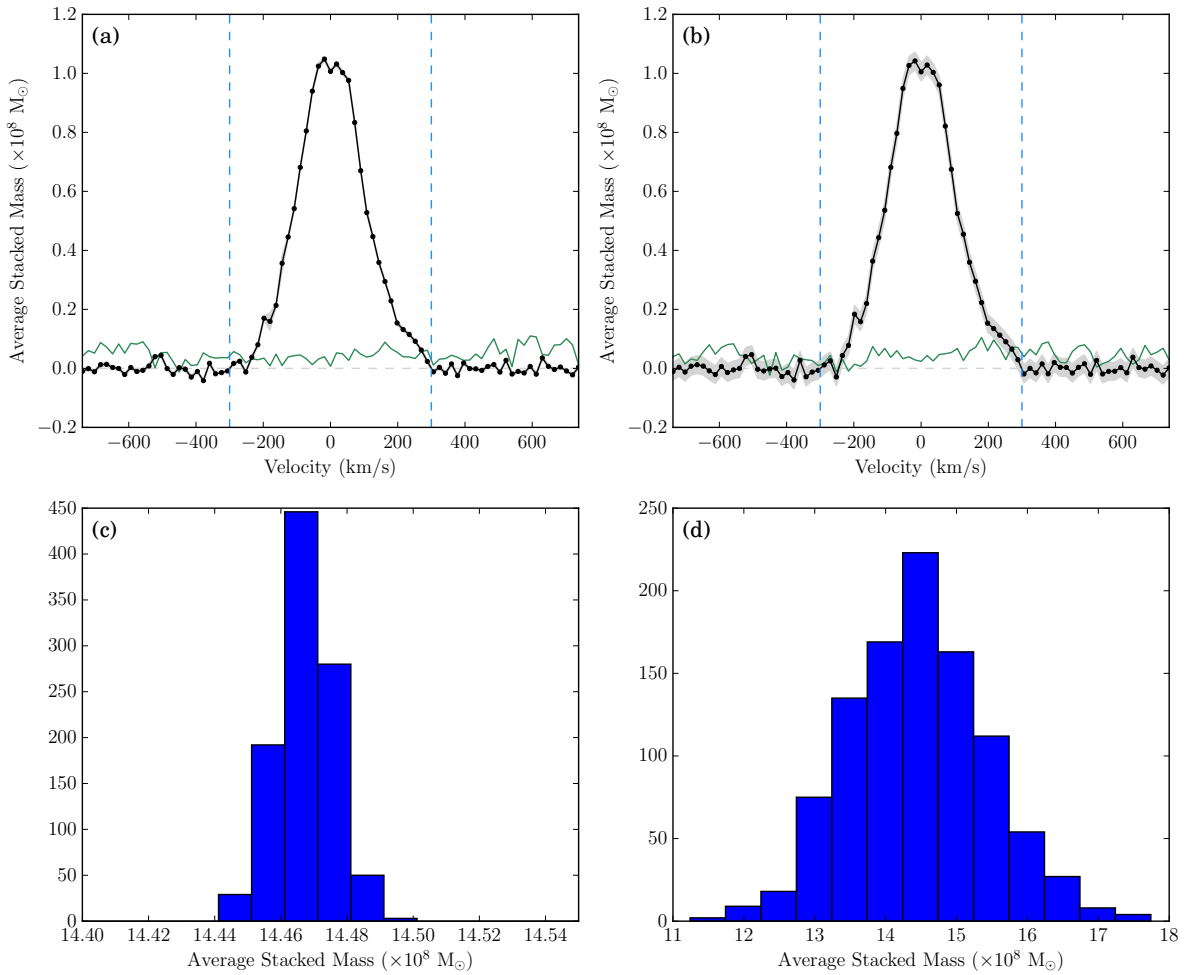


Figure 4.14: *Stacked spectra for all detections and non-detections. Panels* **a** *and* **b** *show the stacked spectra produced using the two methods of uncertainty calculation. Panel* **a** *shows the stacked spectrum produced while taking into account the redshift uncertainties, while panel* **b** *shows the spectrum produced using the DAGJK method of uncertainty calculation. The bottom two panels show the* $\langle M_{\rm HI} \rangle$ *measurements corresponding to the stacked spectra in the top row.*

determined from the stacked spectrum Figure 4.14 is

$$\langle \mathrm{M_{HI}} \rangle = (14.4675 \pm 0.0085(\mathrm{sys}) \pm 0.9789(\mathrm{stat})) \times 10^8\,\mathrm{M_\odot}$$

It is clear from the histograms of the average HI mass produced by HISS when calculating the statistical and systematic uncertainties in Figure 4.14, that the sample variance (quantified by the statistical uncertainty) is the dominant source of error on the HI mass measurement.

Table 4.5 contains the $\langle \mathrm{M_{HI}} \rangle$ and $\langle f_{\rm HI} \rangle$ calculated from stacking the full NIBLES sample in two stellar mass bins as well as further sub-dividing the stellar mass bins into blue and red samples (as was done for the non-detections in Section 4.2.2) . The $\langle \mathrm{M_{HI}} \rangle$ and $\langle f_{\rm HI} \rangle$ quantities are calculated from the stacked spectra for each bin by integrating the stacked flux (which is either in units of $\mathrm{M_{HI}}$ per channel or $f_{\rm HI}$ per channel) over different velocity widths which range from $400\,\mathrm{km \cdot s^{-1}}$ for the blue low stellar mass bin, and $600\,\mathrm{km \cdot s^{-1}}$ for the red high stellar mass bin. The distribution of the stellar masses for the full stacking sample is shown in Figure 4.15; the grey bars indicate all galaxies

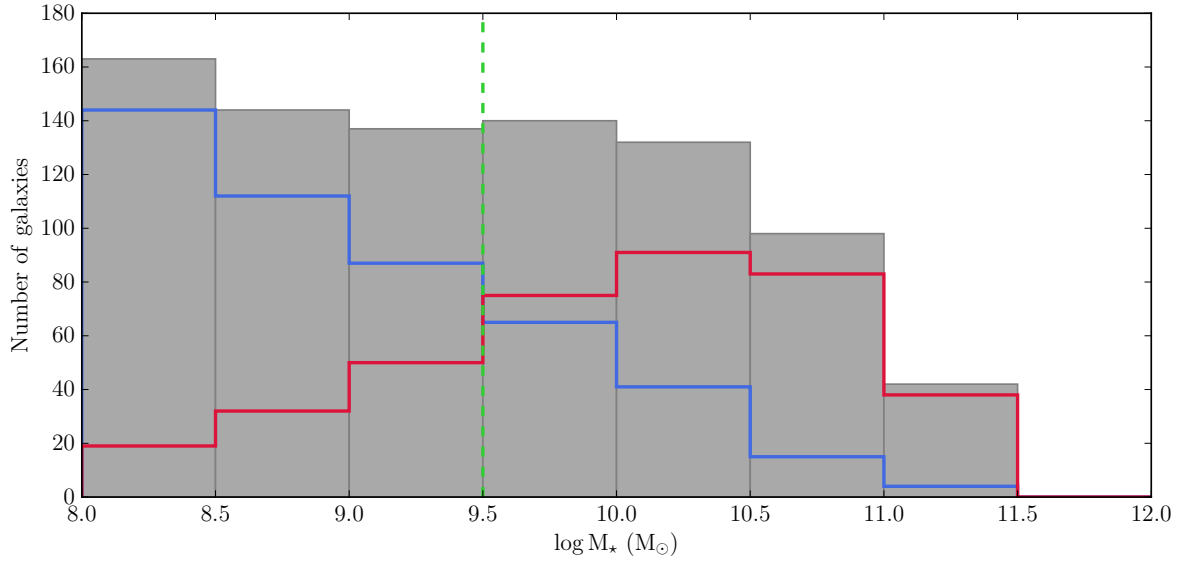


Figure 4.15: *The stellar mass distribution for the full NIBLES stacking catalogue. The grey bars indicate the number of galaxies per bin, while the blue and red lines indicate the number of blue and red galaxies per mass bin. Each bin is* 0.5 *dex.*

in the bin, while the blue and red lines indicate the number of blue and red galaxies per bin. The stacked spectra, from which the data in Table 4.5 are calculated, are plotted in Appendix D.3.1.

Table 4.5: *$\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ from stacking the detections and non-detections in two stellar mass bins. Each stellar mass bin has also been separated into blue and red sub-samples. Quoted with each of the quantities are the systematic and statistical errors.*

	N	$\langle \mathrm{M_{HI}}\rangle$ ($\mathrm{M_\odot}$)	$\langle \mathrm{M_\star}\rangle$ ($\mathrm{M_\odot}$)	$\langle f_{\rm HI}\rangle$
$10^8 \mathrm{M_\odot} < \mathrm{M_\star} < 10^{9.5} \mathrm{M_\odot}$				
All Spectra	444	$6.4750^{\pm0.0063(\mathrm{sys})}_{\pm0.4874(\mathrm{stat})} \times 10^8$	8.199×10^8	$1.2742^{\pm0.0017(\mathrm{sys})}_{\pm0.0986(\mathrm{stat})}$
Blue Sample	343	$6.9769^{\pm0.0091(\mathrm{sys})}_{\pm0.5328(\mathrm{stat})} \times 10^8$	7.206×10^8	$1.4802^{\pm0.0024(\mathrm{sys})}_{\pm0.1210(\mathrm{stat})}$
Red Sample	101	$4.7560^{\pm0.0089(\mathrm{sys})}_{\pm1.0393(\mathrm{stat})} \times 10^8$	1.157×10^9	$0.5710^{\pm0.0020(\mathrm{sys})}_{\pm0.1195(\mathrm{stat})}$
$10^{9.5} \mathrm{M_\odot} < \mathrm{M_\star} < 10^{12} \mathrm{M_\odot}$				
All Spectra	412	$2.3144^{\pm0.0017(\mathrm{sys})}_{\pm0.1888(\mathrm{stat})} \times 10^9$	3.641×10^{10}	$0.12790^{\pm0.00012(\mathrm{sys})}_{\pm0.01054(\mathrm{stat})}$
Blue Sample	125	$3.2933^{\pm0.0028(\mathrm{sys})}_{\pm0.3520(\mathrm{stat})} \times 10^9$	1.914×10^{10}	$0.26031^{\pm0.00032(\mathrm{sys})}_{\pm0.02410(\mathrm{stat})}$
Red Sample	287	$1.8874^{\pm0.0022(\mathrm{sys})}_{\pm0.2138(\mathrm{stat})} \times 10^9$	4.393×10^{10}	$0.070233^{\pm0.000074(\mathrm{sys})}_{\pm0.008104(\mathrm{stat})}$

4.3.1 Investigating the gas scaling relations with NIBLES

Gas scaling relations have been used in conjunction with stacking to study various galaxy properties (e.g. star formation, stellar mass, etc.) and their influence on the HI content of

the galaxies. Brown et al. (2015, B15) found that $NUV - r$ colour was the best tracer of HI content in galaxies with a secondary dependence on stellar mass. Using the NIBLES HI data and HISS to implement the stacking procedure, the B15 gas scaling relations can be extended by an order of magnitude lower in stellar mass.

Following what was done by B15, the full NIBLES sample (detections and non-detections) was separated into bins of stellar mass – the stellar mass distribution of the galaxies is shown in Figure 4.15. The sample is separated into blue and red sub-samples using the same Baldry et al. (2004) criteria as used in Section 4.2.2 to classify the non-detections. Each stellar mass bin is stacked in units of $f_{\rm HI}$ per channel using HISS, the resultant spectra are plotted in Appendix D.3.2. The measurements of the $f_{\rm HI}$ made from the stacked spectra are listed in Table 4.6.

The gas scaling relation that arises from the results in Table 4.6 are compared to the gas scaling relations found by B15, Catinella et al. (2010) and Fabello et al. (2011a) in Section 5.1.2.

Table 4.6: *$\langle f_{\rm HI} \rangle$ from stacking NIBLES spectra in bins of stellar mass. The full sample is also separated into blue and red sub-samples based on the Baldry criteria.*

		N	$\langle \mathrm{M}_\star \rangle$ ($\mathrm{M}_\odot$)	$\langle f_{\rm HI} \rangle$
Full Sample	$10^8 < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^9$	307	3.573×10^8	$1.5796^{\pm0.0026(\rm sys)}_{\pm0.1326(\rm stat)}$
	$10^9 < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{10}$	277	4.045×10^9	$0.40563^{\pm0.00056(\rm sys)}_{\pm0.03573(\rm stat)}$
	$10^{10} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{11}$	230	3.402×10^{10}	$0.086480^{\pm0.000071(\rm sys)}_{\pm0.008237(\rm stat)}$
	$10^{11} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{12}$	42	1.502×10^{11}	$0.03498^{\pm0.00010(\rm sys)}_{\pm0.0079(\rm stat)}$
Blue Sample	$10^8 < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^9$	256	3.452×10^8	$1.7325^{\pm0.0028(\rm sys)}_{\pm0.1527(\rm stat)}$
	$10^9 < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{10}$	152	3.689×10^9	$0.57476^{\pm0.00074(\rm sys)}_{\pm0.05253(\rm stat)}$
	$10^{10} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{11}$	56	2.536×10^{10}	$0.17320^{\pm0.00019(\rm sys)}_{\pm0.01954(\rm stat)}$
	$10^{11} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{12}$	4	1.426×10^{11}	$0.08234^{\pm0.00037(\rm sys)}_{\pm0.01448(\rm stat)}$
Red Sample	$10^8 < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^9$	51	4.181×10^8	$0.8179^{\pm0.0021(\rm sys)}_{\pm0.2241(\rm stat)}$
	$10^9 < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{10}$	125	4.478×10^9	$0.19967^{\pm0.00031(\rm sys)}_{\pm0.03788(\rm stat)}$
	$10^{10} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{11}$	174	3.681×10^{10}	$0.058630^{\pm0.000088(\rm sys)}_{\pm0.007273(\rm stat)}$
	$10^{11} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{12}$	38	1.510×10^{11}	$0.02898^{\pm0.00012(\rm sys)}_{\pm0.00780(\rm stat)}$

One of the main conclusions from B15 was that the gas fraction to stellar mass scaling relation is primarily driven by the correlation between HI gas and $NUV - r$ colour. The the NIBLES sample was separated into bins of 1 mag in $NUV - r$ colour in the range $1 <$ NUV-r (mag) < 6, each bin was then stacked in units of gas fraction per channel using HISS. The stacked spectra for each $NUV - r$ colour bin is plotted in Appendix D.3.3.

Figure 4.16 shows the stellar mass distribution with respect to the $NUV - r$ colour of the full NIBLES sample. The bars are coloured by $NUV - r$ colour (blue for NUV-r< 3.5 and red otherwise). The bars are 0.5 dex on the $\mathrm{M}_\star$ axis and 0.5 mag on the $NUV - r$ axis.

Table 4.7: *$\langle f_{\rm HI} \rangle$ from stacking NIBLES spectra in bins of $NUV - r$ colour.*

	N	$\langle \mathrm{M}_\star \rangle$ ($\mathrm{M}_\odot$)	$\langle f_{\rm HI} \rangle$
$1 <$ NUV-r (mag) < 2	165	1.224×10^9	$2.1160^{\pm 0.0035(\rm sys)}_{\pm 0.0109(\rm stat)}$
$2 <$ NUV-r (mag) < 3	300	7.113×10^9	$0.5881^{\pm 0.0014(\rm sys)}_{\pm 0.0037(\rm stat)}$
$3 <$ NUV-r (mag) < 4	135	3.098×10^{10}	$0.11429^{\pm 0.00064(\rm sys)}_{\pm 0.00101(\rm stat)}$
$4 <$ NUV-r (mag) < 5	104	2.715×10^{10}	$0.17220^{\pm 0.00058(\rm sys)}_{\pm 0.08996(\rm stat)}$
$5 <$ NUV-r (mag) < 6	100	3.819×10^{10}	$0.06958^{\pm 0.00071(\rm sys)}_{\pm 0.04639(\rm stat)}$

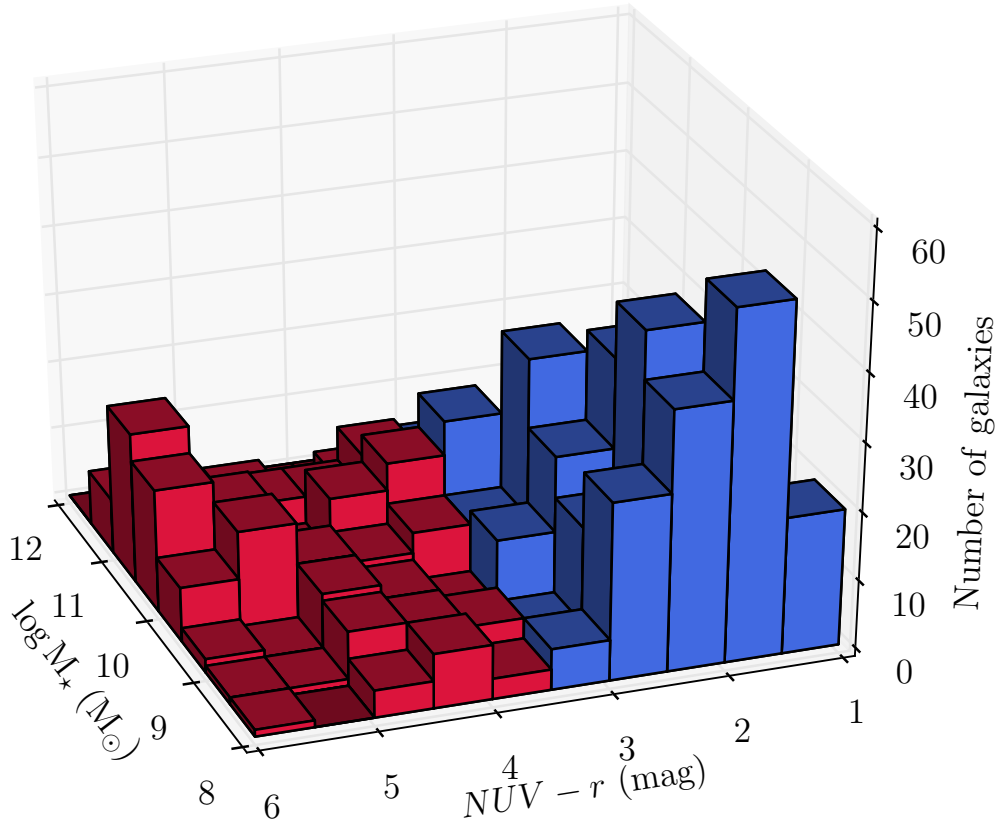


Figure 4.16: *Distribution of the NIBLES galaxies in $NUV - r$ colour and $M_\star$. The sources with NUV-r< 3.5 mag are coloured blue while those with NUV-r> 3.5 mag are coloured red.*

Chapter 5

Discussion and Outlook

This chapter discusses the results presented in Chapter 4, and provides an outlook for future work.

5.1 Discussion

5.1.1 Stacking HI detected and non-detected NIBLES galaxies

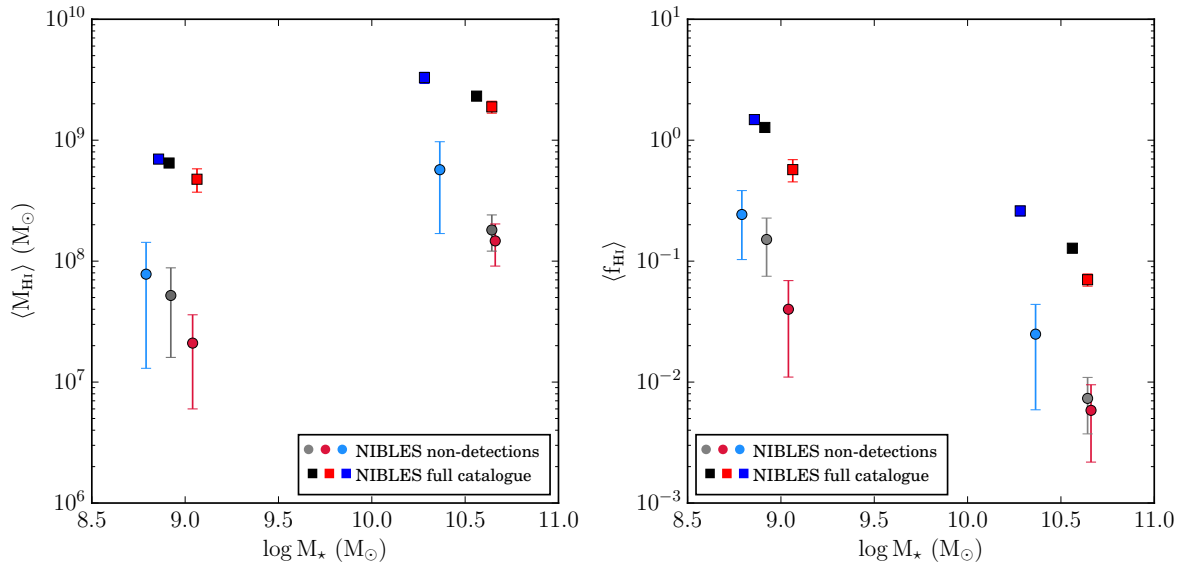


Figure 5.1: *$\langle M_{HI} \rangle$ (left) and $\langle f_{HI} \rangle$ (right) as function of $M_\star$ for the full NIBLES sample as well as the sub-sample of* HI *undetected galaxies. Each sample has been split into two stellar mass bins: $10^8 < M_\star\,(M_\odot) < 10^{9.5}$ and $10^{9.5} < M_\star\,(M_\odot) < 10^{12}$. Shown here are the mean $M_\star$ of the galaxies in each of the two stellar mass bins, for each of the various samples. The round data points represent the quantities obtained from stacking only the non-detections, and the square points represent the averages of the full sample. The blue, red and grey/black data points represent the blue, red and full sub-samples of the non-detected sample and full catalogue. The error bars represent the statistical uncertainties.*

Both the full sample of NIBLES galaxies, as well as the sub-sample of HI undetected galaxies, were separated into stellar mass sub-samples that span the gas-richness ($M_\star \sim 10^{9.5}\,M_\odot$) scale. Kannappan et al. (2013) claimed that galaxies with stellar masses

below the gas-richness scale are more gas-rich than galaxies with stellar masses above the threshold. This claim is supported by the $\langle f_{\rm HI} \rangle$ measurements made by stacking the NIBLES spectra in the two stellar mass bins ($10^8 < {\rm M}_\star\,({\rm M}_\odot) < 10^{9.5}$ and $10^{9.5} < {\rm M}_\star\,({\rm M}_\odot) < 10^{12}$). The galaxies in the low stellar mass bin have a lower $\langle {\rm M_{HI}} \rangle$ than those in the higher stellar mass bin. However, in terms of gas-richness, the low stellar mass galaxies have a higher $\langle f_{\rm HI} \rangle$ than the high stellar mass galaxies.

The square data points in Figure 5.1 represent the stacked spectra comprised of both detected and non-detected galaxies, while the round data points represent the stacked spectra of only non-detections. The ${\rm M}_\star$ component of each set of data points in Figure 5.1 represents the mean of each sample, rather than the median of the stellar mass bin, this was done to be more representative of the underlying stellar mass distribution of each sub-sample (the stellar mass distribution is shown in Figures 4.13 and 4.15)

The mean of a sample is typically representative of the mode. In Figure 5.1, this is clear by how the grey and black data points, which represent all the spectra in each stellar mass bin, are more tightly correlated to the blue data points in the low stellar mass bin and to the red data points in the high stellar mass bin. Given that the NIBLES sample was chosen in a manner that has it uniformly distributed in stellar mass (there was no *a priori* colour selection), it is possible to conclude that this phenomenon is representative of the underlying physics. Kannappan et al. (2013) point out that below the so-called bi-modality scale (${\rm M}_\star \sim 10^{10.5}\,{\rm M}_\odot$), galaxies are predominantly blue (star-forming) and above, are mostly red (non-star-forming).

Figure 5.1 also highlights the difference in stacking only non-detections compared to stacking all HI spectra. It is clear from both panels of Figure 5.1 that stacking only the non-detections provides a lower limit on the sample average determined by stacking both detections and non-detections together.

5.1.2 Investigating the gas scaling relations with NIBLES

Gas scaling relations have been used in conjunction with stacking to study various galaxy properties (e.g. star formation, stellar mass, etc.) and their influence on the HI content of the galaxies. Two gas scaling relations were studied using the NIBLES sample: the stellar mass scaling relation, and the $NUV - r$ scaling relation.

The full NIBLES sample was separated into stellar mass bins which had widths of 1 dex. For the NIBLES sample, the $NUV - r$ values in Figure 5.3 represent the mean value in each bin. The $\langle f_{\rm HI} \rangle$ for each stellar mass bin is shown by the round blue data points in the left panel of Figure 5.2. In Figure 5.2, the NIBLES gas fractions are compared to those from Catinella et al. (2010) and B15 which are a result of stacking spectra obtained using the Arecibo Telescope. The Arecibo multi-beam data has a known flux offset from NIBLES (see Section 4.1.3 and van Driel et al. (2016)), thus the Catinella et al. (2010) and B15 gas fractions are scaled by the ALFALFA/NRT flux ratio (1.45 ± 0.17). The right panel shows the NIBLES and B15 samples separated into red and blue populations. No direct comparison can be made between the B15 colour samples and the NIBLES colour samples as the two samples have been defined differently. The B15 red

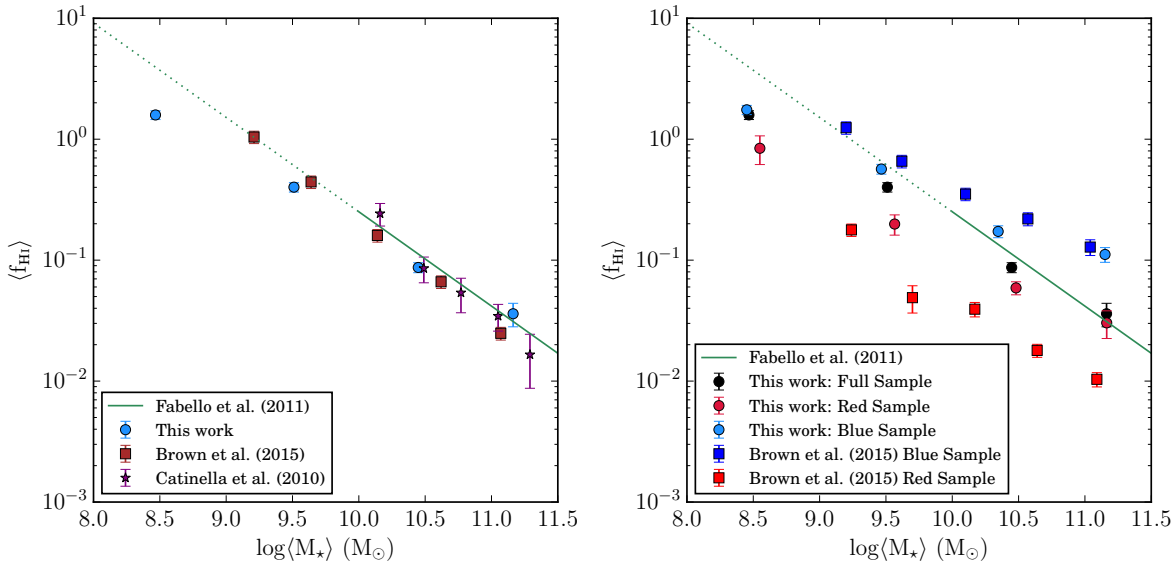


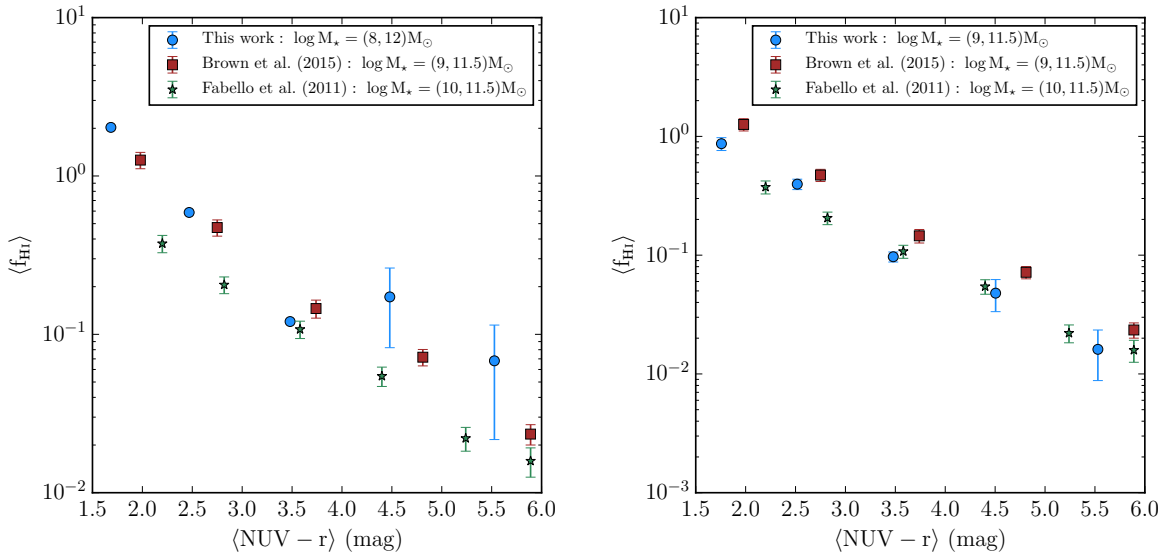
Figure 5.2: *Left: $\langle f_{\rm HI} \rangle$ as a function $M_\star$ for the NIBLES detections and non-detections. Plotted with the NIBLES results are $\langle f_{\rm HI} \rangle$ from B15 who stacked sub-samples of ALFALFA galaxies, and Catinella et al. (2010) who stacked sub-samples of GASS galaxies. The green line is taken Fabello et al. (2011a) who fitted the slope of the gas scaling relation for a sub-sample of ALFALFA galaxies, the solid line represents the stellar mass range of their sample. Right: The round data points represent that NIBLES sample, and the square data points represent the blue and red samples from B15. The errors on the NIBLES data points are the statistical uncertainties.*

sample is defined as $NUV - r$ in the range $5 <$ NUV-r (mag) < 8 and the blue sample is $1 <$ NUV-r (mag) < 3. The NIBLES red and blue samples are defined according to their $u - r$ colour as per the Baldry et al. (2004) colour divider. The galaxy samples that are classified by different colour cuts largely overlap (see Figure B.3) and the gas fractions therefore follow the same trends.

The NIBLES gas fractions agree well with adjusted gas fractions from B15, Fabello et al. (2011a) and Catinella et al. (2010) for $M_\star > 10^9\,M_\odot$ (left panel Figure 5.2). At the low mass end ($M_\star \sim 10^8\,M_\odot$), the NIBLES $\langle f_{\rm HI} \rangle$ diverges from the trend set by the high mass galaxies. The right panel of Figure 5.2 indicates that the gas fraction at $M_\star \sim 10^8\,M_\odot$ is dominated by the blue galaxies, however without more low stellar mass ($M_\star < 10^8\,M_\odot$) gas fractions, it is unclear if the divergence is due to a flattening of the gas scaling relation or simply an artefact of this dataset due to the optical selection criteria. However without access to the data used by Brown et al. (2015), it is difficult to fully investigate the differences in our results.

According to B15, the gas fraction to stellar mass scaling relation is primarily driven by the correlation between HI gas and $NUV - r$ colour. Thus, the NIBLES sample was separated into bins of $NUV - r$ colour 1 mag wide in the range $1 <$ NUV-r (mag) < 6.

The gas fraction as a function of the $NUV - r$ colour is plotted in Figure 5.3, the B15 and Fabello et al. (2011a) gas fractions have been scaled according to the ALFALFA/NRT flux ratio (see Section 4.1.3). Figure 5.3a shows the gas scaling relation for the entire NIBLES sample ($10^8 < M_\star\,(M_\odot) < 10^{12}$), while Figure 5.3b shows the same relation for the NIBLES galaxies with the same stellar masses as B15 ($10^9 < M_\star\,(M_\odot) < 10^{11.5}$): the



(a) *$\langle f_{\rm HI}\rangle$ as a function of $NUV-r$ colour for NIBLES galaxies with stellar masses in the range $10^8 < M_\star\,(M_\odot) < 10^{12}$.*

(b) *$\langle f_{\rm HI}\rangle$ as a function of $NUV-r$ colour for NIBLES galaxies with stellar masses in the range $10^9 < M_\star\,(M_\odot) < 10^{12}$.*

Figure 5.3: *Both panels show $\langle f_{\rm HI}\rangle$ as a function $M_\star$ for the full NIBLES sample.Plotted with the NIBLES results are $\langle f_{\rm HI}\rangle$ from B15 and Fabello et al. (2011a) who stacked sub-samples of ALFALFA galaxies. The statistical uncertainties for the NIBLES results are the errorbars in these plots.*

gas fraction decreases with redder colours (increasing $NUV-r$). It should be noted that the Fabello et al. (2011a) data span a stellar mass range of $10^{10} < \mathrm{M}_\star\,(\mathrm{M}_\odot) < 10^{11.5}$. In Figure 5.3a, there is clear agreement with B15 and Fabello et al. (2011a), however in Figure 5.3b, where the stellar mass range spanned by the NIBLES data is the same as B15, the NIBLES gas scaling relation agrees with Fabello et al. (2011a) but is systematically lower than B15 which could be due to different underlying stellar mass distributions in each $NUV-r$ colour bin.

5.2 Conclusion

This work has detailed the creation of a new HI stacking package in Python (HI Stacking Software: HISS). HISS takes the HI spectra for a sample of galaxies along with the accompanying galaxy catalogue containing the sample redshifts to produce an average HI spectrum from which a number of average properties (e.g. $\langle M_{HI} \rangle$, $\langle f_{HI} \rangle$) for the sample may be extracted. HISS also offers the user a choice of two built-in error analysis methods, and the ability to characterise the shape of the stacked spectrum through fitting a variety of functions.

HISS has been used to analyse HI spectra from the Nançay Interstellar Baryon Legacy Extragalactic Survey (NIBLES). NIBLES is an optically selected targeted HI galaxy survey of the local Universe using the 100 m Nançay Radio Telescope. The main aim of NIBLES was to study the HI properties of galaxies as a function of the stellar mass. The first aim of this work was to characterise the average HI mass of the galaxies for which there was no direct HI detection. The NIBLES spectra were separated into detected and non-detected samples using a quantitative method that classified spectra as detected based on criteria where the flux in a number of consecutive channels was above a certain threshold. Using the stacking technique implemented by HISS, the average HI mass for the non-detected sample was found to be

$$\langle M_{HI} \rangle = 9.24^{\pm 0.45(sys)}_{\pm 3.00(stat)} \times 10^7 \, M_\odot$$

which is two orders of magnitude below the 5σ detection limit for NIBLES. While this determination of the undetected HI mass allows us to study the HI in samples of galaxies below the survey detection threshold, the average mass of all non-detections does not provide much information about the sample which is why the sample was separated into sub-samples by stellar mass and colour for further investigation.

Combining the non-detections and the detections provides an opportunity to study the average gas properties of the entire sample. Due to the wide stellar mass range spanned by the NIBLES data ($10^8 < M_\star \, (M_\odot) < 10^{12}$), we were able to extend the previously studied stellar mass gas scaling relations (Brown et al., 2015; Fabello et al., 2011a; Catinella et al., 2010) by an order of magnitude lower in stellar mass. We find good agreement with previous results for high stellar mass, however we see hints of a possible flattening of the slope in the scaling relation at low stellar mass.

5.3 Outlook

Looking to the future, there are various aspects of this work that can be extended. For the HI Stacking Software, the current command line interface can (and will) be supplemented with a graphical user interface (GUI) to make the software more user-friendly. HISS is ready for release to the astronomy community pending the creation of a detailed user-friendly user manual. The software will be made public through GitHub as this platform has been designed to easily share open-source software while facilitating the tracking of software updates and allowing for other users to contribute to the software package.

Stacking the NIBLES spectra has provided insight into the gas content of galaxies over a wider range of stellar masses than has previously been studied. Since only the sources with reliable stellar masses and optical photometry were used, the optical photometry could be redone and stellar masses recalculated for the unreliable measurements

to increase sample statistics.

This work has shown that stacking is a powerful technique that can be successfully exploited to study the HI content in galaxies. Future surveys such as LADUMA on MeerKAT will be able to reduce analysis times by making use of data analysis tools such as the stacking software presented in this work. LADUMA has already identified stacking as the tool of choice for high redshift ($z > 0.5$) studies which means that HISS is poised to play an integral role in the upcoming HI stacking studies, particularly those involving LADUMA data.

References

Alam S., et al., 2015, Astrophys. J. Suppl. Ser., 219, 12

Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, Astrophys. J., 600, 681

Barnes D. G., et al., 2001, Mon. Not. R. Astron. Soc., 322, 486

Bell E., et al., 2004, Astrophys. J., 608, 752

Bianchi L., GALEX Team 2000, Mem. della Soc. Astron. Ital. Vol. 71, p.1117, 71, 1117

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, Astron. J., 136, 2846

Blyth S.-L., 2014, in 7th PHISCC Work. Deep HI with LADUMA Surv.. No. March. Cape Town

Blyth S. L., Baker A. J., Holwerda B. W., 2016, in MeerKAT Sci. Work.. Cape Town

Bouwens R. J., et al., 2012a, Astrophys. J., 752, L5

Bouwens R. J., et al., 2012b, Astrophys. J., 754, 83

Braun R., 2012, Astrophys. J., 749, 87

Brinchmann J., Charlot S., Heckman T. M., Kauffmann G., Tremonti C., White S. D. M., 2004, ArXiv eprints

Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, Mon. Not. R. Astron. Soc., 452, 2479

Catinella B., Cortese L., 2014, Mon. Not. R. Astron. Soc., 446, 3526

Catinella B., et al., 2010, Mon. Not. R. Astron. Soc., 403, 683

Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463

Chengalur J. N., Braun R., Wieringa M., 2001, Astron. Astrophys., 372, 768

Cucciati O., et al., 2011, Astron. . . . , 31, 22

Dahlen T., Mobasher B., Dickinson M., Ferguson H. C., Giavalisco M., Kretchmer C., Ravindranath S., 2007, Astrophys. J., 654, 172

Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, Mon. Not. R. Astron. Soc., 433, 1398

Elson E. C., Blyth S. L., Baker A. J., 2016, Mon. Not. R. Astron. Soc., p. stw1291

Fabello S., Catinella B., Giovanelli R., Kauffmann G., Haynes M. P., Heckman T. M., Schiminovich D., 2011a, Mon. Not. R. Astron. Soc., 411, 993

Fabello S., Kauffmann G., Catinella B., Giovanelli R., Haynes M. P., Heckman T. M., Schiminovich D., 2011b, Mon. Not. R. Astron. Soc., 416, 1739

Fabello S., Kauffmann G., Catinella B., Li C., Giovanelli R., Haynes M. P., 2012, Mon. Not. R. Astron. Soc., 427, 2841

Fernández X., et al., 2013, Astrophys. J., 770, L29

Fernández X., et al., 2016, Astrophys. J. Lett., 824, 7

Freudling W., et al., 2011, Astrophys. J., 727, 40

Giovanelli R., Haynes M. P., Kent B. R., 2005, Astron. J., p. 37

Gruppioni C., et al., 2013, Mon. Not. R. Astron. Soc., 432, 23

Haynes M. P., et al., 2011, Astron. J., 142, 170

Holwerda B. W., Blyth S.-L., Baker a. J., 2011, Proc. Int. Astron. Union, 7, 496

Hoppmann L., Staveley-Smith L., Freudling W., Zwaan M. A., Minchin R. F., Calabretta M. R., 2015, Mon. Not. R. Astron. Soc., 452, 3726

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

Jaffé Y. L., Poggianti B. M., Verheijen M. A. W., Deshev B. Z., van Gorkom J. H., 2013, Mon. Not. R. Astron. Soc., 431, 2111

Jaffé Y. L., et al., 2016, p. 21

Jones M. G., Haynes M. P., Giovanelli R., Papastergis E., 2015, arXiv:1510.03854 [astro-ph], p. 11

Joseph T., 2008, PhD thesis, University of Cape Town

Kannappan S. J., et al., 2013, Astrophys. J., 777, 42

Kauffmann G., White S. D. M., Heckman T. M., Menard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, Mon. Not. R. Astron. Soc. Vol. 353, Issue 3, pp. 713-731., 353, 713

Kennicutt R. C. J., 1989, Astrophys. J., 344, 685

Kennicutt, Jr. R. C., 1998, Astrophys. J., 498, 541

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, Mon. Not. R. Astron. Soc., 363, 2

Kott P. S., 2001, J. Off. Stat., 17, 521

Lah P., et al., 2007, Mon. Not. R. Astron. Soc., 376, 1357

Lah P., et al., 2009, Mon. Not. R. Astron. Soc., 399, 1447

Lehnert M. D., van Driel W., Minchin R., 2016, Astron. Astrophys., 590, 1

Leroy A. K., et al., 2012, Astron. J., 144, 3

Mac Low M.-M., Ferrara A., 1999, Astrophys. J., 513, 142

Madau P., Dickinson M., 2014, Annu. Rev. Astron. Astrophys., 52, 415

Maddox N., Hess K. M., Blyth S.-L., Jarvis M. J., 2013, Mon. Not. R. Astron. Soc., 433, 2613

Magnelli B., Elbaz D., Chary R., Dickinson M., Le Borgne D., Frayer D., Willmer C., 2011, Astron. Astrophys., 528, A35

Magnelli B., et al., 2013, Astron. Astrophys., 553, 1

Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, Astrophys. J., 723, 1359

Matthews L. D., van Driel W., 2000, Astron. Astrophys. Suppl. Ser., 143, 421

Meyer M. J., et al., 2004, The HIPASS catalogue - I. Data presentation (`arXiv:0406384`), doi:10.1111/j.1365-2966.2004.07710.x, `http://arxiv.org/abs/astro-ph/0406384`

Meyer S. A., Meyer M., Obreschkow D., Staveley-Smith L., 2016, Mon. Not. R. Astron. Soc., 455, 3136

Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, Astron. Astrophys., 505, 1087

Noterdaeme P., et al., 2012, Astron. Astrophys., 547, L1

O'Neil K., 2004, Astron. J., 128, 2080

Obreschkow D., Rawlings S., Road K., 2009a, Mon. Not. R. Astron. Soc., 394, 1857

Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009b, Astrophys. J., 698, 1467

Obreschkow D., Klockner H.-R., Heywood I., Levrier F., Rawlings S., 2009c, Astrophys. J., 703, 1890

Prochaska J. X., Wolfe A. M., 2009, Astrophys. J., 696, 1543

Prochaska J. X., et al., 2005, Astrophys. J., 635, 123

Putman M., Peek J., Joung M., 2012, Annu. Rev. Astron. Astrophys., 50, 491

Rao S. M., Turnshek D. A., Nestor D. B., 2006, Astrophys. J., 636, 610

Reddy N. A., Steidel C. C., 2009, Astrophys. J., 692, 778

Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, Mon. Not. R. Astron. Soc., 435, 2693

Rhee J., Lah P., Chengalur J. N., Briggs F. H., Colless M., 2016, Mon. Not. R. Astron. Soc., 12, 1

Robotham A. S. G., Driver S. P., 2011, Mon. Not. R. Astron. Soc., 413, 2570

Rodighiero G., et al., 2011, Cosmos, 4, 1

Sánchez-Ramírez R., et al., 2015, Mon. Not. R. Astron. Soc., 456, 4488

Sancisi R., Fraternali F., Oosterloo T., Van Der Hulst T., 2008, Astron. Astrophys. Rev., 15, 189

Sanders D. B., Mazzarella J. M., Kim D.-C., Surace J. A., Soifer B. T., 2003, Astron. J., 126, 1607

Schawinski K., et al., 2014, Mon. Not. R. Astron. Soc., 440, 889

Schechter P., 1976, Astrophys. J., 203, 203

Schenker M. A., et al., 2013, Astrophys. J., 768, 196

Schiminovich D., et al., 2005, Astrophys. J., 619, 47

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

Stoughton C., et al., 2002, Astron. J., 123, 485

Strateva I., Ivezic Z., Knapp G. R., Narayanan V. K., Strauss M. A., 2001,] 10.1086/323301

Takeuchi T. T., Yoshikawa K., Ishii T. T., 2003, Astrophys. J., 587, L89

Taylor E. N., et al., 2014, Mon. Not. R. Astron. Soc., 446, 2144

Verheijen M., van Gorkom J. H., Szomoru A., Dwarakanath K. S., Poggianti B. M., Schiminovich D., 2007, Astrophys. J., 668, L9

Walter F., Brinks E., Blok W. J. G. D., Bigiel F., Kennicutt R. C., Thornley M. D., Leroy A., 2008, Astron. J., 136, 2563

Westmeier T., Jurek R., Obreschkow D., Koribalski B. S., Staveley-Smith L., 2014, Mon. Not. R. Astron. Soc., 438, 1176

Wieringa M. H., de Bruyn A. G., Katgert P., 1992, Astron. Astrophys. (ISSN 0004-6361), vol. 256, no. 2, p. 331-342., 256, 331

Wong O. I., et al., 2006, Mon. Not. R. Astron. Soc., 371, 1855

Wyder T. K., et al., 2005, Astrophys. J., 619, L15

Wyder T. K., et al., 2007, Astrophys. J. Suppl. Ser., 173, 293

York D. G., et al., 2000, Astron. J., 120, 1579

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, Astron. Astrophys., 556, A141

Zwaan M. A., Briggs F. H., Sprayberry D., Sorar E., 1997, Astrophys. J., 490, 173

Zwaan M., van Dokkum P., Verheijen M., Briggs F., 2001, Gas Galaxy Evol., 240, 640

Zwaan M. A., et al., 2003, Astron. J., 125, 2842

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, Mon. Not. R. Astron. Soc. Lett., 359, L30

de Blok W. J. G., Fraternali F., Heald G., Adams E. A. K., Bosma A., Koribalski B. S., 2015, Adv. Astrophys. with Sq. Km. Array, pp 1 – 22

van Driel W., et al., 2009, "Proceedings Panor. Radio Astron. Wide-f. 1-2 GHz Res. galaxy Evol. June 2-5 2009. Groningen

van Driel W., et al., 2016, Astron. Astrophys., p. 71

van der Hulst J. M., van Albada T. S., Sancisi R., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H., eds, Astron. Soc. Pacific Conf. Ser.. p. 451, `http://adsabs.harvard.edu/abs/2001ASPC..240..451V`

Appendices

Appendix A

Derivation of the Significance Calculation

To calculate the p-value and significance of a stacked spectrum, the following method is implemented in HISS:

1. Calculate the χ^2 value for a single Gaussian fit to the data and the null hypothesis (a horizontal line at approximately 0) using Equation A.1.
2. Calculate the log likelihood statistic[1]: $D = -2\ln(\Lambda)(= \chi^2_{null} - \chi^2_g)$
3. Compare D to the χ^2-distribution with the appropriate degrees of freedom, which in this case is 3 (the single Gaussian has 3 free parameters and the null hypothesis is a horizontal line with no free parameters).
4. Use Equation A.4 with D, the aforementioned log-likelihood statistic, and $ndf = 3$ to determine the p-value and then Equation A.11 to calculate the corresponding significance level.

Equation A.4 and Equation A.11 are implemented using built in functions from the SciPy library (that have been sanity-checked).

A.1 χ^2-distribution

The first step to determining the significance of a detection is to determine the χ^2 statistic for the expected shape of the detection (in this case, a single Gaussian), as well as for the null hypothesis (a horizontal line at 0). The χ^2 of a fit can be determined by:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{A.1}$$

where O_i is the observed value and E_i is the expected value. The p-value for a particular χ^2 statistic is calculated by integrating the χ^2 probability density function from the observed χ^2 value to infinity. This is shown for a χ^2 value from a fit with 3 degrees of freedom in Figure A.1, the p-value is highlighted by the shaded green area below the curve. The probability density for the χ^2-distribution is given by:

$$f_k(x) = \frac{1}{2\Gamma(\frac{k}{2})} \left(\frac{x}{2}\right)^{\frac{k}{2}-1} \mathrm{e}^{-x/2} \tag{A.2}$$

[1] `http://en.wikipedia.org/wiki/Likelihood-ratio_test`

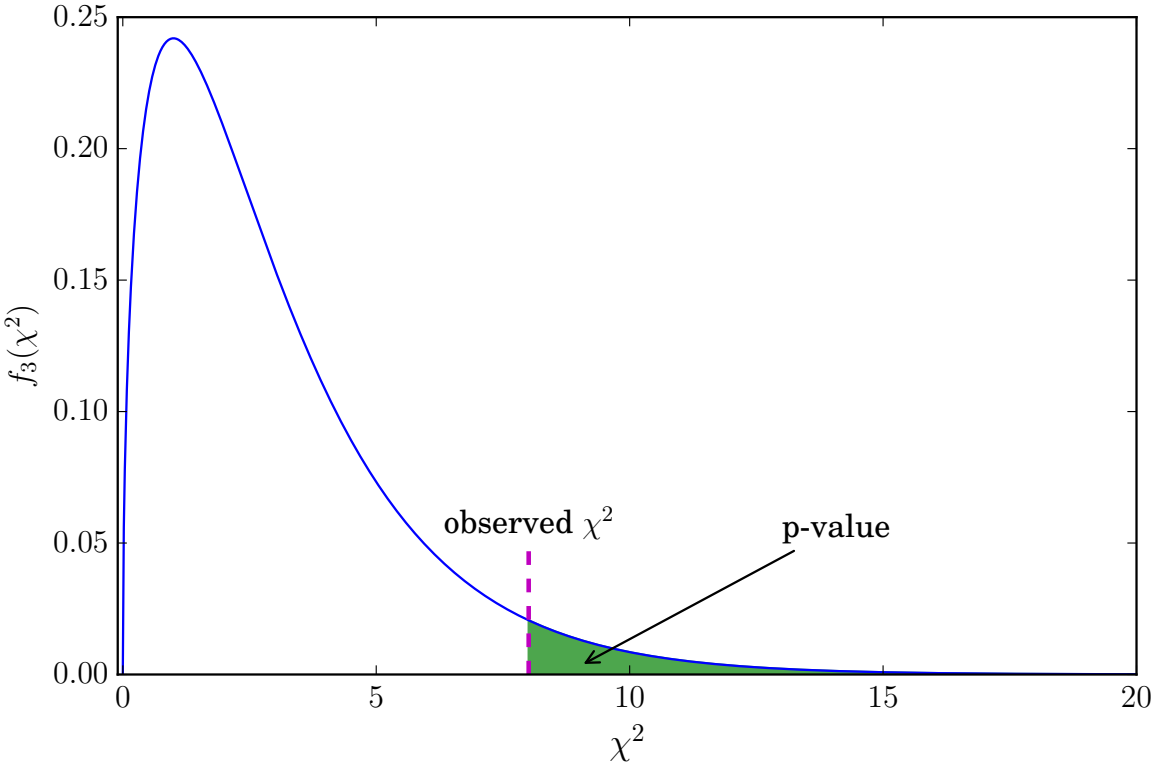


Figure A.1: *Probability distribution for χ^2-distribution with 3 degrees of freedom (this is given by Equation A.2). For an observed χ^2 value one can determine the associated p-value by integrating the the χ^2-distribution from that particular χ^2 value to infinity.*

where k is number of degrees of freedom[2].

A.2 p-values

The p-value is useful in the context of determining the significance of a result relative to a null hypothesis. The p-value is the probability of obtaining a result equal to the null hypothesis (in rough terms)[3]. Calculating a p-value: it is the area below the χ^2-distribution to the right of the oberved χ^2 value (this is shown by the green shaded area in Figure A.1).

The area under the χ^2-distribution is given by the cumulative distribution function

$$\begin{aligned} F(x;k) &= \int_0^x \frac{t^{\frac{r}{2}-1}\mathrm{e}^{-t/2}dt}{\Gamma(\frac{r}{2})2^{r/2}} \\ &= \frac{\gamma\left(\frac{r}{2},\frac{x}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \\ &= P\left(\frac{k}{2},\frac{x}{2}\right) \end{aligned} \tag{A.3}$$

where $\gamma(a,x)$ is the lower incomplete gamma function and $P(a,x)$ the regularised (normalised incomplete) gamma function. However Equation A.3 will give the area to the left

[2] `https://en.wikipedia.org/wiki/Chi-squared_distribution`
[3] `https://en.wikipedia.org/wiki/P-value`

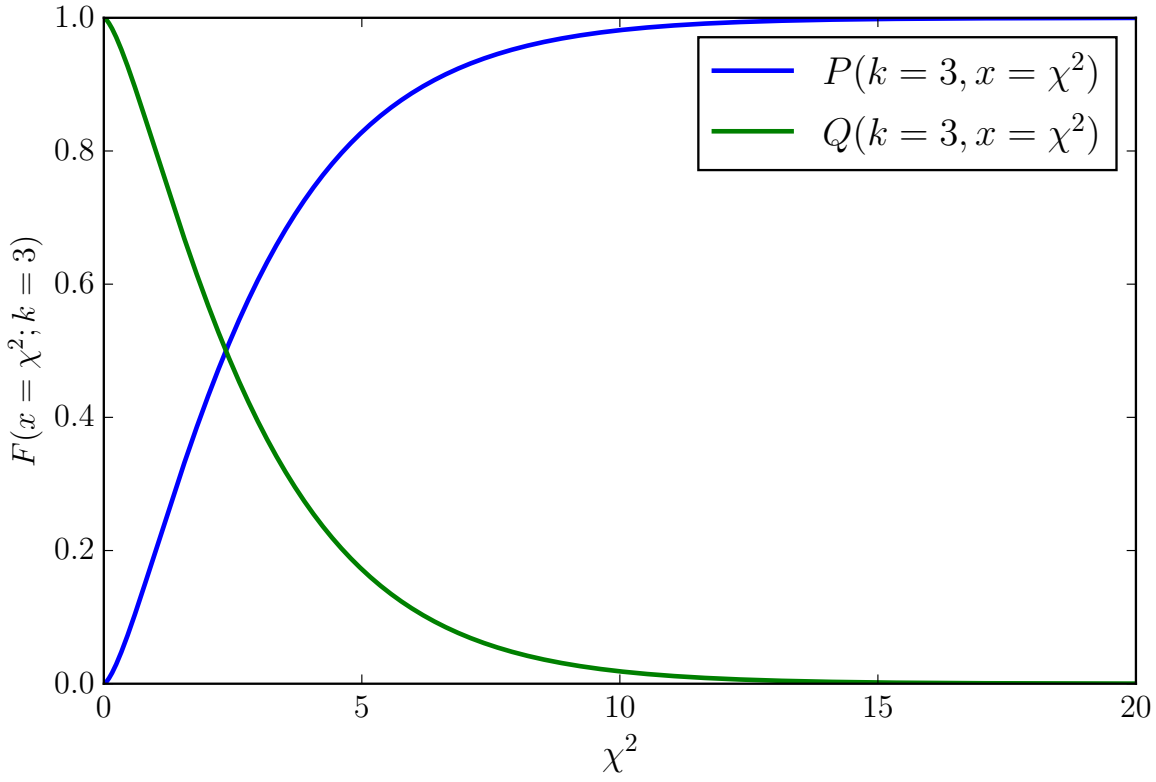


Figure A.2: *The cumulative distribution function (and its complement) – $P(\frac{k}{2}, \frac{x}{2})$ and $Q(\frac{k}{2}, \frac{x}{2})$ – for the χ^2-distribution with 3 degrees of freedom.*

of the observed χ^2 value, to determine the area to the right of the observed χ^2 value, we use the complement of Equation A.3 – this will also give the p-value:

$$\begin{aligned} p &= 1 - P\left(\frac{ndf}{2}, \frac{\chi^2}{2}\right) \\ &= Q\left(\frac{ndf}{2}, \frac{\chi^2}{2}\right) \end{aligned} \tag{A.4}$$

where *ndf* is the number of degrees of freedom. Figure A.2 shows both $P(\frac{k}{2}, \frac{x}{2})$ and $Q(\frac{k}{2}, \frac{x}{2})$.

A.3 Significance

So the p-value will tell us about the probability of more extreme result (than what is observed) asumming that the null hypothesis is true, but it is the significance of the observed result that is more interesting and meaningful. To calculate the significance, we use Equation A.3 which will give us the area to the left of the observed χ^2 value – this area can be equated to the area beneath a gaussian encapsulated by $(\mu - n\sigma, \mu + n\sigma)$.

The probability density function for the standard $(\sigma = 1, \mu = 0)$ normal distribution[4] (plotted in Figure A.3) is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \tag{A.5}$$

[4] `https://en.wikipedia.org/wiki/Normal_distribution`

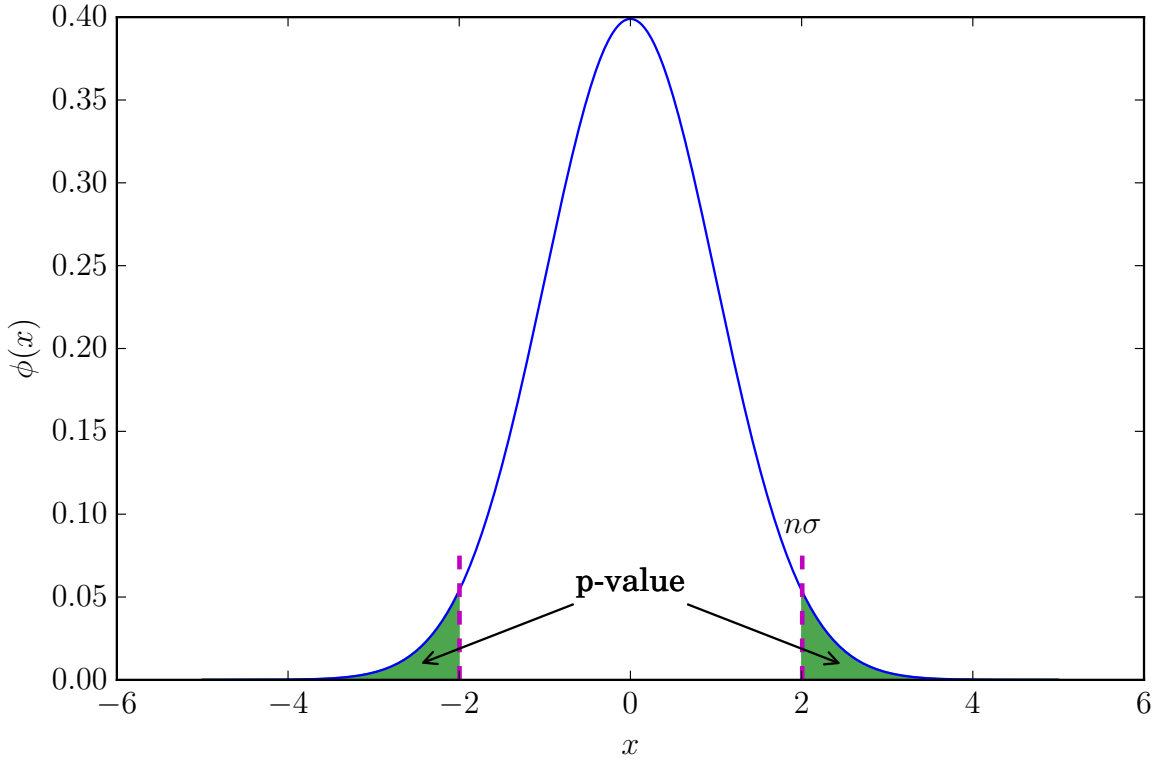


Figure A.3: *Plotted above is a probability density function for the standard normal distribution. The equivalent p-value shown in Figure A.1 is shown above in shaded green. The significance is determined by solving for n.*

And the corresponding cumulative distribution function (area under the graph) is given by,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt \tag{A.6}$$

The error function is defined as the probability of a random variable with a normal distribution of mean 0 and a variance of $\frac{1}{2}$ and in the range $[-x, x]$,

$$\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \mathrm{e}^{-t^2} dt \tag{A.7}$$

Equation A.6 and Equation A.7 can be related by:

$$\Phi(x) = \frac{1}{2} \left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}} \right) \right] \tag{A.8}$$

For the general normal distribution with mean μ and variance σ^2:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma} \right) = \frac{1}{2} \left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}} \right) \right] \tag{A.9}$$

To find the area of the unshaded region, we want the probability in the region $(\mu -$

$n\sigma, \mu + n\sigma)$.

$$\begin{aligned} \text{Area} &= F(\mu + n\sigma) - F(\mu - n\sigma) \\ &= \Phi(n) - \Phi(-n) \text{ given that } \sigma = 1 \text{ and } \mu = 0 \\ &= \text{erf}\left(\frac{n}{\sqrt{2}}\right) \end{aligned} \tag{A.10}$$

To find the significance, n, we need to solve Equation A.10 for n given that the area is $P\left(\frac{ndf}{2}, \frac{\chi^2}{2}\right)$:

$$n = \sqrt{2} \times \text{erf}^{-1}\left[P\left(\frac{ndf}{2}, \frac{\chi^2}{2}\right)\right] \tag{A.11}$$

where erf^{-1} is the inverse error function.

Appendix B

Extra Figures

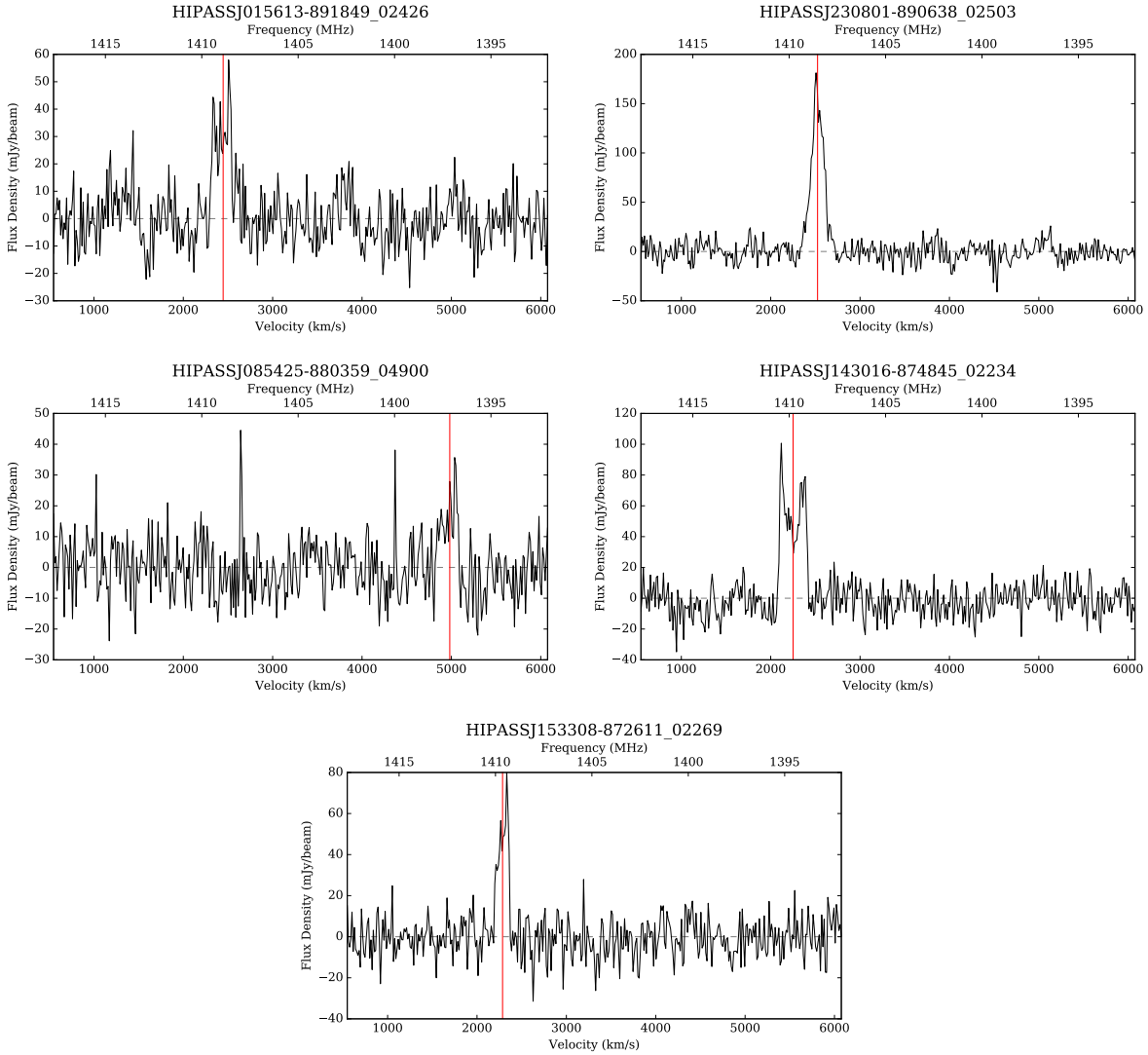


Figure B.1: *The five HIPASS Spectra used in HISS package test. Both frequency and velocity spectral axes are shown on each plot. The red vertical line indicates the central velocity of each galaxy obtained from the HICAT catalogue (Meyer et al., 2004).*

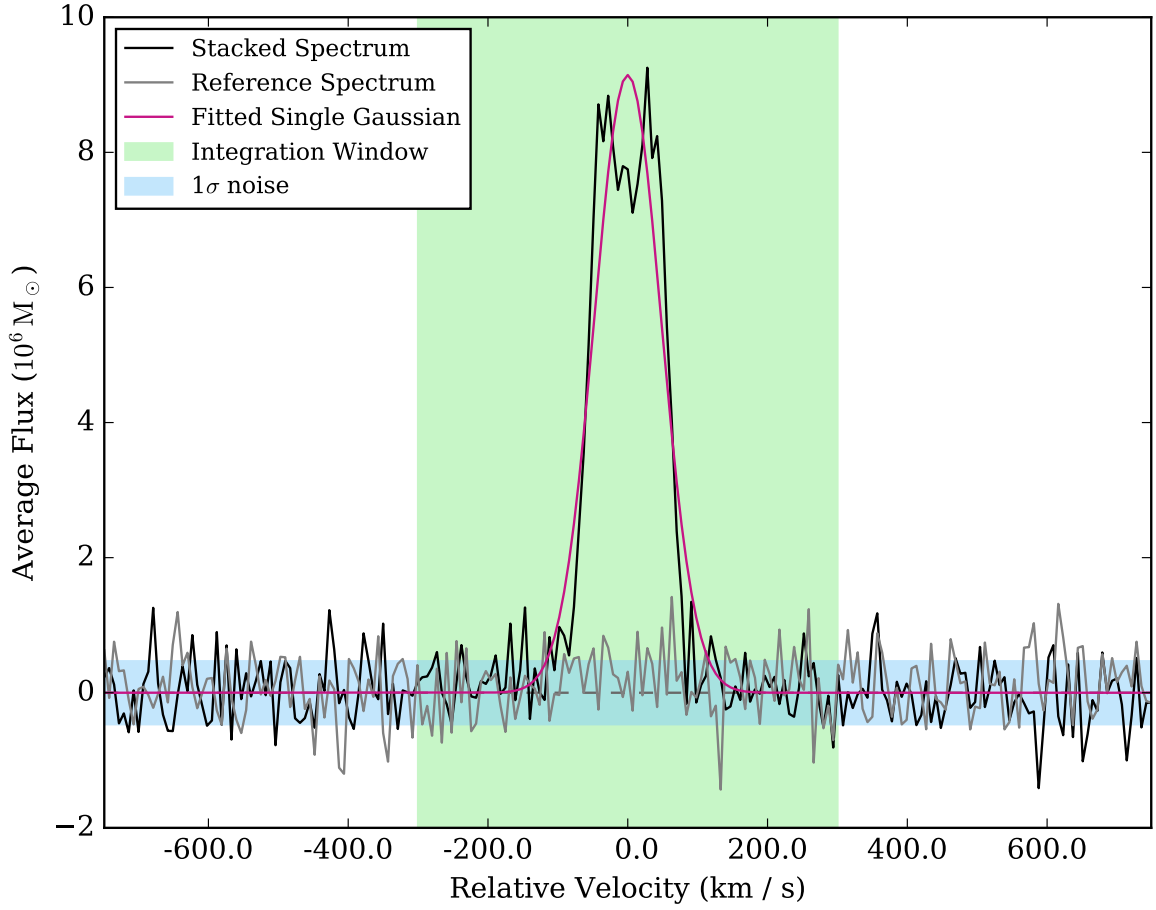


Figure B.2: *Diagnostic Plot 1: this is the first plot of the stacked spectrum displayed to the user.*

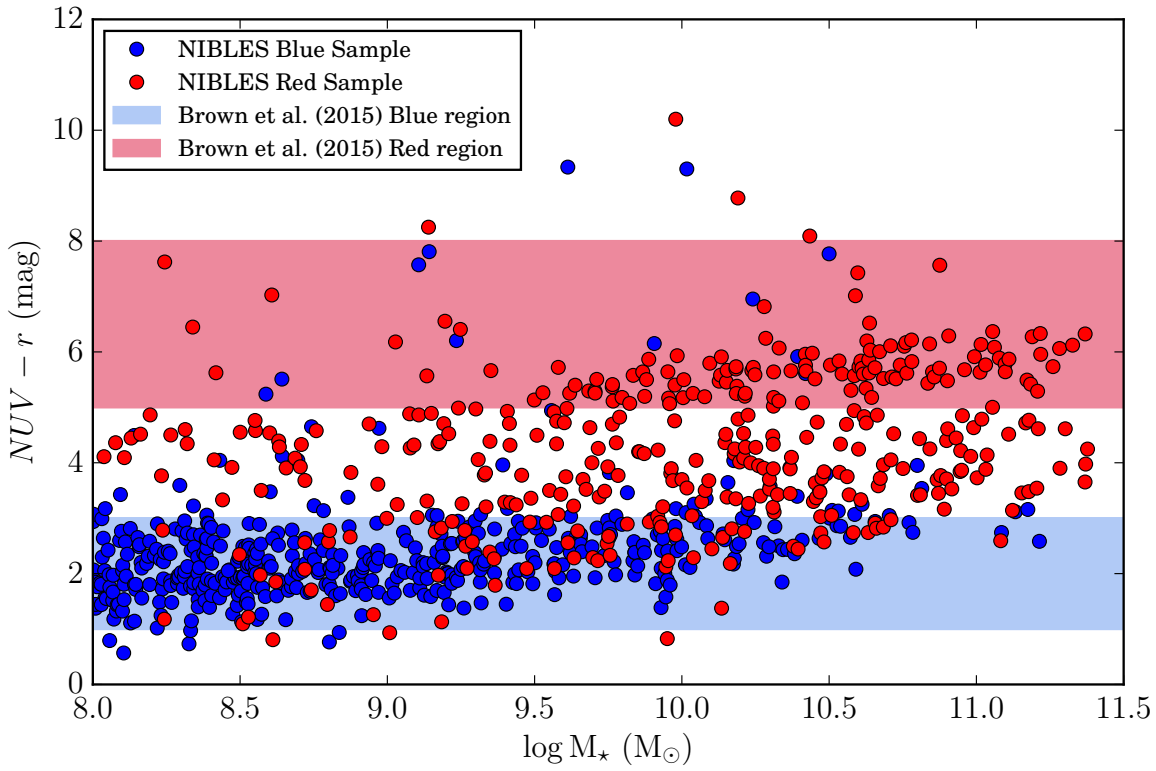


Figure B.3: *$NUV - r$ colour for the NIBLES sample plotted as a function of the stellar mass. The data points for each source are coloured blue or red according to the Baldry colour classification. The shaded regions indicate the $NUV - r$ magnitude range that Brown et al. (2015) classify as blue and red.*

Appendix C

NIBLES Stacking Sample

C.1 NIBLES Non-detected HI Sources

Table C.1: *CRUMBS Data*

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			$\rm km \cdot s^{-1}$	mag	mag	mag	$M_\odot$
1818	03 01 11.90	00 18 10.50	PGC 1162351	2779 ± 2	1.24	2.44	−16.06	7.94
1926	11 23 18.84	03 57 18.78	CGCG 039-161	1582 ± 19	1.91	3.76	−16.18	8.23
1951	11 47 06.99	03 06 23.09	ASK 74477	1016 ± 2	1.28	2.27	−13.51	6.71
1995	12 08 24.18	03 00 47.77	ASK 075814	880 ± 2	1.30	2.38	−12.73	6.58
2012	12 19 21.12	04 46 24.49	ASK 173829	2065 ± 2	1.61	3.43	−15.87	8.09
2025	12 23 09.95	00 25 37.42	PGC 1165632	2420 ± 1	0.98	1.68	−15.12	7.32
2244	14 04 15.84	04 06 43.85	UGC 08986	1239 ± 3	2.19	4.88	−17.79	9.08
2248	14 11 24.01	−03 10 02.51	ASK 199473	1628 ± 1	0.99	1.85	−14.34	7.04
2256	14 17 07.45	04 50 13.39	ASK 99987	1643 ± 1	1.07	1.64	−14.37	7.09
2272	14 25 08.99	−01 06 48.52	CGCG 019-019	2643 ± 3	2.11	4.53	−17.94	9.21
2274	14 28 08.66	01 49 25.61	ASK 082514	1728 ± 29	1.40	2.54	−14.56	6.95
2344	14 50 59.86	02 20 16.40	ASK 84013	1527 ± 27	2.12	3.95	−14.07	7.47
2445	15 24 50.12	03 04 53.25	SHOC 505	1753 ± 1	0.97	1.81	−15.12	7.56
2463	23 30 04.82	00 33 40.11	PGC 1140348	2452 ± 2	1.59	2.97	−15.11	7.77
2525	09 35 44.04	31 42 19.68	NGC 2918	6756 ± 2	2.65	6.06	−22.64	11.28
2535	14 30 25.58	11 55 40.77	NGC 5644	7649 ± 2	2.72	6.12	−22.87	11.33
2541	13 53 38.43	36 08 02.57	NGC 5352	7956 ± 3	2.81	5.73	−22.64	11.26
2565	00 35 59.83	−10 07 18.07	NGC 0163	5917 ± 4	2.70	5.63	−22.05	11.01
2574	11 00 35.40	12 09 41.65	NGC 3491	6351 ± 2	2.72	6.37	−21.98	11.06
2589	12 19 32.09	49 48 56.71	UGC 07367	4093 ± 2	2.60	5.83	−21.45	10.78
2597	12 19 32.59	56 44 11.69	NGC 4271	4756 ± 2	2.82	–	−21.84	10.93
2614	12 12 17.27	13 12 18.72	NGC 4168	2273 ± 2	2.44	5.34	−21.43	10.63
2615	12 32 47.65	63 56 21.18	NGC 4512	2516 ± 1	2.72	5.73	−20.75	10.60
2661	10 36 38.44	14 10 15.97	NGC 3300	3017 ± 1	2.54	5.84	−21.02	10.53
2675	12 08 09.62	10 22 43.96	NGC 4119	1651 ± 2	2.28	4.04	−19.74	10.05
2682	12 27 13.34	12 44 05.22	NGC 4425	1891 ± 1	2.39	5.25	−20.42	10.22
2688	12 24 05.01	11 13 05.04	NGC 4352	2083 ± 1	2.36	5.19	−20.12	10.08
2710	12 27 02.54	15 27 41.34	NGC 4421	1551 ± 1	2.58	5.25	−20.15	10.04
2715	14 05 12.42	55 44 30.67	NGC 5475	1647 ± 1	2.54	5.40	−19.58	9.98
2742	12 15 59.87	66 13 51.00	NGC 4221	1314 ± 1	2.51	5.58	−19.48	9.83
2743	10 12 41.25	03 07 45.79	NGC 3156	1266 ± 29	1.99	4.32	−19.14	9.42
2744	11 06 32.10	11 23 07.51	NGC 3524	1357 ± 1	2.36	5.26	−19.08	9.71
2756	11 08 40.38	57 13 48.71	NGC 3530	1876 ± 1	2.42	–	−18.95	9.69
2769	12 25 18.78	64 56 00.52	NGC 4391	1320 ± 1	2.52	5.72	−18.73	9.58
2770	13 00 10.57	12 28 59.92	NGC 4880	1362 ± 3	2.24	4.50	−19.29	9.51

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			km · s^{-1}	mag	mag	mag	$M_\odot$
2775	12 22 04.11	12 47 14.95	NGC 4306	1520 ± 3	2.21	4.39	−18.70	9.35
2789	11 21 24.99	03 00 50.23	NGC 3643	1742 ± 2	2.28	5.26	−18.67	9.53
2790	12 38 17.87	13 06 35.64	NGC 4584	1715 ± 3	2.01	3.78	−18.96	9.46
2795	12 41 59.35	12 56 34.27	NGC 4620	1125 ± 3	2.09	4.99	−18.63	9.24
2798	15 06 35.03	02 00 18.38	NGC 5841	1256 ± 1	2.34	5.57	−18.19	9.13
2805	12 48 55.36	14 54 28.40	IC 3806	1401 ± 5	2.14	4.06	−18.21	9.31
2831	12 32 14.21	10 15 05.18	IC 3468	1300 ± 4	2.18	4.89	−18.52	9.15
2890	12 11 07.76	14 16 29.33	IC 3032	1184 ± 5	2.02	4.58	−16.83	8.55
2891	13 55 38.75	04 59 06.18	NGC 5360	1180 ± 3	1.72	2.73	−16.84	8.32
2899	12 17 01.11	09 24 27.19	IC 3097	1253 ± 7	1.94	4.39	−16.88	8.63
2906	12 16 52.36	14 30 52.50	IC 3096	1264 ± 3	1.84	3.50	−17.03	8.57
2910	12 10 12.57	64 45 38.33	MCG +11-15-032	1349 ± 5	1.89	3.33	−16.76	8.44
2914	12 33 13.43	09 23 50.50	IC 3487	1074 ± 4	1.78	3.22	−16.45	8.34
2924	12 41 46.71	11 29 18.20	IC 3665	1227 ± 29	2.18	3.89	−16.73	–
2961	12 32 39.01	10 05 31.46	VCC 1446	1196 ± 13	2.00	4.86	−15.83	8.19
2962	12 12 18.99	15 28 59.14	VCC 0050	1209 ± 12	1.85	4.45	−15.87	8.13
2975	11 35 18.14	58 53 18.78	MCG +10-17-017	1037 ± 2	1.71	2.87	−15.41	7.92
2979	12 16 02.22	46 43 58.32	MCG +08-22-086	1061 ± 7	1.96	4.11	−15.62	8.04
2984	12 52 33.75	−01 43 48.75	ASK 14845	1129 ± 3	1.74	3.17	−15.77	8.04
2985	12 22 50.85	13 35 34.31	VCC 0592	1102 ± 8	2.09	4.36	−15.43	8.08
2996	12 24 05.26	10 04 03.89	PGC 40315	1370 ± 29	2.15	–	−16.05	8.13
3009	12 02 55.54	55 49 05.68	ASK 239047	1034 ± 12	1.86	3.82	−14.98	7.78
3011	12 18 40.14	45 54 34.96	ASK 319587	1036 ± 8	1.77	3.84	−14.61	7.61
3015	12 08 10.74	55 44 46.81	PGC 2512985	1115 ± 2	1.50	2.91	−15.13	7.61
3023	12 15 08.55	14 58 18.78	VCC 0137	1152 ± 7	1.96	–	−14.73	7.72
3030	12 45 35.27	01 59 19.21	ASK 77777	1169 ± 29	2.06	–	−14.73	7.78
3031	12 00 02.44	42 47 23.03	PGC 2206215	987 ± 6	2.04	4.43	−14.32	7.60
3047	12 47 16.10	11 45 36.79	IC 3775	1093 ± 13	2.04	4.69	−15.36	7.76
3055	12 46 07.25	−03 16 08.84	ASK 14107	1051 ± 16	2.06	–	−14.07	7.46
3062	11 47 51.36	53 50 48.00	PGC 2452508	1019 ± 39	1.74	–	−14.07	7.29
3063	11 40 33.03	57 33 35.12	ASK 297088	1020 ± 7	1.92	–	−13.53	7.30
3073	11 21 47.57	57 20 48.22	PGC 2563515	1081 ± 1	1.17	1.67	−13.84	6.89
3077	12 06 38.67	66 18 34.38	PGC 4567536	1180 ± 11	1.43	2.09	−14.15	7.14
3082	10 43 51.82	58 38 39.09	ASK 211533	1172 ± 1	1.11	1.51	−13.26	7.28
3083	10 01 09.46	08 46 55.61	ASK 277703	1265 ± 2	1.41	2.43	−14.03	7.14
3089	09 40 03.27	44 59 31.65	ASK 264262	1369 ± 1	1.30	2.14	−14.18	7.12
3096	11 46 04.48	56 33 56.13	ASK 296630	1021 ± 5	1.24	2.53	−13.26	6.93
3097	11 46 43.28	57 13 57.82	PGC 4015804	1029 ± 1	0.87	1.41	−13.52	6.56
3101	11 54 41.22	46 36 36.35	ASK 348521	1009 ± 1	0.50	0.79	−13.70	6.14
3120	12 41 29.16	00 43 11.39	ASK 1622	1175 ± 3	1.92	1.76	−13.45	7.02
3132	10 10 14.97	46 17 44.26	ASK 209206	1099 ± 1	0.95	1.64	−13.35	6.55
3134	11 25 26.75	65 46 07.14	SHOC 323	1118 ± 0	0.68	1.28	−13.61	6.34
3196	11 42 19.54	57 53 45.93	MCG +10-17-045	1327 ± 3	1.83	–	−15.71	7.54
3207	12 10 03.99	68 41 41.99	ASK 65617	2450 ± 2	1.92	1.68	−15.53	6.63
3306	23 20 28.21	15 04 20.91	2MASX J23202822+1504211	3828 ± 2	2.30	4.75	−18.77	9.57
3346	00 25 52.49	−09 39 42.37	2MASXJ00255209-0939420	5201 ± 2	2.65	5.60	−20.78	10.54
3371	03 34 17.75	−06 15 53.99	NGC 1361	5254 ± 2	2.60	5.51	−21.44	10.73
3381	07 36 16.80	33 07 21.76	IC 2201	4694 ± 1	2.79	–	−20.54	10.59
3387	08 25 17.47	03 56 00.90	CGCG 032-022	4169 ± 2	2.45	5.38	−20.26	10.19
3388	04 06 14.72	−05 39 06.84	IC 2031	4823 ± 3	1.68	3.03	−20.01	10.05
3406	00 04 35.22	00 50 54.98	2MASX J00043518+0050549	6392 ± 4	2.17	4.34	−18.54	9.57
3411	00 13 45.42	−09 56 03.67	ASK 126038	3822 ± 2	1.38	–	−16.79	8.24
3426	01 31 01.08	13 03 15.54	2MASX J01310103+1303155	6206 ± 2	2.63	6.25	−20.18	10.28

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			km · s^{-1}	mag	mag	mag	$M_\odot$
3438	07 25 24.12	42 25 59.11	ASK 475551	3013 ± 1	1.04	1.72	−16.12	7.89
3439	07 40 22.74	23 16 29.95	UGC 03960	2255 ± 2	2.31	5.11	−19.36	9.77
3459	08 47 21.99	42 23 55.65	ASK 191682	2980 ± 1	1.26	2.30	−16.13	8.03
3467	08 50 25.20	32 37 18.86	ASK 266943	2221 ± 1	1.24	1.99	−15.15	7.21
3472	13 39 38.33	42 32 44.84	ASK 309421	2644 ± 1	1.21	1.86	−15.39	7.65
3478	13 36 20.08	38 06 53.60	ASK 508595	5157 ± 2	1.40	–	−17.37	8.53
3483	13 39 03.16	43 33 58.60	CGCG 218-031	2740 ± 2	2.41	5.29	−18.94	9.71
3491	09 08 54.37	01 25 57.21	CGCG 006-006	4918 ± 3	1.68	2.31	−19.00	9.30
3498	14 19 29.09	35 34 01.11	ASK 392553	3207 ± 1	1.36	2.46	−16.81	8.39
3506	00 29 33.17	−09 04 50.74	IC 0022	6049 ± 2	2.53	5.71	−21.22	10.66
3513	00 23 47.67	00 39 40.27	ASK 29153	5422 ± 1	1.21	1.91	−17.26	6.36
3520	13 52 53.20	37 41 22.28	2MASX J13525317+3741219	5317 ± 2	2.55	5.72	−20.18	10.24
3527	13 56 55.58	05 09 06.83	CGCG 046-013	1517 ± 10	2.11	4.54	−16.85	8.60
3530	13 48 15.24	37 42 24.92	ASK 512367	2366 ± 1	1.55	2.57	−15.54	7.75
3531	13 57 21.10	26 12 27.27	PGC 1767195	2391 ± 4	1.27	2.11	−15.02	7.52
3532	14 03 52.44	55 35 32.40	ASK 301585	1969 ± 0	0.95	1.01	−14.24	6.64
3542	08 19 24.29	21 00 12.82	2MASX J08192430+2100125	3906 ± 4	2.54	3.52	−18.79	9.67
3546	07 45 01.45	43 55 52.36	2MASX J07450145+4355525	3736 ± 1	1.30	2.50	−17.01	8.43
3548	08 15 38.04	21 15 37.23	ASK 483519	3476 ± 1	0.98	1.61	−16.51	7.94
3556	08 51 56.67	16 56 41.32	UGC 04639	8549 ± 2	2.70	4.77	−22.44	11.18
3565	09 14 13.07	43 08 05.97	KUG 0910+433	4232 ± 2	1.90	–	−20.15	10.09
3570	15 33 36.76	33 21 33.65	ASK 313250	1999 ± 2	1.10	–	−14.85	7.40
3572	03 30 17.18	00 55 12.71	CGCG 390-096	6412 ± 2	2.75	5.71	−21.50	10.87
3581	03 37 10.37	−05 35 05.09	2MASX J03371039-0535055	8347 ± 2	2.61	5.48	−21.77	10.90
3609	01 33 52.56	13 42 09.39	ASK 43205	2599 ± 1	0.50	–	−15.04	6.60
3622	22 51 43.16	00 24 54.98	2MASX J22514319+0024547	7807 ± 2	2.84	5.43	−21.33	10.84
3643	00 25 07.43	00 18 45.63	UM 240	3269 ± 1	0.76	2.09	−16.57	7.66
3650	01 29 43.99	−01 14 29.10	UGC 01072	5190 ± 2	2.59	5.84	−21.23	10.61
3654	03 58 19.35	−04 43 49.56	ASK 57478	5294 ± 2	1.60	1.62	−18.21	8.93
3659	11 58 49.18	55 18 24.81	PGC 2832101	947 ± 27	2.14	–	−14.08	7.36
3665	07 22 46.74	41 39 29.65	UGC 03818	7002 ± 2	2.44	3.72	−21.19	10.67
3666	07 26 35.39	43 17 46.85	UGC 03844	3130 ± 2	2.70	5.65	−20.57	10.42
3668	07 24 33.15	41 57 02.18	2MASX J07243311+4157021	7960 ± 3	2.30	3.81	−19.49	9.96
3669	07 24 19.58	38 45 38.73	2MASX J07241961+3845387	7244 ± 2	2.46	5.18	−19.19	9.80
3670	07 21 00.79	41 03 06.58	2MASX J07210079+4103065	6903 ± 2	1.43	2.54	−18.69	9.22
3673	07 27 28.08	43 27 50.08	ASK 476000	6623 ± 1	1.13	2.14	−17.60	8.70
3691	12 28 08.61	12 05 35.85	IC 0794	1921 ± 3	2.37	5.40	−19.16	9.64
3698	01 26 42.25	−01 05 16.00	ASK 32676	4370 ± 8	2.29	4.71	−17.32	9.19
3710	14 01 57.67	46 18 42.90	NGC 5439	1895 ± 4	2.02	3.28	−18.59	9.42
3711	14 29 14.46	44 41 56.37	CGCG 247-032	2409 ± 2	2.48	5.50	−18.94	9.71
3716	11 52 24.07	32 24 13.96	NGC 3935	3066 ± 2	2.19	4.28	−20.30	10.22
3721	07 21 23.42	41 41 59.03	ASK 475472	7001 ± 2	1.05	1.84	−17.45	8.48
3735	00 19 47.33	00 35 26.86	2MASXJ00194732+0035273	5306 ± 3	1.94	3.82	−19.65	9.75
3762	00 39 55.98	−08 51 11.54	ASK 127226	4122 ± 1	1.18	1.87	−16.61	8.26
3779	01 28 59.57	00 33 42.93	UGC 01062	5464 ± 2	2.60	5.64	−21.76	10.85
3783	01 13 41.83	00 06 09.76	UGC 00771	5152 ± 2	2.55	4.34	−20.84	10.54
3784	01 15 54.29	13 21 12.05	CGCG 436-016	4184 ± 2	2.59	5.66	−20.14	10.24
3795	00 25 17.31	14 34 40.89	ASK 147884	5224 ± 1	1.06	1.81	−16.86	8.11
3801	01 03 01.72	−01 06 39.54	CGCG 384-061	5237 ± 2	2.51	5.20	−20.12	10.21
3805	01 22 46.01	00 25 35.09	2MASX J01224599-0025343	5182 ± 3	2.22	3.75	−18.95	9.61
3808	01 15 22.22	00 09 30.44	2MASX J01152220-0009299	5196 ± 4	2.32	4.97	−18.15	9.30
3813	01 04 25.32	−09 22 59.33	PGC 991736	4653 ± 1	1.41	2.49	−17.36	8.51
3818	01 27 25.12	−08 41 21.54	ASK 130312	3974 ± 1	1.31	2.16	−15.97	7.91

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			km · s^{-1}	mag	mag	mag	$M_\odot$
3832	01 38 47.38	01 04 18.45	UGC 01169	4964 ± 2	3.14	4.94	−20.34	10.58
3841	01 48 53.13	13 25 26.20	2MASXi J0148531+132526	4659 ± 2	2.17	4.93	−18.29	9.41
3842	01 49 53.83	12 58 33.74	2MASX J01495381+1258340	4691 ± 6	2.48	5.66	−18.14	9.35
3846	01 55 16.26	−08 51 15.39	KUG 0152-090	3950 ± 2	1.56	3.48	−17.30	8.60
3848	01 55 59.75	00 11 08.04	PGC 3111971	3637 ± 1	1.16	1.80	−17.01	8.19
3906	02 31 41.35	−09 18 01.60	NGC 0960	4867 ± 1	2.81	5.55	−20.85	10.71
3946	03 20 40.80	−07 23 39.89	NGC 1303	5461 ± 1	2.41	5.57	−21.33	10.59
3947	03 17 48.53	−07 37 00.65	NGC 1286	4312 ± 1	2.51	5.51	−20.80	10.45
3951	03 17 06.67	−06 30 07.76	2MASX J03170667-0630082	4008 ± 4	2.06	–	−19.74	9.91
3986	03 39 07.13	−06 27 02.88	2MASX J03390711-0627028	5381 ± 4	2.16	3.50	−20.03	10.08
3997	03 38 04.15	−05 21 09.07	ASK 56542	4019 ± 1	1.48	2.37	−17.25	8.56
4004	03 45 56.46	−05 08 31.59	ASK 56909	3893 ± 1	1.19	1.81	−16.09	7.96
4020	04 12 11.55	−04 52 27.14	ASK 57882	9047 ± 4	1.39	1.66	−18.88	9.19
4026	04 01 33.54	−06 20 12.46	PGC 1034032	5722 ± 1	1.50	2.61	−17.56	8.77
4030	07 27 19.37	44 25 38.37	2MASX J07271936+4425388	9742 ± 2	2.58	–	−20.87	10.89
4031	07 26 24.68	43 02 54.78	ASK 475844	7583 ± 2	1.70	2.95	−18.64	9.30
4060	07 41 54.78	38 29 53.74	2MASX J07415484+3829537	3507 ± 1	1.06	1.73	−17.27	8.31
4061	07 44 05.51	46 42 34.21	ASK 476552	3105 ± 1	1.15	–	−16.46	8.00
4093	08 18 08.76	23 02 51.31	IC 2269	3980 ± 2	3.01	5.73	−19.38	10.18
4094	08 19 32.11	21 23 39.50	IC 2293	4090 ± 3	1.89	3.17	−20.00	9.98
4096	08 00 40.82	39 52 13.81	CGCG 207-017	3975 ± 2	2.49	5.64	−19.30	9.87
4106	08 17 15.93	24 53 56.81	2MASXJ08171594+2453569	1898 ± 7	2.17	5.62	−16.08	8.42
4109	08 23 24.07	22 16 00.45	ASK 484591	2055 ± 13	2.46	4.60	−16.06	8.31
4129	08 37 26.59	40 02 08.11	UGC 04498	7091 ± 2	2.48	4.43	−21.61	10.81
4161	08 45 25.41	15 19 46.05	PGC 4176159	1614 ± 8	1.82	3.10	−13.74	7.30
4170	09 14 34.33	30 08 27.06	UGC 04869	6852 ± 2	2.81	4.72	−22.16	11.17
4175	09 10 11.89	50 24 04.75	NGC 2767	4925 ± 2	2.79	6.11	−20.98	10.71
4177	09 01 28.62	03 43 14.13	NGC 2729	3819 ± 1	2.71	5.98	−20.61	10.44
4201	09 14 57.34	06 00 18.61	ASK 261057	1370 ± 29	1.71	3.50	−14.88	7.56
4205	09 20 00.56	39 53 33.28	ASK 207278	1439 ± 2	1.38	2.74	−13.91	7.06
4214	09 39 34.69	48 25 18.56	UGC 05145	7378 ± 3	2.62	4.61	−21.97	11.21
4222	09 57 35.59	45 13 47.69	UGC 05345	3696 ± 2	2.53	3.70	−20.18	10.30
4224	09 46 48.95	09 44 10.08	CGCG 063-066	2997 ± 2	2.41	5.17	−19.66	9.95
4225	09 38 34.45	08 53 16.44	2MASX J09383447+0853159	3228 ± 2	2.38	0.83	−19.46	9.95
4231	09 48 41.60	59 15 39.59	2MASX J09484156+5915394	2231 ± 2	1.84	–	−18.19	9.31
4247	09 40 45.83	32 28 20.80	KUG 0937+327	1293 ± 1	1.24	2.36	−14.47	7.13
4309	10 37 37.91	37 27 20.30	NGC 3304	6937 ± 2	2.71	4.52	−21.97	11.11
4316	10 58 37.57	09 03 01.62	UGC 06062	2623 ± 2	2.59	5.69	−19.98	10.18
4319	10 35 04.85	46 33 41.23	UGC 05744	3356 ± 3	1.58	2.70	−19.85	9.66
4323	10 50 45.47	28 28 08.72	NGC 3400	1412 ± 1	2.36	5.73	−18.78	–
4348	10 56 19.93	17 05 06.00	PGC 4257755	960 ± 1	1.16	2.10	−12.88	6.57
4374	11 24 06.48	38 51 23.09	2MASX J11240641+3851226	1979 ± 3	2.45	6.18	−17.29	9.03
4387	11 04 56.82	17 38 30.51	PGC 4263693	907 ± 6	1.65	3.90	−14.01	7.39
4388	11 07 09.91	28 22 31.90	PGC 1833985	1003 ± 1	1.21	2.89	−14.19	7.09
4407	11 51 28.06	35 26 03.72	UGC 06827	3095 ± 2	2.47	5.58	−20.18	10.13
4414	11 51 13.43	50 09 24.84	NGC 3922	930 ± 3	2.21	6.41	−18.20	9.25
4428	11 37 44.42	54 02 44.70	ASK 236929	868 ± 12	1.52	2.83	−14.78	7.50
4435	11 50 59.61	47 57 49.48	ASK 348705	942 ± 3	1.23	2.25	−13.46	6.77
4448	12 05 59.46	02 29 53.63	CGCG 041-039	6098 ± 2	2.68	5.68	−21.29	10.96
4459	12 11 02.73	12 06 14.40	IC 0767	1877 ± 2	2.05	4.71	−18.82	9.41
4462	12 08 55.08	41 44 27.12	UGC 07129	925 ± 2	1.85	3.61	−17.43	8.97
4465	12 22 19.53	14 45 38.81	UGC 07436	982 ± 4	2.29	4.58	−17.13	8.76
4516	12 41 39.41	12 14 50.61	IC 3663	927 ± 10	1.99	4.09	−15.90	8.11

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			km · s^{-1}	mag	mag	mag	$M_\odot$
4517	12 58 09.02	14 51 32.19	UGC 08081	853 ± 16	1.92	–	−15.45	7.87
4522	12 51 00.22	01 50 45.17	PGC 1207820	953 ± 13	2.27	5.20	−14.58	7.57
4524	12 30 02.62	09 24 11.93	ASK 275272	899 ± 13	1.77	3.93	−13.75	7.33
4541	13 08 42.15	52 01 00.69	CGCG 271-017	4705 ± 2	2.62	5.86	−20.96	10.63
4542	13 14 46.04	53 49 13.27	CGCG 271-026	4711 ± 2	2.61	–	−20.77	10.53
4595	13 56 58.06	45 58 23.61	UGC 08876	2059 ± 1	2.56	5.50	−19.28	9.88
4598	13 48 56.27	59 50 08.64	UGC 08741	2019 ± 6	2.00	–	−18.70	9.44
4606	13 46 41.39	59 49 22.58	2MASX J13464142+5949225	1865 ± 2	1.87	4.33	−17.11	8.71
4609	13 53 19.84	05 46 17.93	UGC 08799	1046 ± 16	2.14	4.50	−15.71	8.26
4615	13 58 45.03	24 09 05.03	2MASX J13584501+2409048	970 ± 2	1.42	2.82	−14.26	7.34
4616	13 52 05.64	05 45 54.13	PGC 4389504	984 ± 2	1.87	2.99	−14.44	7.52
4617	13 34 06.94	09 15 43.27	PGC 4544337	1112 ± 11	1.64	2.96	−14.15	7.30
4634	14 03 47.16	35 44 30.04	UGC 08984	3805 ± 1	2.58	5.96	−20.61	10.42
4640	14 07 39.28	54 47 40.60	CGCG 272-032	2315 ± 1	2.33	5.29	−19.11	9.68
4644	14 09 54.94	56 49 21.36	CGCG 295-037	1790 ± 4	2.17	4.38	−17.85	9.18
4648	14 04 42.99	55 26 06.72	2MASX J14044299+5526064	1537 ± 2	1.86	4.29	−16.96	8.63
4656	14 02 36.07	39 13 13.28	PGC 50011	1326 ± 1	1.00	1.99	−15.15	7.13
4668	14 57 11.25	52 20 45.82	UGC 09629	7822 ± 2	2.78	5.00	−21.82	11.05
4687	14 48 58.97	33 11 34.95	CGCG 193-002	1700 ± 2	1.45	2.42	−16.89	8.33
4694	14 32 08.73	38 31 22.59	ASK 324027	1377 ± 2	1.62	–	−14.90	7.76
4696	14 54 19.38	38 48 29.91	CGCG 221-008	8685 ± 3	2.75	5.77	−21.76	11.03
4698	14 31 49.74	27 53 30.30	PGC 4553986	4415 ± 4	1.76	2.13	−16.91	8.32
4701	14 30 07.17	08 42 16.09	ASK 457251	1427 ± 35	1.23	2.73	−13.66	6.88
4720	15 11 48.38	46 15 13.90	CGCG 249-011	5454 ± 2	2.20	4.33	−20.41	10.31
4728	15 13 29.18	58 30 33.56	Mrk 0847	2543 ± 3	1.60	3.18	−18.96	9.41
4745	15 00 33.03	02 13 49.14	PGC 3350778	1312 ± 29	1.94	4.23	−15.32	7.73
4749	15 27 53.06	25 38 37.53	PGC 1744110	1470 ± 1	1.07	–	−13.92	6.91
4806	16 02 16.34	49 12 11.42	UGC 10150	6135 ± 6	2.12	3.64	−20.36	10.16
4808	16 02 11.63	07 05 09.97	CGCG 051-021	2654 ± 2	2.24	4.76	−19.45	9.98
4822	16 00 13.83	17 50 53.88	PGC 1543427	2053 ± 1	1.13	1.71	−15.91	7.71
4838	16 58 50.69	40 07 16.18	CGCG 225-021	8658 ± 2	2.73	6.08	−22.04	11.06
4846	16 41 06.57	46 00 14.63	IC 1226	5312 ± 4	2.35	3.95	−20.32	10.26
4855	16 35 20.71	17 45 55.19	Mrk 0886	2740 ± 1	1.12	–	−18.21	8.94
4861	16 34 39.88	46 57 37.26	2MASXi J1634398+465736	2452 ± 1	1.24	2.40	−16.74	8.26
4908	17 01 29.63	22 43 52.46	ASK 406590	2820 ± 1	2.22	–	−15.79	8.11
4989	21 16 24.75	10 16 24.07	CGCG 426-029	5175 ± 2	2.29	4.51	−20.22	10.15
4990	21 04 51.97	00 26 52.75	CGCG 374-042	4129 ± 2	2.45	5.64	−19.62	9.95
4999	21 13 07.73	01 13 47.13	2MASX J21130776+0113480	4157 ± 2	1.30	2.42	−18.41	8.97
5020	21 50 38.41	13 17 17.48	CGCG 427-027	6045 ± 2	2.59	–	−21.61	10.88
5025	21 50 27.56	12 38 10.37	2MASX J21502753+1238103	6507 ± 2	2.48	–	−19.53	9.98
5040	21 52 22.47	−01 10 15.96	ASK 21400	4770 ± 1	1.15	1.80	−16.83	8.14
5041	21 54 47.73	00 13 45.67	ASK 21641	2984 ± 1	1.33	2.23	−16.05	7.88
5048	22 01 41.64	11 51 24.41	CGCG 428-014	8944 ± 3	3.13	–	−21.95	11.33
5070	22 04 08.80	00 55 31.78	ASK 22153	4825 ± 16	2.28	3.91	−17.21	8.66
5105	22 33 49.80	00 28 58.47	ASK 23755	4642 ± 1	1.09	–	−17.02	8.22
5117	23 18 08.39	00 23 25.63	CGCG 380-023	8775 ± 2	2.71	5.49	−22.40	11.17
5118	23 20 29.08	−01 00 08.81	CGCG 380-035	9370 ± 2	2.75	5.89	−22.03	11.08
5120	23 23 27.04	14 19 33.09	CGCG 431-053	7651 ± 2	2.78	5.55	−21.40	10.86
5164	23 33 27.02	14 20 06.86	CGCG 432-002	5735 ± 3	2.16	3.91	−20.74	10.37
5165	23 43 55.95	00 13 05.01	CGCG 381-031	6627 ± 1	2.48	5.59	−20.28	10.25
5170	23 56 11.28	00 01 45.24	2MASX J23561127-0001452	6739 ± 4	2.52	4.70	−19.07	9.76
5195	13 28 33.04	32 04 09.73	NGC 5166B	4838 ± 3	2.40	5.80	−19.96	10.09
5210	23 47 03.79	14 50 30.38	CGCG 432-030	6622 ± 2	2.19	3.39	−20.79	10.39

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			km · s^{-1}	mag	mag	mag	$M_\odot$
5217	23 52 42.63	−11 00 53.07	ASK 124704	5374 ± 1	1.31	1.79	−17.20	8.38
5241	12 10 32.83	58 18 14.90	NGC 4149	3057 ± 3	2.35	4.10	−20.07	10.37
5248	12 17 08.21	27 47 42.96	PGC 1818175	971 ± 10	1.82	4.31	−15.43	7.88
5252	00 00 47.89	14 16 39.17	2MASX J00004789+1416390	10847 ± 3	2.45	4.05	−20.82	10.71
5256	00 21 35.29	01 09 56.33	CGCG 383-002	5465 ± 1	2.45	5.67	−20.01	10.15
5258	00 15 05.37	15 03 59.70	2MASX J00150533+1503597	5428 ± 2	2.13	–	−19.86	10.00
5260	00 10 25.49	14 17 23.06	KUG 0007+140	5452 ± 4	1.71	2.92	−19.34	9.50
5279	00 31 36.55	−10 53 17.43	2MASXJ00313653-1053178	5075 ± 2	2.56	5.66	−20.52	10.37
5286	00 40 26.90	00 02 43.53	2MASX J00402686-0002431	5847 ± 2	1.68	2.82	−19.11	9.46
5296	00 37 59.82	00 42 35.95	ASK 29753	4167 ± 1	1.44	2.90	−16.69	8.39
5305	01 12 57.42	00 20 41.94	NGC 0429	5628 ± 2	2.66	6.10	−21.28	10.76
5309	01 18 44.86	00 54 22.80	CGCG 385-063	5251 ± 3	2.52	3.68	−20.08	10.09
5310	01 26 25.91	−08 33 51.12	2MASXJ01262589-0833510	5336 ± 1	2.35	5.25	−20.50	10.18
5321	01 26 49.69	00 58 31.08	Mrk 0570	4853 ± 3	1.68	1.78	−19.27	8.91
5323	01 13 02.84	00 40 40.62	ASK 31963	4500 ± 2	1.20	1.59	−16.71	8.59
5326	01 09 42.67	00 57 13.29	ASK 31197	5203 ± 3	1.48	2.38	−16.83	8.24
5334	01 40 53.63	−10 16 41.56	2MASXJ01405362-1016415	5296 ± 1	2.57	5.76	−20.75	10.50
5335	01 31 04.76	13 38 58.79	CGCG 436-067	5889 ± 2	2.76	–	−20.40	10.49
5337	01 51 57.52	−08 31 03.44	IC 0170	5192 ± 2	2.49	5.18	−20.45	10.31
5340	01 42 30.89	00 49 10.03	CGCG 386-032	5189 ± 2	1.58	1.99	−19.35	9.43
5344	01 34 37.53	−01 03 08.03	ASK 33113	4787 ± 7	1.89	3.26	−17.99	9.25
5353	01 54 41.85	00 36 41.71	PGC 1170628	4618 ± 5	1.50	2.21	−16.79	8.24
5369	02 20 42.04	−08 08 26.66	2MASX J02204200-0808252	4710 ± 3	1.77	2.94	−18.94	9.44
5437	03 02 35.01	−07 21 45.38	ASK 55199	5986 ± 3	1.21	1.82	−16.96	8.29
5440	03 02 11.39	−01 09 59.60	ASK 37576	4079 ± 2	1.40	2.81	−16.36	8.13
5453	03 42 11.75	00 21 13.02	UGC 02832	7661 ± 1	1.31	–	−19.97	7.84
5473	04 06 44.26	−04 26 50.54	2MASX J04064426-0426510	11300 ± 2	1.40	2.48	−19.98	9.69
5481	07 19 51.50	41 24 58.04	2MASX J07195149+4124585	10470 ± 6	2.31	–	−19.70	10.05
5483	07 25 04.45	40 35 37.61	ASK 475402	10200 ± 1	0.96	1.64	−18.91	8.84
5507	07 35 24.17	36 38 36.47	ASK 44962	3902 ± 1	1.08	1.25	−17.11	8.23
5508	07 46 19.80	39 51 26.07	ASK 45258	3895 ± 2	1.56	2.53	−17.09	8.57
5522	08 03 28.94	33 27 44.59	KUG 0800+336	11794 ± 2	1.74	3.16	−22.64	11.17
5525	08 20 09.90	27 05 36.51	UGC 04341	5869 ± 2	2.64	–	−21.54	10.92
5526	08 26 18.06	27 50 24.47	IC 2365	6019 ± 1	2.55	–	−21.66	10.77
5528	08 19 51.89	20 59 05.90	NGC 2560	4883 ± 2	2.65	6.22	−21.37	10.78
5547	08 22 15.81	46 41 30.00	2MASX J08221577+4641309	2179 ± 1	1.41	2.36	−16.36	8.26
5549	08 01 32.23	21 22 48.46	ASK 363277	2104 ± 3	1.17	2.33	−15.93	7.86
5559	08 17 11.12	44 46 37.88	ASK 87147	8330 ± 2	1.44	1.85	−18.63	7.42
5575	08 58 30.48	02 55 31.74	IC 2426	3887 ± 1	2.34	5.07	−19.75	9.96
5580	08 36 28.38	18 21 38.96	CGCG 089-052	4006 ± 2	1.64	2.84	−18.74	9.28
5586	08 34 20.18	50 27 08.94	2MASX J08342016+5027088	3433 ± 2	1.61	2.51	−17.13	8.63
5593	08 36 41.13	05 16 24.90	ASK 259050	3127 ± 1	0.76	1.13	−15.88	7.54
5597	08 44 42.72	29 32 43.09	ASK 281794	2116 ± 2	1.15	1.87	−14.73	7.26
5602	08 49 01.38	29 29 17.51	KUG 0845+296	8239 ± 1	0.69	0.44	−17.70	6.80
5656	09 46 45.58	09 15 51.71	CGCG 063-065	4030 ± 2	2.53	4.20	−19.34	9.85
5667	09 52 35.11	08 11 56.62	ASK 277328	2725 ± 3	1.35	2.13	−17.32	8.53
5668	09 46 28.57	00 26 03.56	ASK 51	1810 ± 1	1.45	2.70	−16.92	8.34
5679	09 46 53.04	31 47 44.68	KUG 0943+320	1425 ± 3	1.39	2.72	−14.79	7.47
5689	10 21 21.12	24 20 29.04	UGC 05591	11118 ± 2	2.71	5.29	−22.33	11.21
5690	10 22 06.03	21 03 54.37	CGCG 124-015	11738 ± 3	2.77	–	−22.63	11.33
5694	10 01 31.06	37 12 14.20	IC 2530	6559 ± 2	2.64	6.17	−21.55	10.76
5696	10 25 47.70	26 34 14.63	CGCG 154-013	5023 ± 2	2.56	5.60	−21.12	10.62
5699	10 00 45.21	04 44 03.56	UGC 05383	3968 ± 2	2.64	5.11	−20.20	10.33

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			km · s^{-1}	mag	mag	mag	$M_\odot$
5701	10 04 31.75	60 06 28.71	NGC 3102	3038 ± 2	2.44	5.46	−20.01	10.15
5703	10 02 00.95	59 15 08.32	CGCG 289-027	2802 ± 4	2.09	4.72	−19.12	9.60
5735	10 47 26.58	06 02 53.24	NGC 3376	5816 ± 2	2.72	5.62	−21.42	10.76
5769	10 35 11.05	25 27 04.02	PGC 4243890	1298 ± 1	0.84	1.73	−13.52	6.62
5816	11 38 36.24	33 52 04.62	CGCG 186-015	10844 ± 3	2.87	6.32	−22.59	11.37
5821	11 57 42.90	57 33 57.00	UGC 06939	4915 ± 2	2.56	6.20	−21.20	10.63
5822	11 39 03.33	00 12 21.67	IC 0716	5413 ± 2	2.59	3.96	−20.79	10.68
5825	11 34 11.68	12 30 44.33	NGC 3731	3195 ± 2	2.62	5.64	−20.35	10.31
5828	11 59 49.61	30 50 39.92	IC 2986	3108 ± 2	2.46	5.17	−19.65	10.00
5839	11 56 56.36	60 19 45.89	CGCG 292-051	1391 ± 2	1.33	2.19	−16.76	8.40
5841	11 44 24.49	60 53 02.67	2MASX J11442453+6053022	1497 ± 1	1.45	2.57	−16.42	8.41
5843	11 56 07.08	64 21 02.64	2MASXi J1156072+642102	1410 ± 1	1.77	4.49	−16.34	8.14
5850	11 58 11.02	58 09 23.32	2MASX J11581094+5809237	954 ± 2	1.54	2.85	−14.68	7.59
5853	11 35 38.17	31 10 28.68	ASK 498176	970 ± 49	1.11	1.91	−13.66	6.77
5875	12 10 57.87	63 54 55.22	UGC 07179	2616 ± 3	2.60	–	−19.70	6.20
5883	12 09 22.26	13 59 32.74	IC 3019	1675 ± 14	2.17	–	−18.16	9.07
5884	12 20 24.30	05 34 22.14	NGC 4282	937 ± 3	2.50	4.86	−17.14	9.11
5885	12 15 12.56	14 25 58.37	IC 3065	995 ± 3	2.09	4.70	−17.39	8.94
5888	12 21 04.81	11 45 16.56	IC 3192	1446 ± 11	2.03	4.05	−16.33	8.41
5889	12 07 12.13	63 53 26.64	MCG +11-15-024	1425 ± 3	1.81	3.05	−16.43	8.39
5896	12 18 31.88	13 11 28.18	VCC 0293	956 ± 30	3.20	3.23	−14.08	–
5903	12 12 55.18	44 05 27.42	ASK 319304	968 ± 1	1.15	2.06	−13.19	6.74
5913	12 51 16.52	53 41 46.66	IC 0830	4847 ± 2	2.56	6.00	−21.30	10.67
5921	12 35 20.23	06 06 54.33	NGC 4543	2374 ± 1	2.41	5.20	−19.48	9.87
5923	12 30 10.32	10 46 46.12	NGC 4482	1835 ± 3	2.10	4.92	−19.42	9.57
5926	12 36 54.85	12 31 12.27	IC 3586	1723 ± 4	2.00	4.34	−18.55	9.17
5931	12 40 13.38	12 52 29.19	IC 3635	1560 ± 8	2.26	4.08	−16.87	8.69
5936	12 39 37.71	10 58 32.68	VCC 1803	1355 ± 21	2.48	2.78	−15.88	8.24
5946	12 59 45.14	00 52 16.68	PGC 1132599	1229 ± 29	1.44	2.43	−13.88	7.06
5948	12 55 52.01	63 36 39.30	IC 0836	2658 ± 5	2.28	6.59	−17.87	7.62
5951	12 30 16.73	09 05 06.92	IC 3430	2115 ± 1	1.75	3.59	−17.88	7.64
5989	13 09 36.90	31 40 34.06	PGC 1958740	2159 ± 4	1.28	2.10	−15.29	7.56
5999	13 48 34.66	37 06 47.91	CGCG 190-059	10255 ± 2	2.54	5.87	−22.49	11.11
6016	13 50 41.47	40 16 45.63	CGCG 218-060	2300 ± 2	2.57	5.25	−18.66	9.62
6033	13 35 37.17	14 21 39.43	PGC 1456087	993 ± 21	1.57	2.93	−14.26	7.38
6037	13 30 47.82	39 54 45.70	ASK 355809	1236 ± 1	1.29	2.26	−13.60	6.86
6071	14 07 24.28	59 30 09.45	2MASX J14072420+5930095	1820 ± 1	1.27	2.16	−15.78	7.83
6078	14 02 08.70	61 24 44.64	ASK 108562	1691 ± 2	1.07	1.79	−14.49	7.14
6086	14 54 32.75	02 57 59.13	NGC 5776	8165 ± 3	2.74	5.96	−22.65	11.22
6107	14 45 15.84	00 09 34.17	ASK 8160	1664 ± 4	1.60	2.83	−16.87	8.40
6118	14 45 36.28	34 10 43.89	ASK 394208	1642 ± 6	1.28	2.54	−14.70	7.15
6120	14 31 53.00	03 22 48.27	2MASX J23244466+0101490	1529 ± 1	0.98	2.31	−13.90	6.94
6128	15 23 35.25	09 20 46.14	CGCG 077-110	10703 ± 2	2.69	6.33	−22.45	11.22
6135	15 18 43.82	41 51 55.67	NGC 5914	5441 ± 2	2.65	5.76	−21.23	10.74
6156	15 27 44.49	09 41 56.85	ASK 421256	1820 ± 1	1.29	2.32	−15.95	7.96
6157	15 12 24.05	02 04 48.23	ASK 84838	1682 ± 10	2.00	4.34	−16.04	8.32
6392	21 15 16.19	09 53 47.21	ASK 138334	5239 ± 1	0.97	1.51	−17.38	8.38
6483	23 24 23.54	15 26 36.37	2MASX J23242356+1526362	7658 ± 2	2.59	5.08	−20.49	10.41
6498	23 16 03.79	13 47 05.74	ASK 143916	4504 ± 6	1.63	2.33	−16.81	8.38
6508	23 41 25.92	00 39 43.89	2MASX J23412592-0039437	10864 ± 2	2.77	5.52	−20.93	10.68
6510	23 53 31.01	00 44 24.61	CGCG 381-058	6925 ± 3	2.72	5.31	−20.51	10.57
6513	23 40 32.70	00 33 03.84	2MASXJ23403276-0033029	6928 ± 3	1.52	2.18	−20.07	9.57
6546	07 40 42.64	23 37 34.47	CGCG 117-065	7343 ± 2	2.76	–	−21.17	10.76

Table C.1: continued from previous page

Source	RA	Dec	Name	$V_{\rm opt}$	$u-r$	NUV $-r$	M_r	$M_\star$ [log]
	(J2000)			$\rm km\cdot s^{-1}$	mag	mag	mag	$M_\odot$
6562	07 33 27.72	41 53 08.78	2MASX J07332776+4153084	5888 ± 2	2.30	5.07	−19.35	9.82
6572	07 37 59.17	33 56 27.05	ASK 85066	7108 ± 2	1.62	–	−17.83	8.87
6573	07 36 59.47	30 29 53.49	PGC 1906163	4546 ± 2	1.35	–	−17.25	8.54
6735	08 33 55.25	47 08 03.99	CGCG 237-006	4474 ± 2	2.25	–	−19.67	9.95
6741	08 39 00.68	40 20 45.98	UGC 04513	7410 ± 2	2.74	7.56	−21.74	10.88
6747	08 46 45.23	06 57 35.86	UGC 04594	8732 ± 3	2.60	–	−21.91	11.01
6748	08 47 49.73	18 54 42.83	IC 2399	4225 ± 7	1.98	3.24	−18.87	9.44
6759	13 57 29.53	09 57 03.24	CGCG 074-017	6969 ± 2	2.39	–	−20.53	10.29

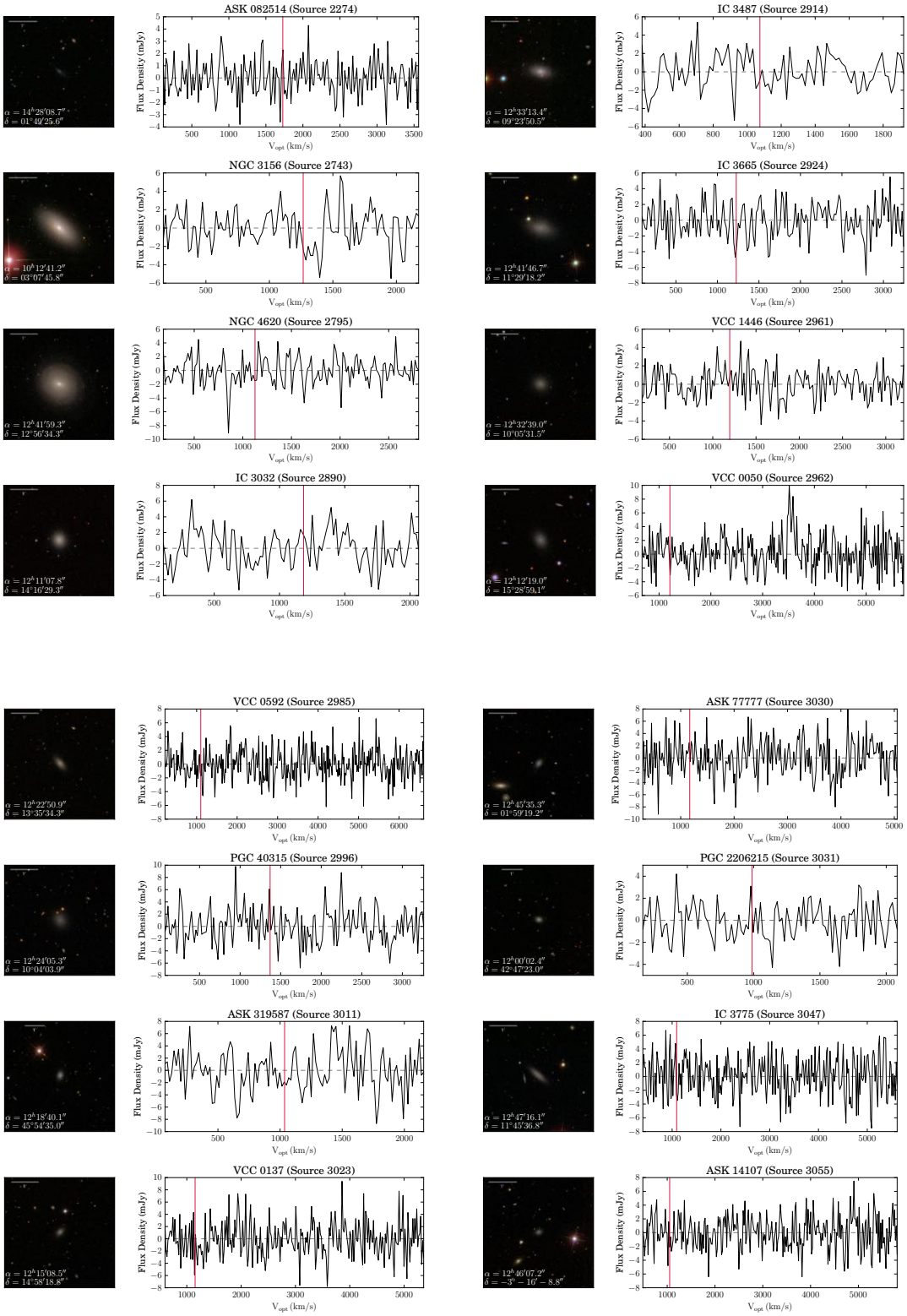


Figure C.1: *Optical images and* HI *spectra for galaxies in CRUMBS Stacking Sample.*

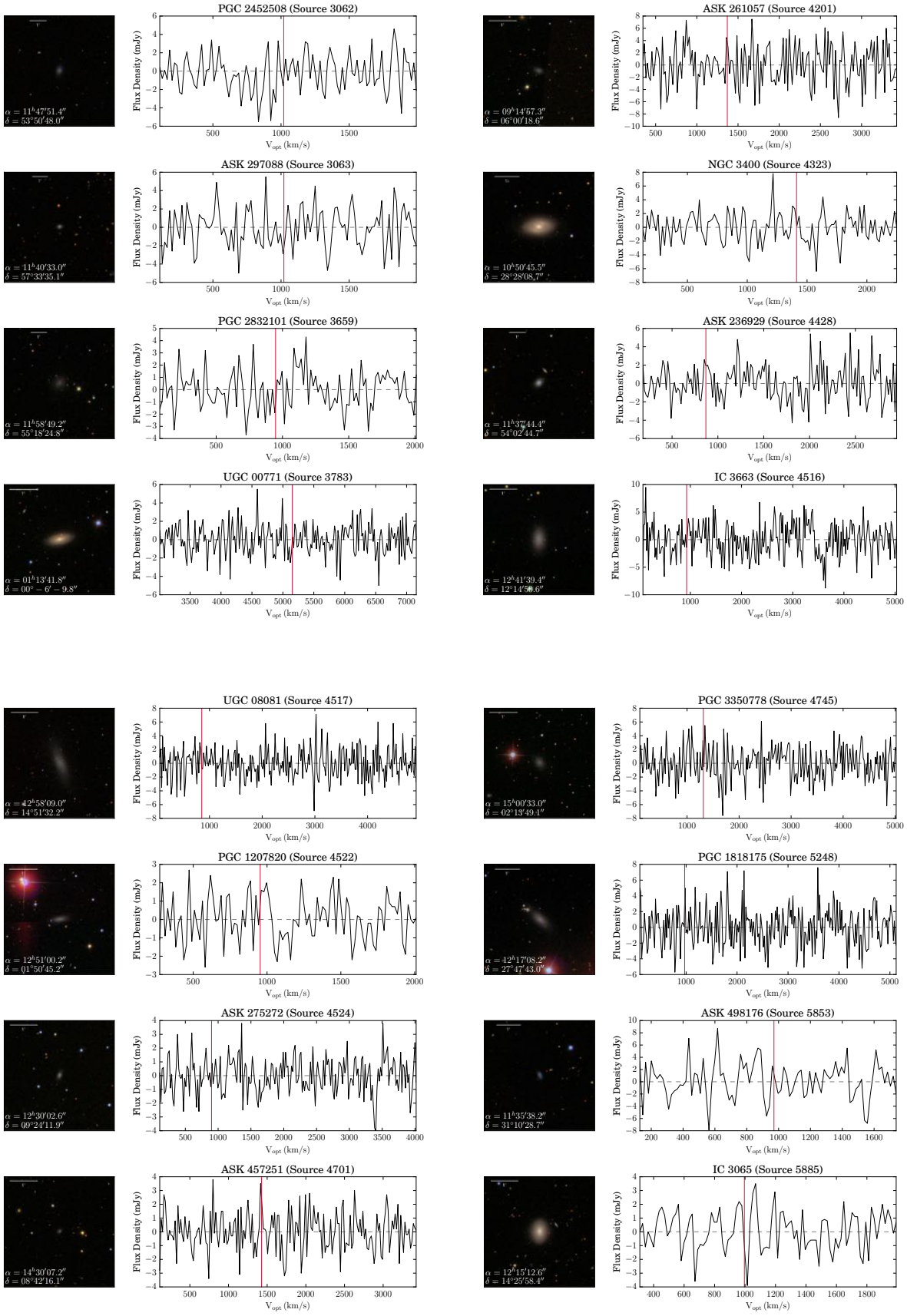


Figure C.1: *(continued) CRUMBS Stacking Sample*

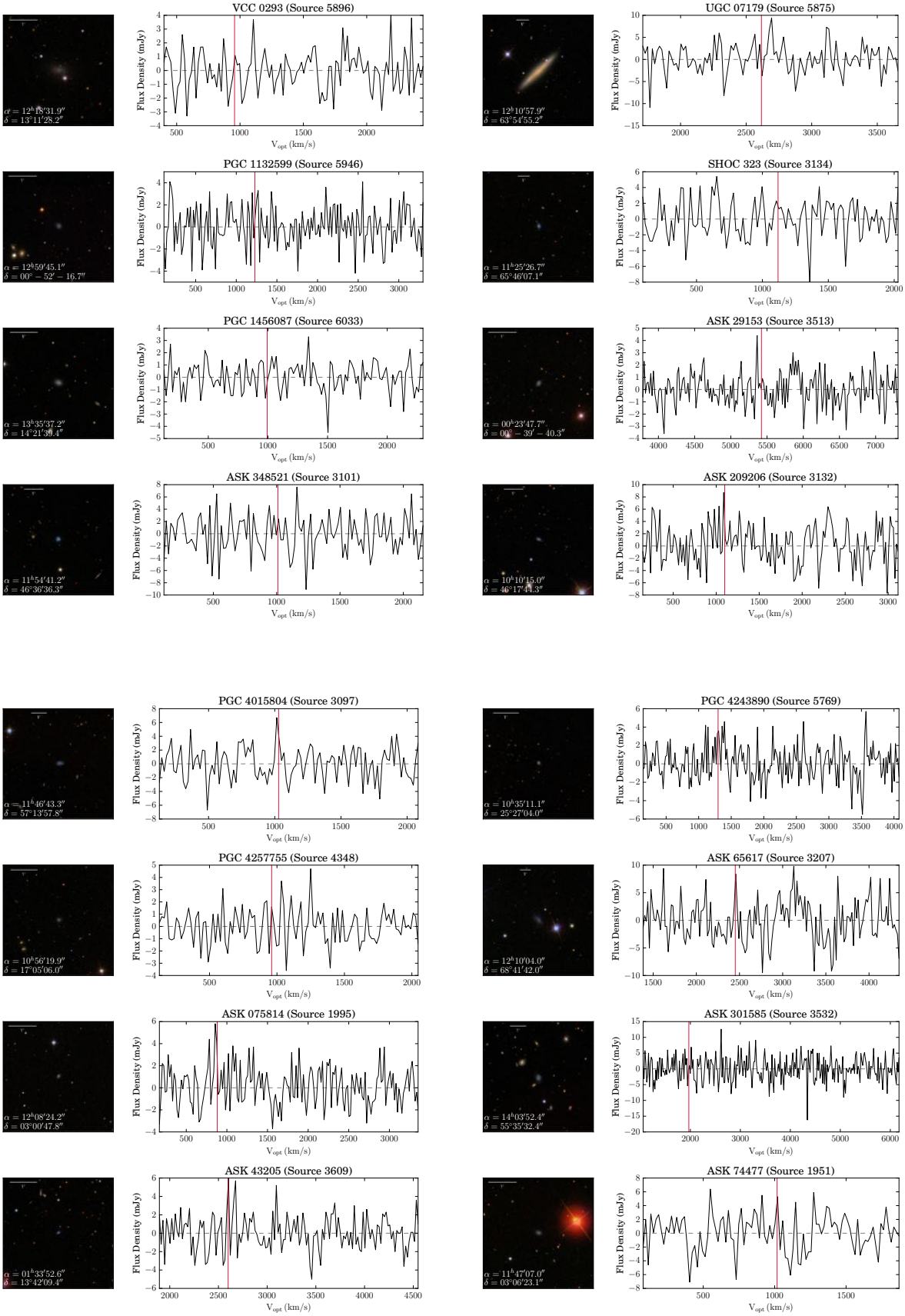


Figure C.1: *(continued) CRUMBS Stacking Sample*

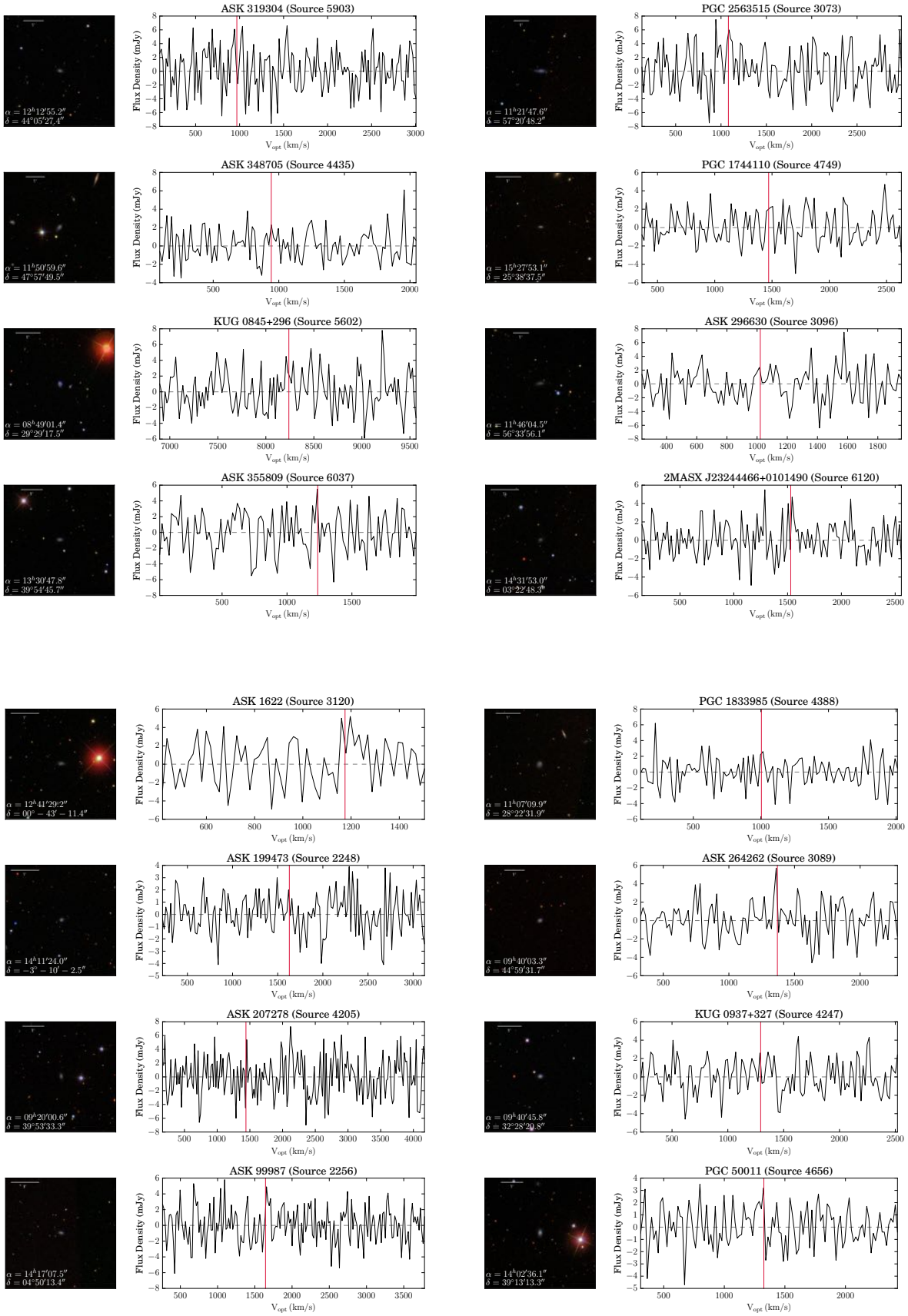


Figure C.1: *(continued) CRUMBS Stacking Sample*

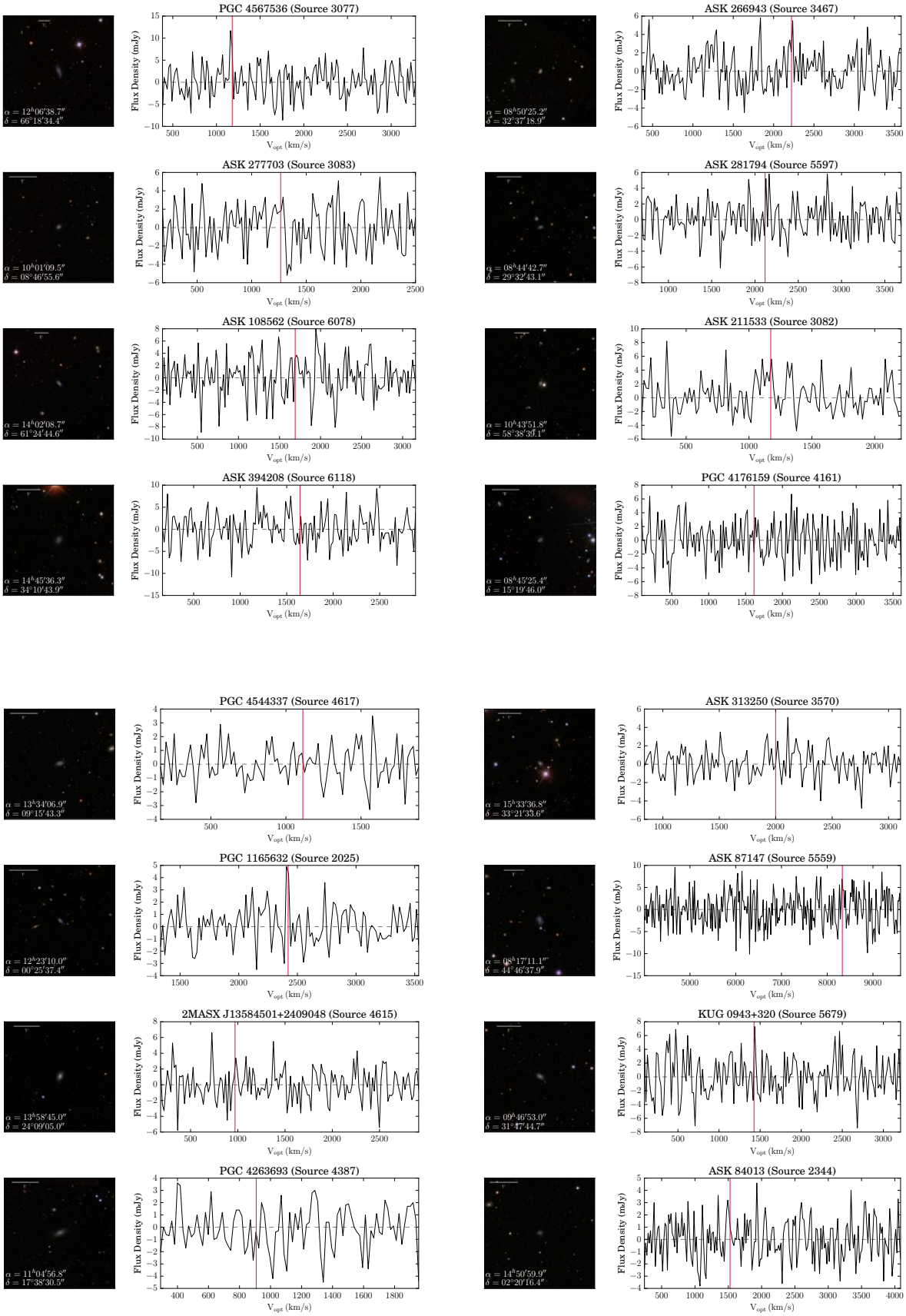


Figure C.1: *(continued) CRUMBS Stacking Sample*

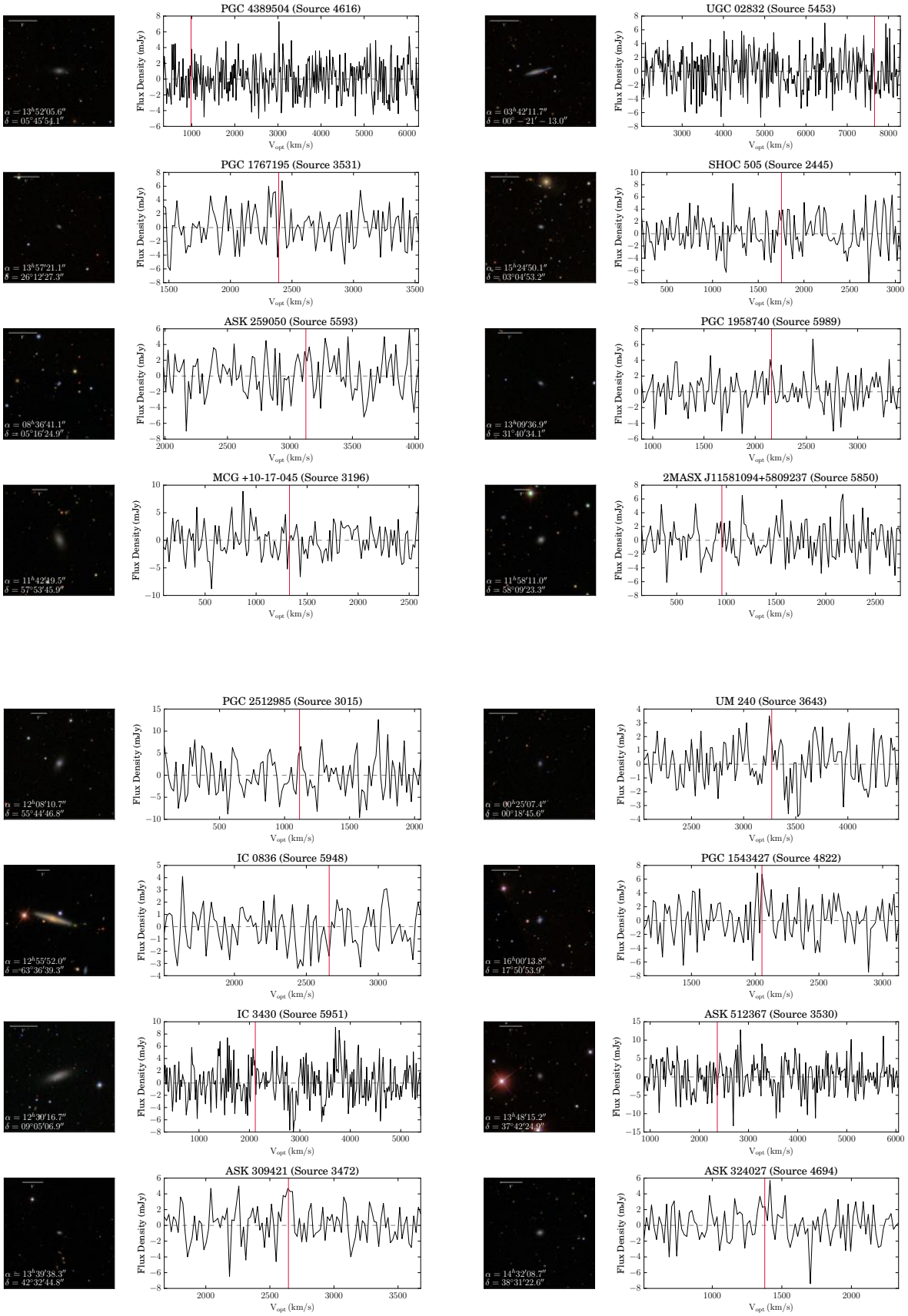


Figure C.1: *(continued) CRUMBS Stacking Sample*

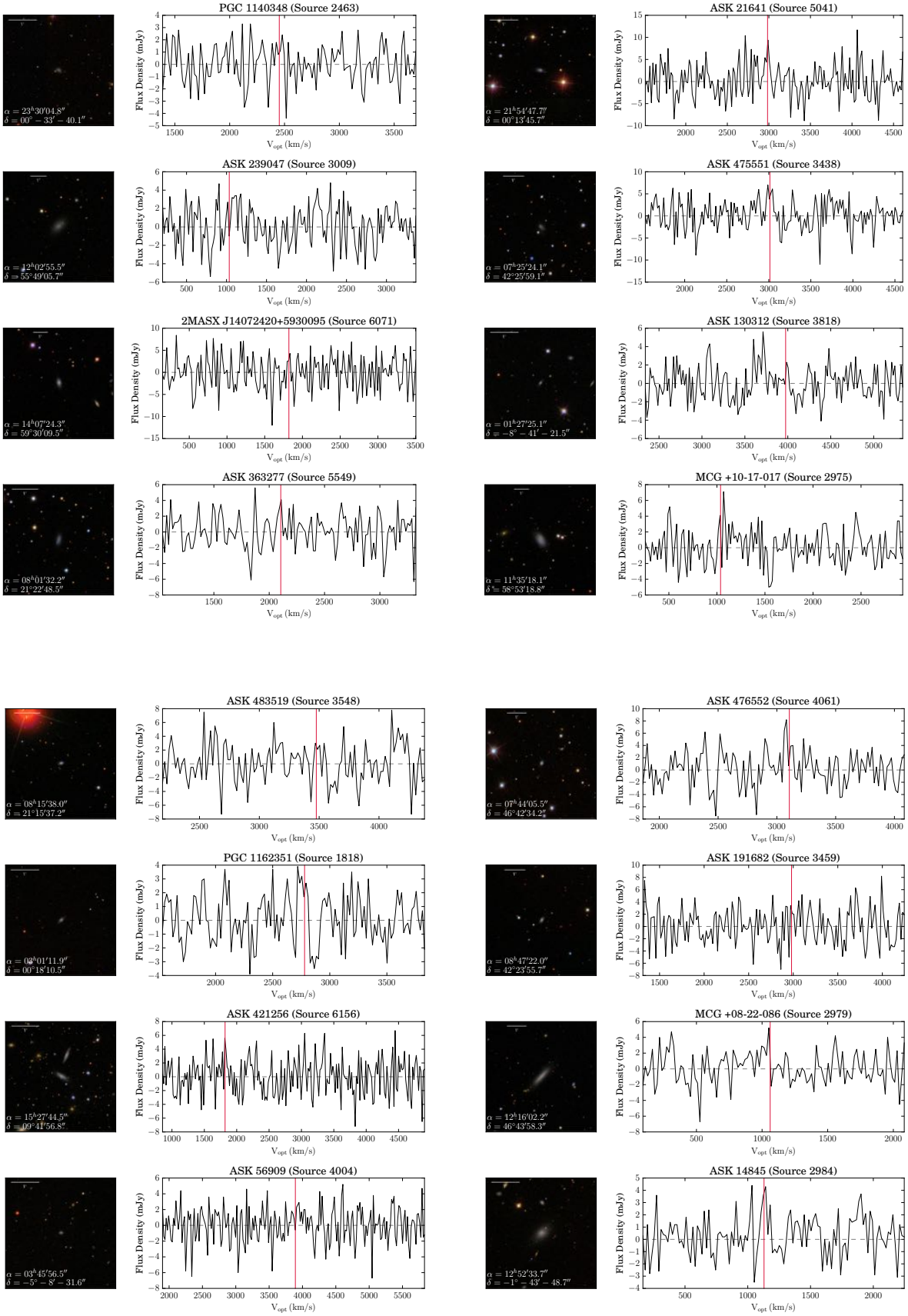


Figure C.1: *(continued) CRUMBS Stacking Sample*

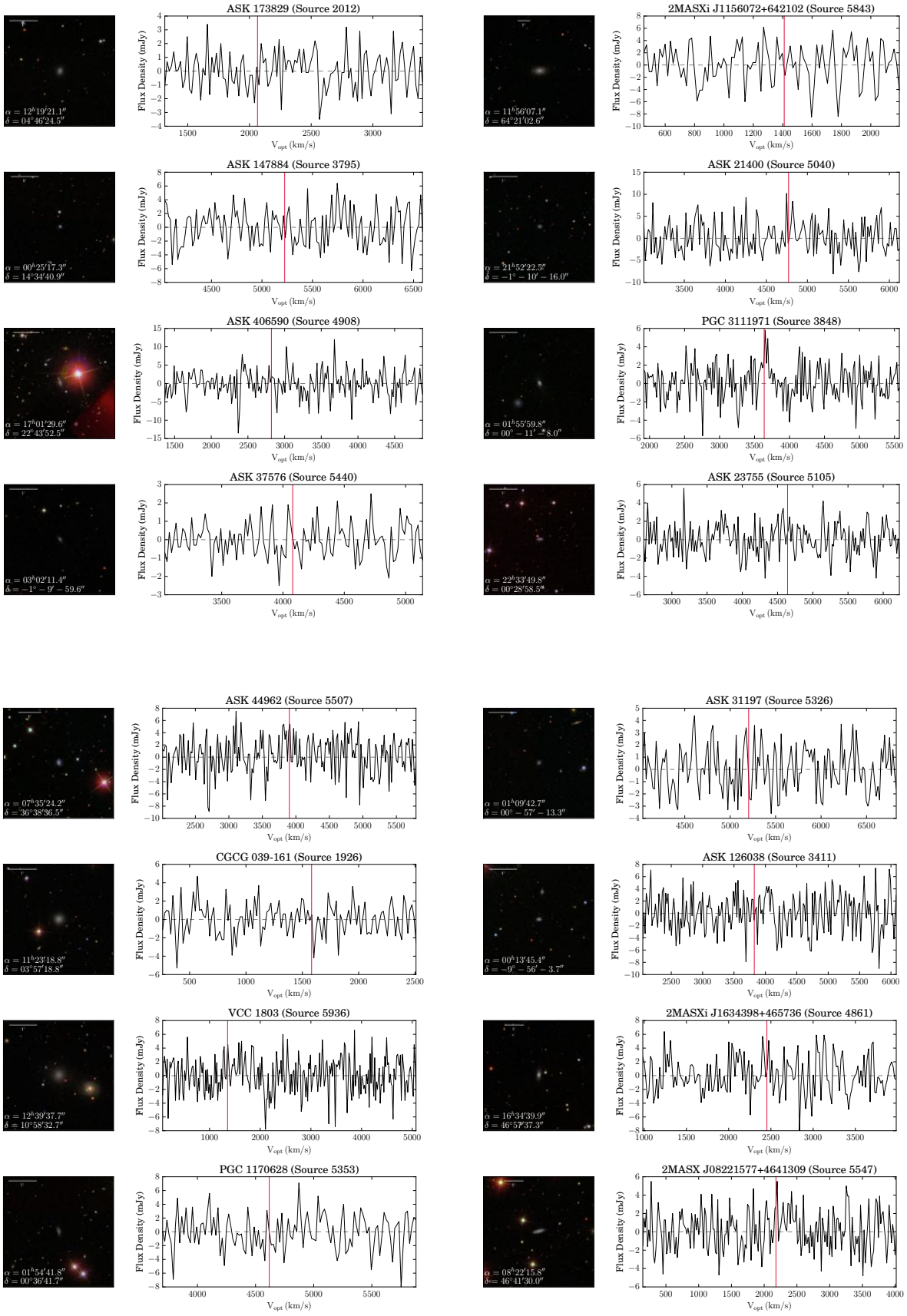


Figure C.1: *(continued) CRUMBS Stacking Sample*

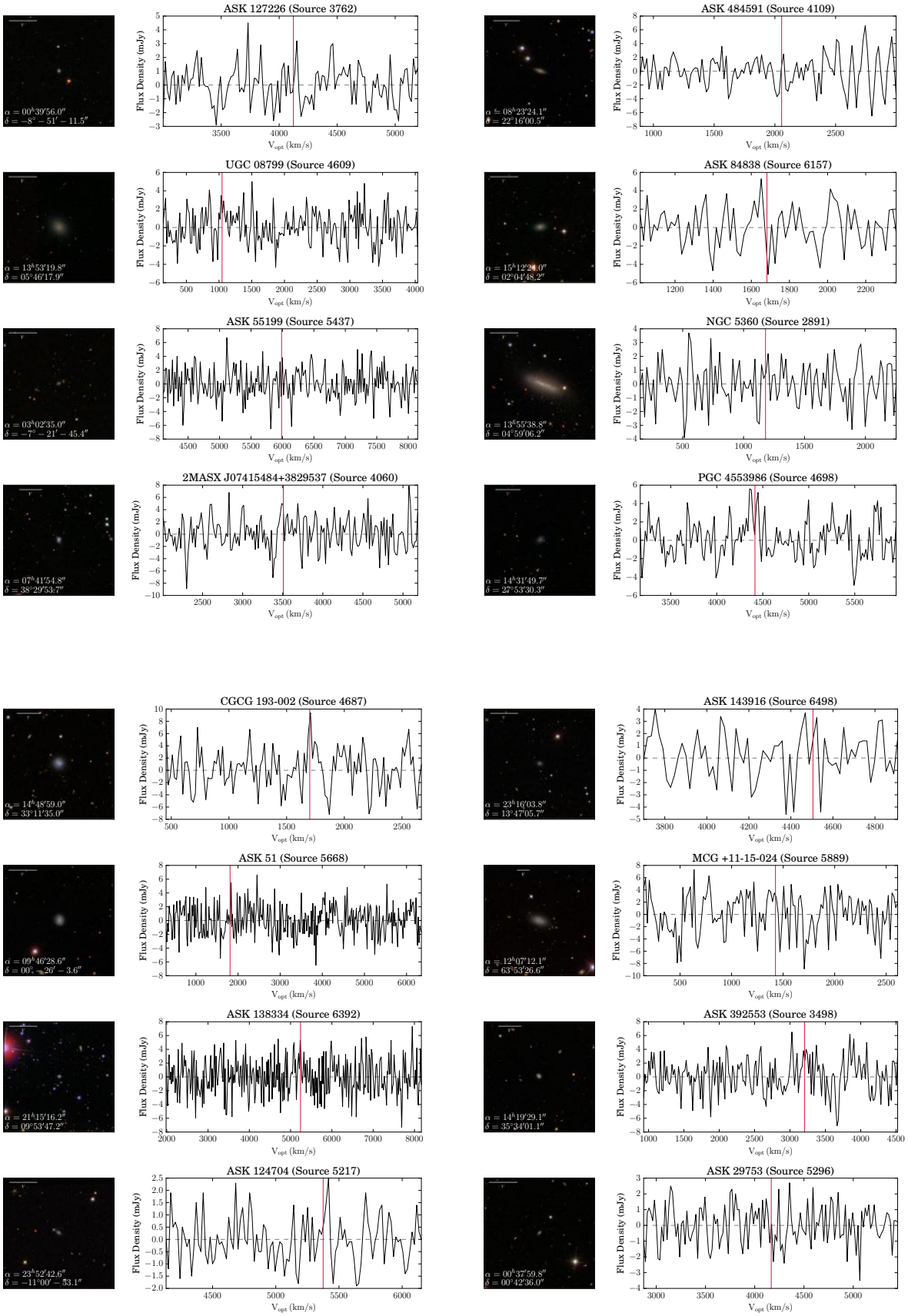


Figure C.1: *(continued) CRUMBS Stacking Sample*

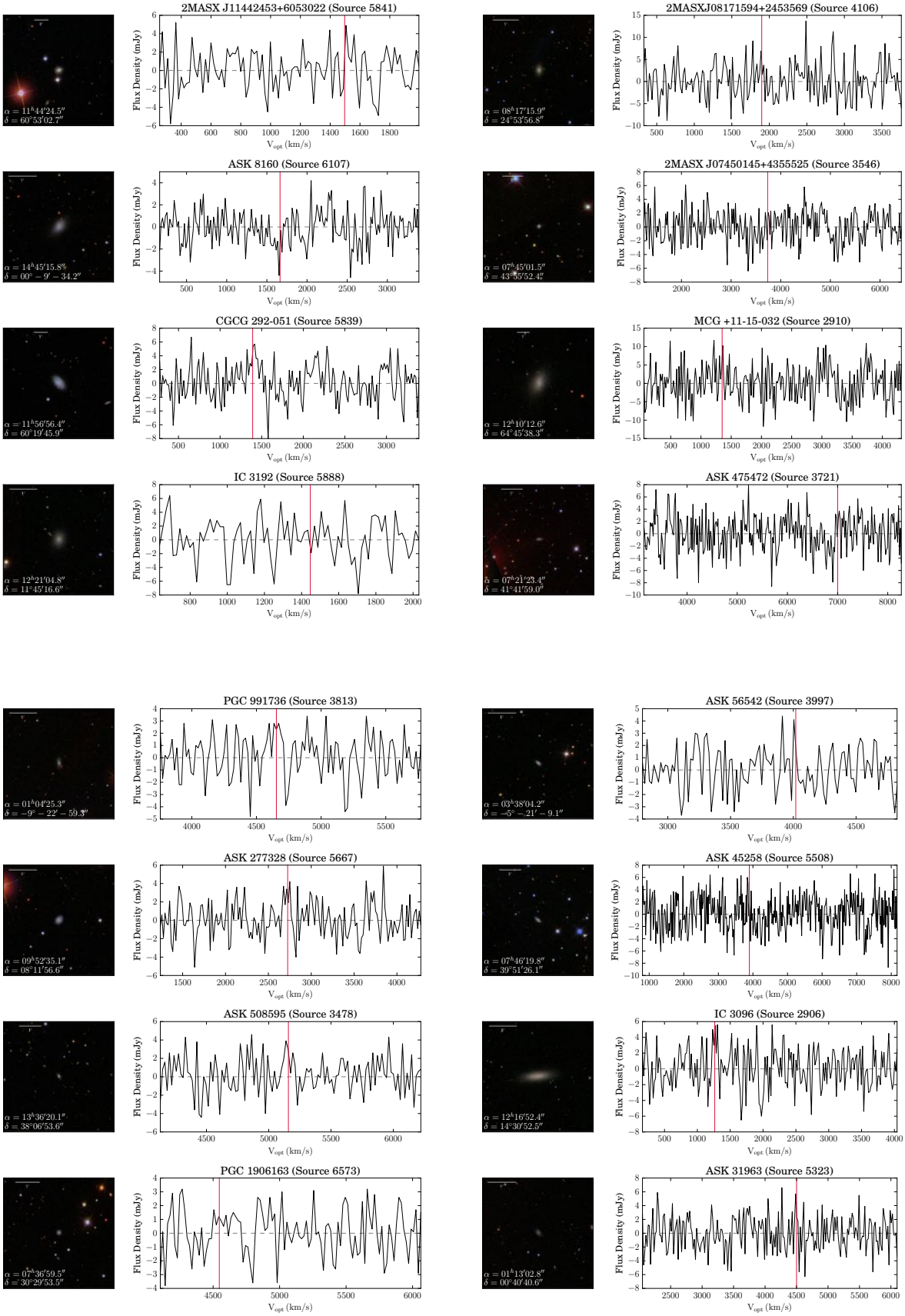


Figure C.1: *(continued) CRUMBS Stacking Sample*

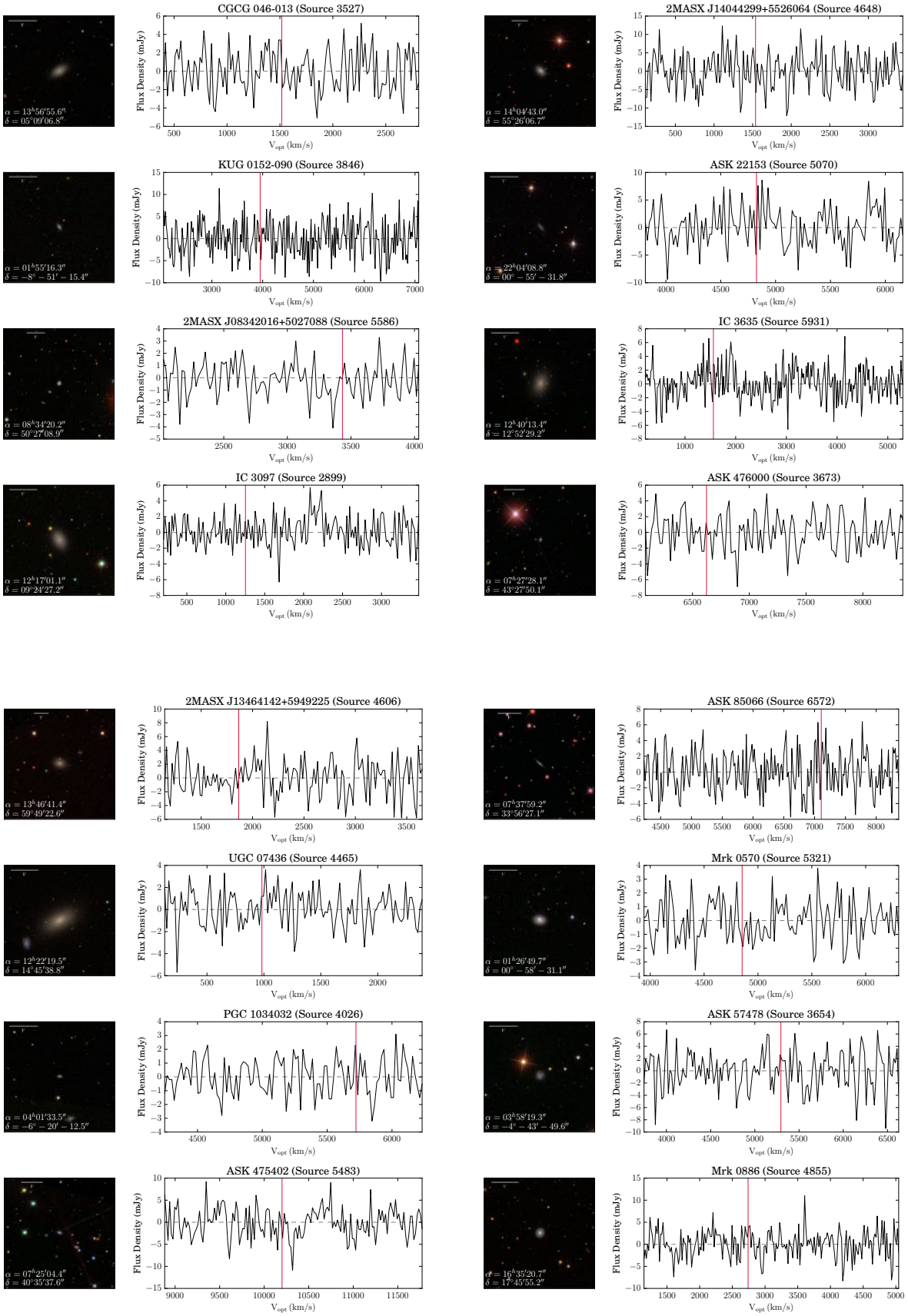


Figure C.1: *(continued) CRUMBS Stacking Sample*

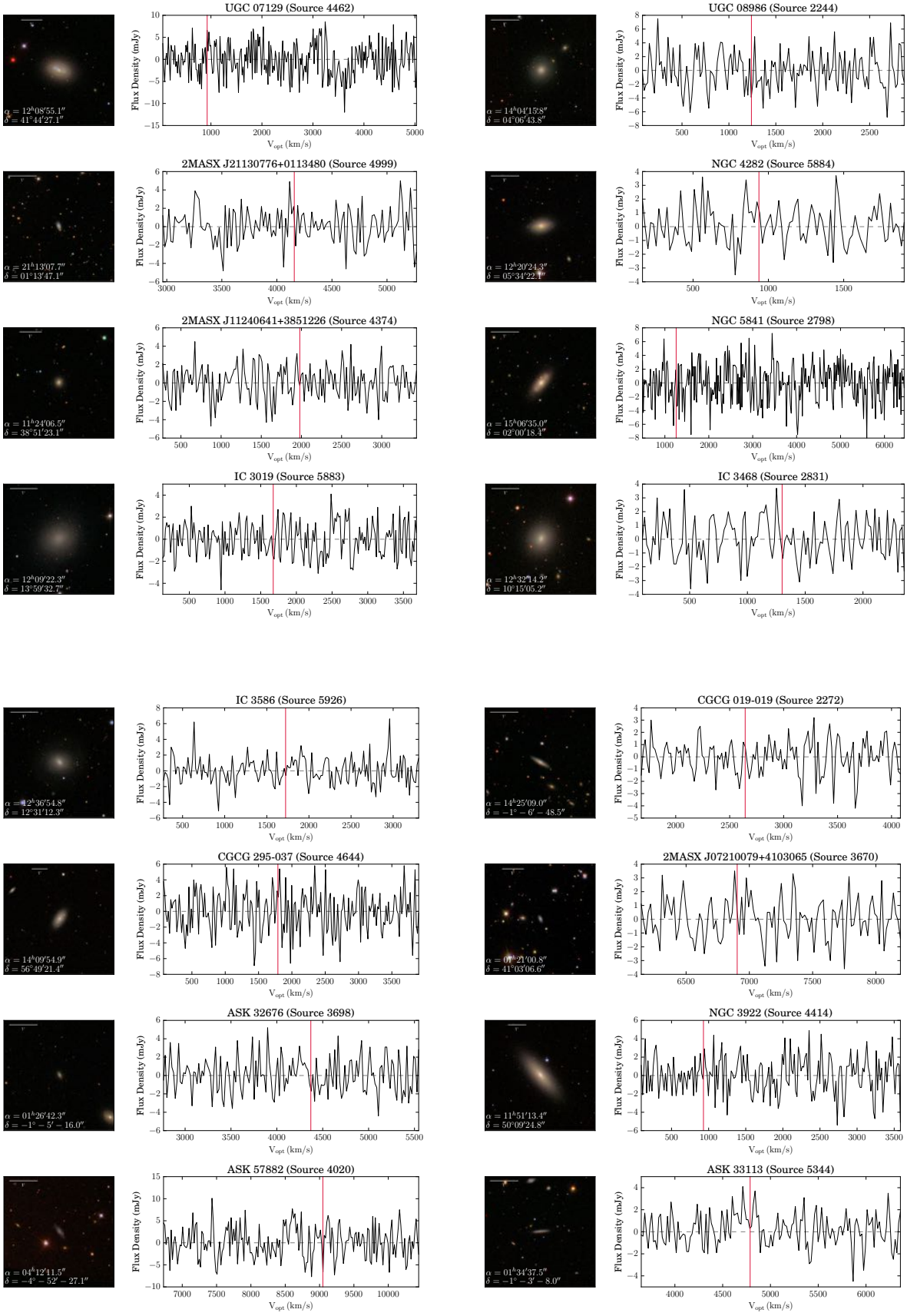


Figure C.1: *(continued) CRUMBS Stacking Sample*

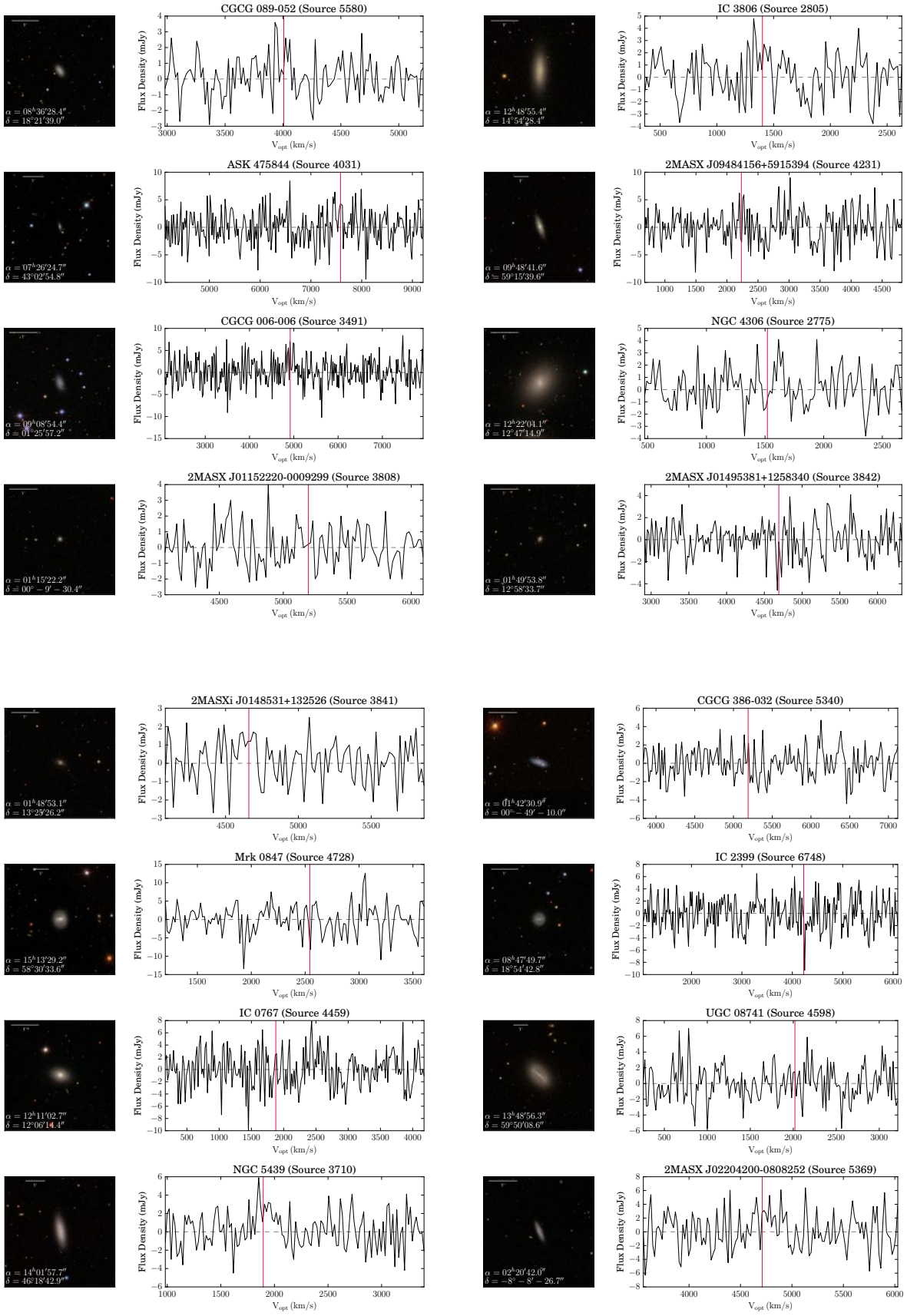


Figure C.1: *(continued) CRUMBS Stacking Sample*

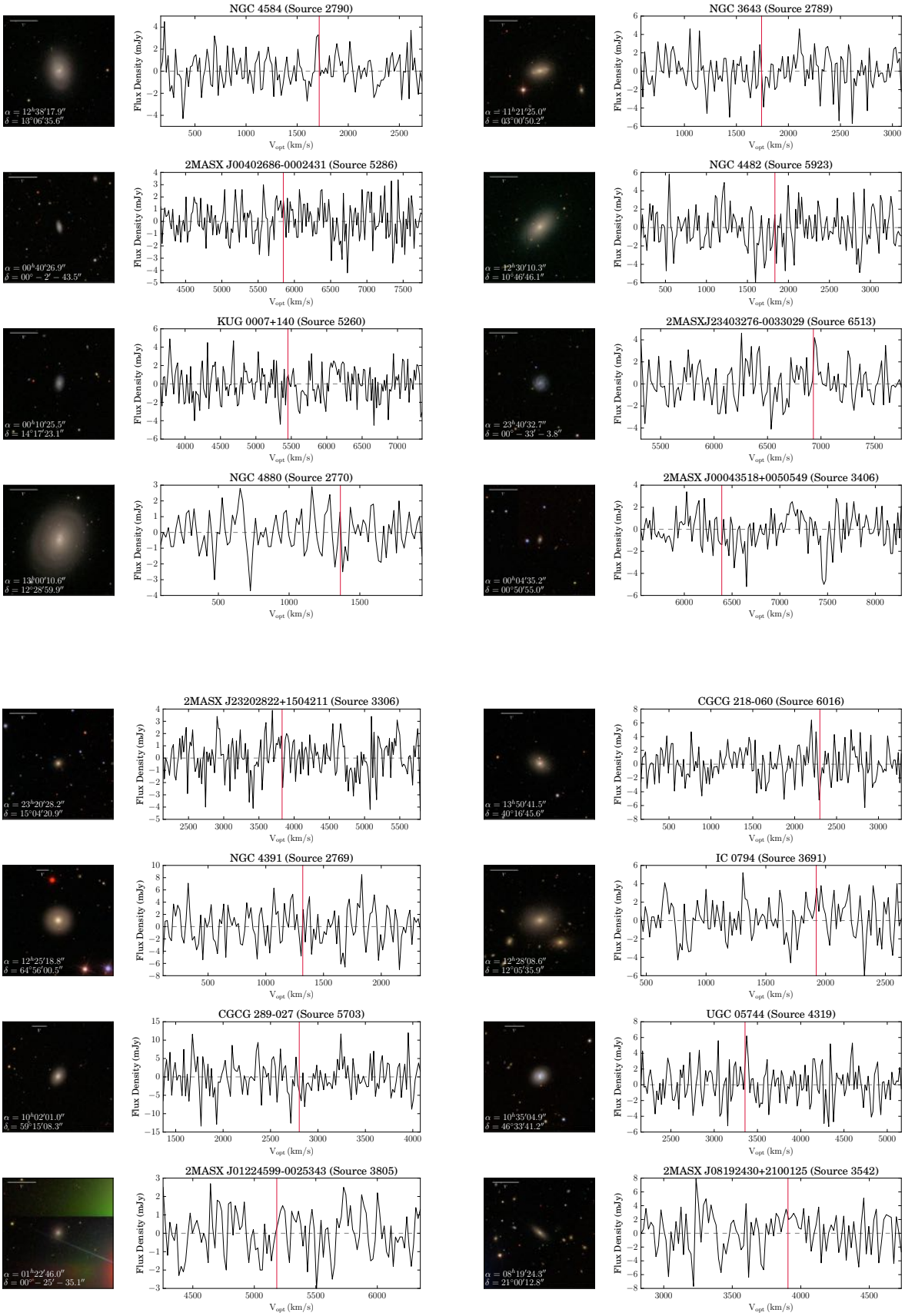


Figure C.1: *(continued) CRUMBS Stacking Sample*

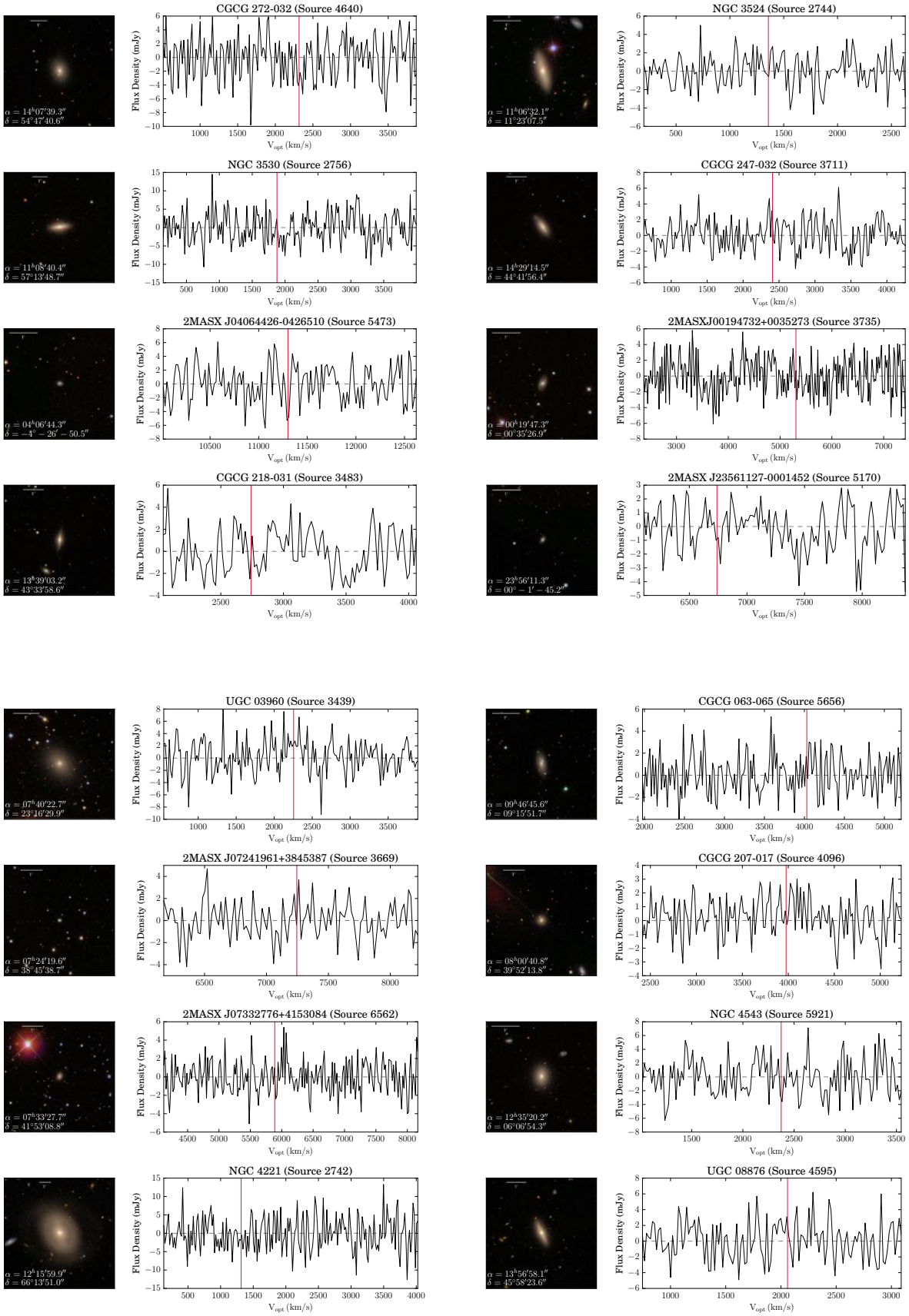


Figure C.1: *(continued) CRUMBS Stacking Sample*

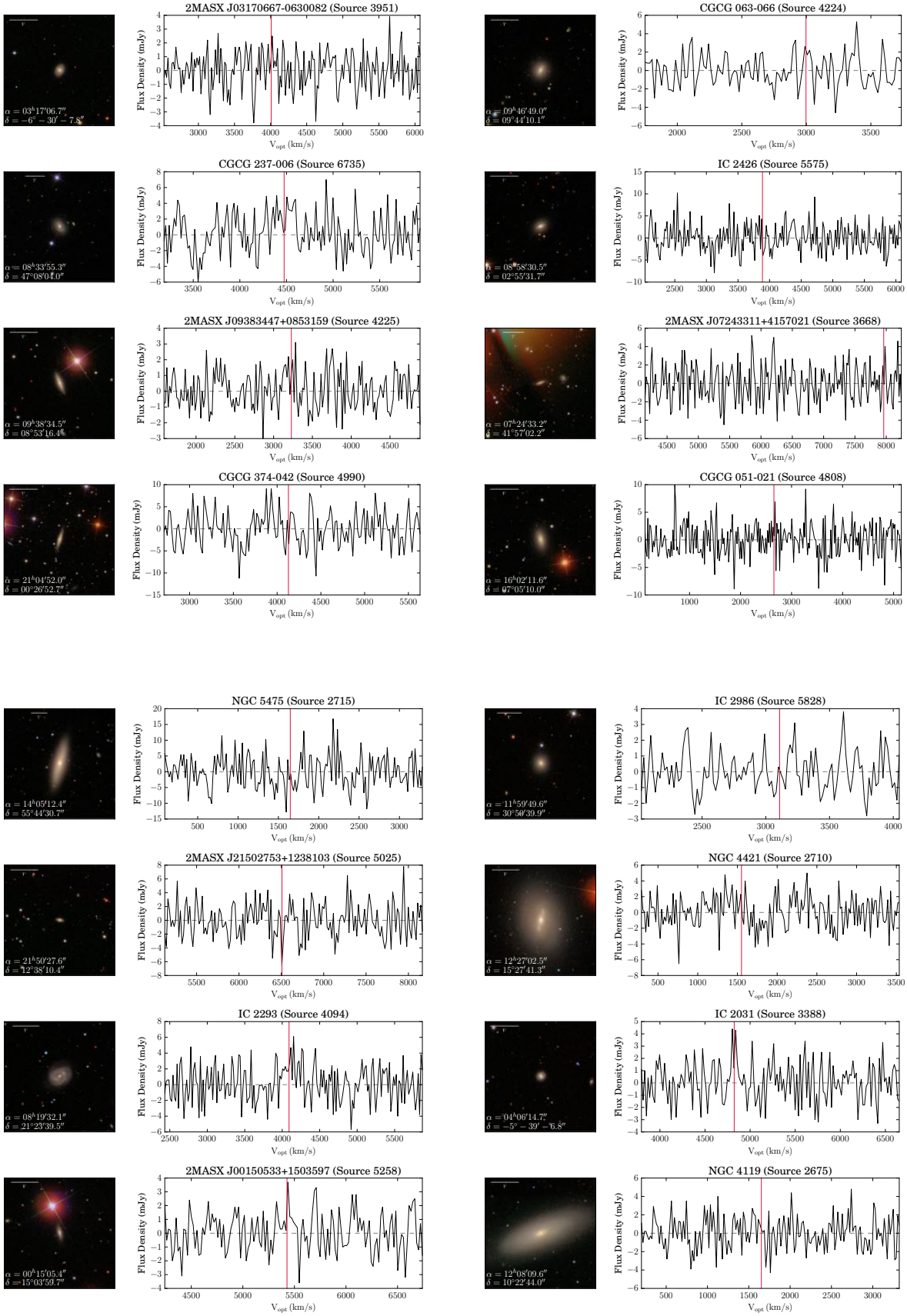


Figure C.1: *(continued) CRUMBS Stacking Sample*

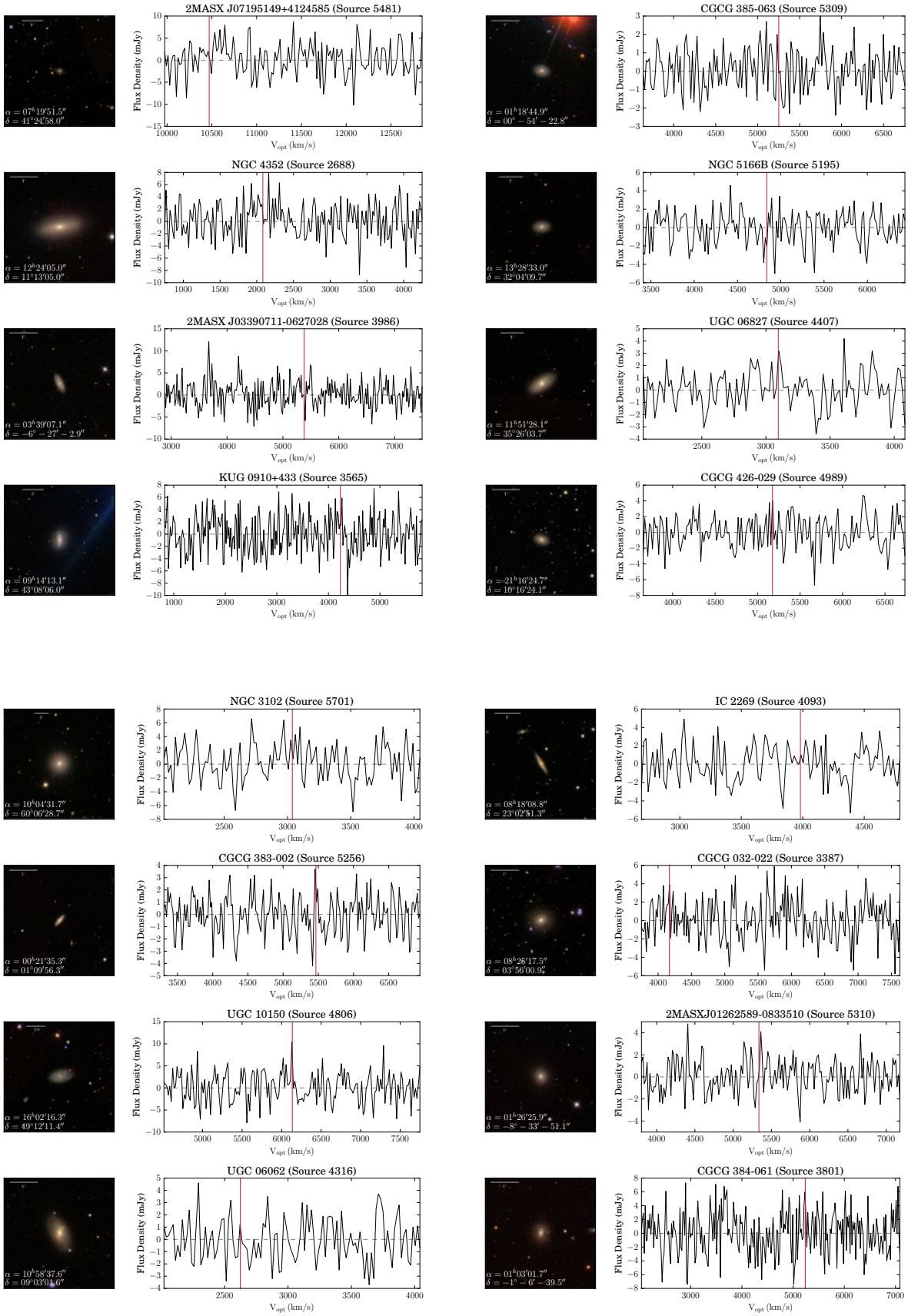


Figure C.1: *(continued) CRUMBS Stacking Sample*

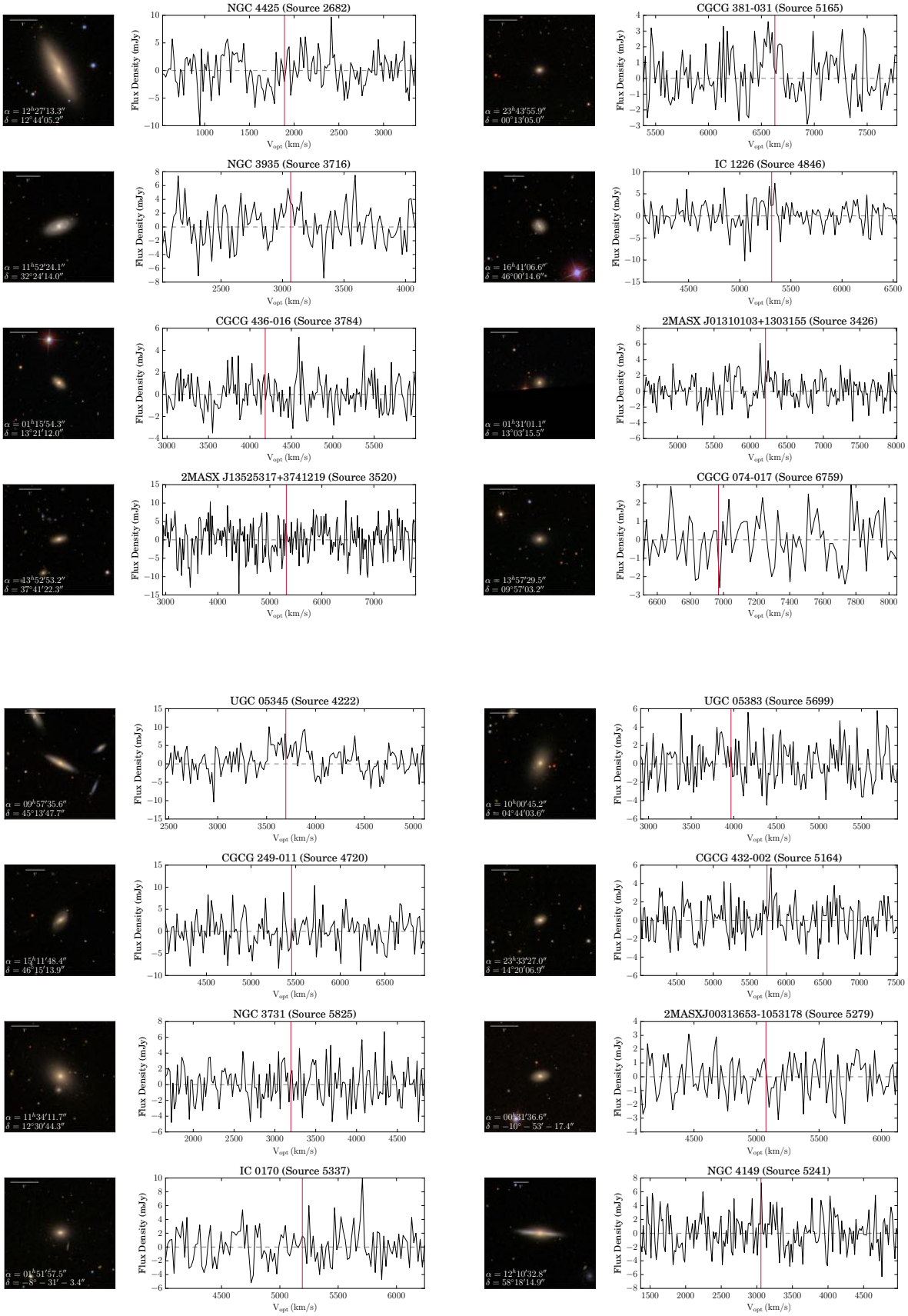


Figure C.1: *(continued) CRUMBS Stacking Sample*

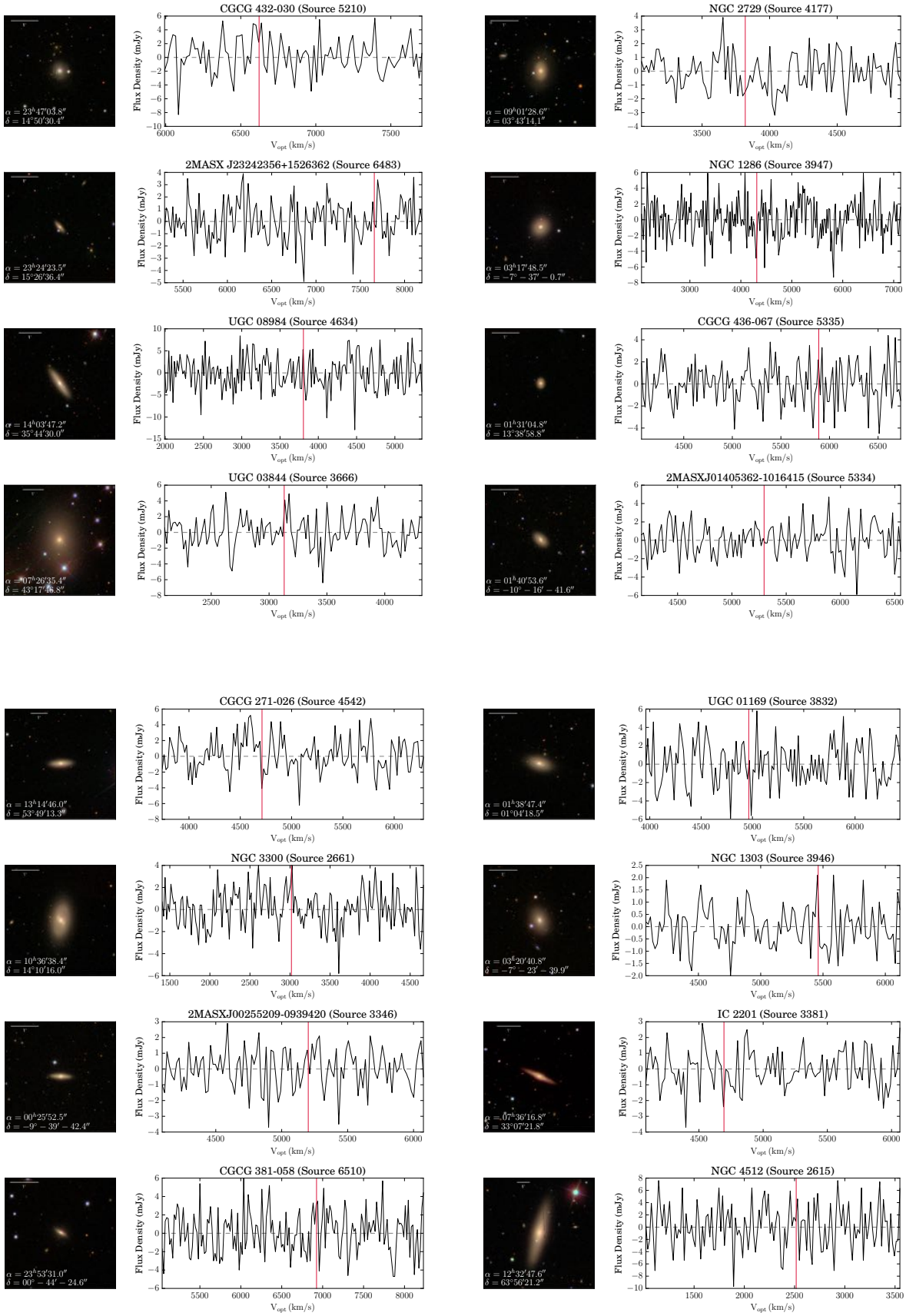


Figure C.1: *(continued) CRUMBS Stacking Sample*

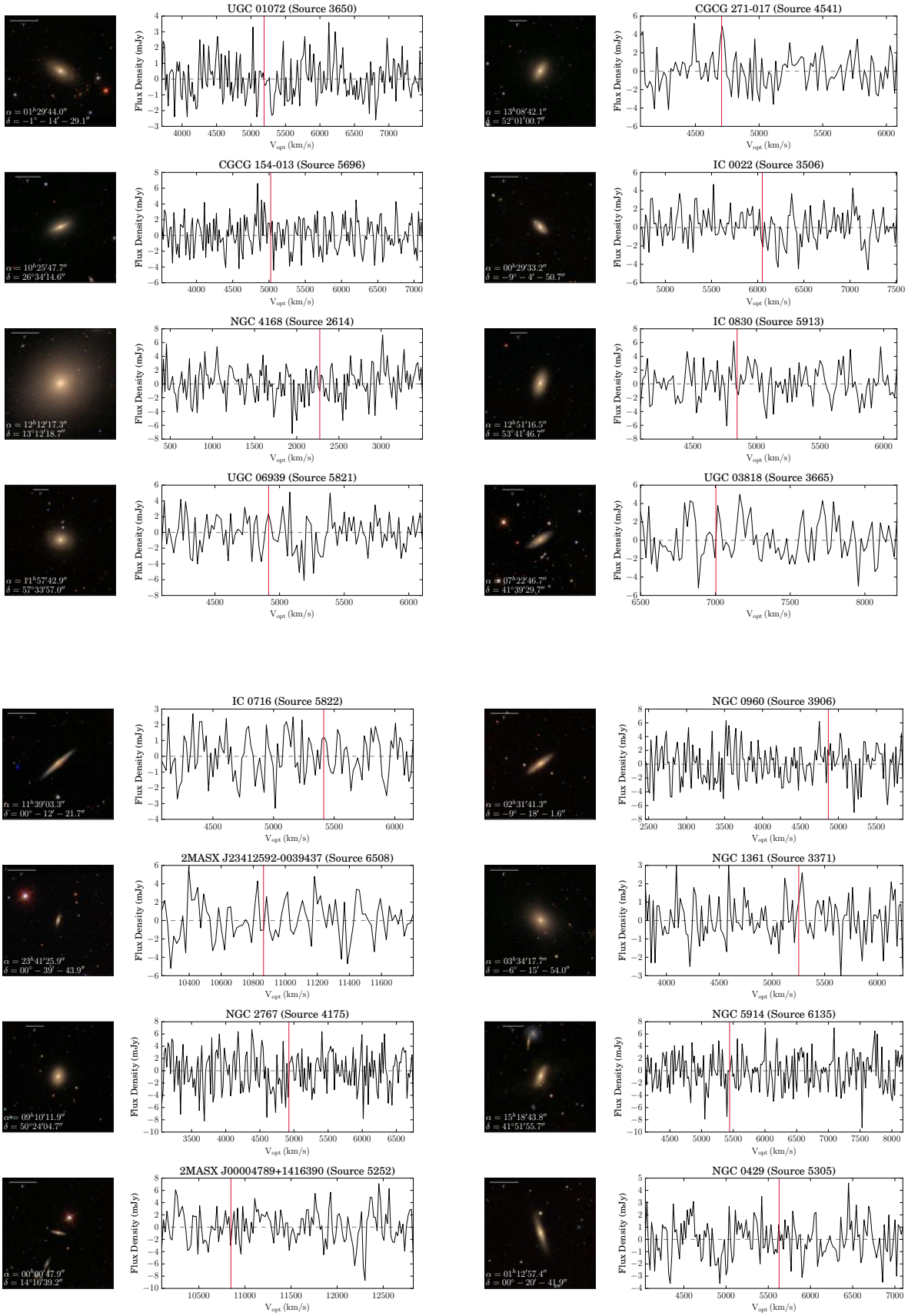


Figure C.1: *(continued) CRUMBS Stacking Sample*

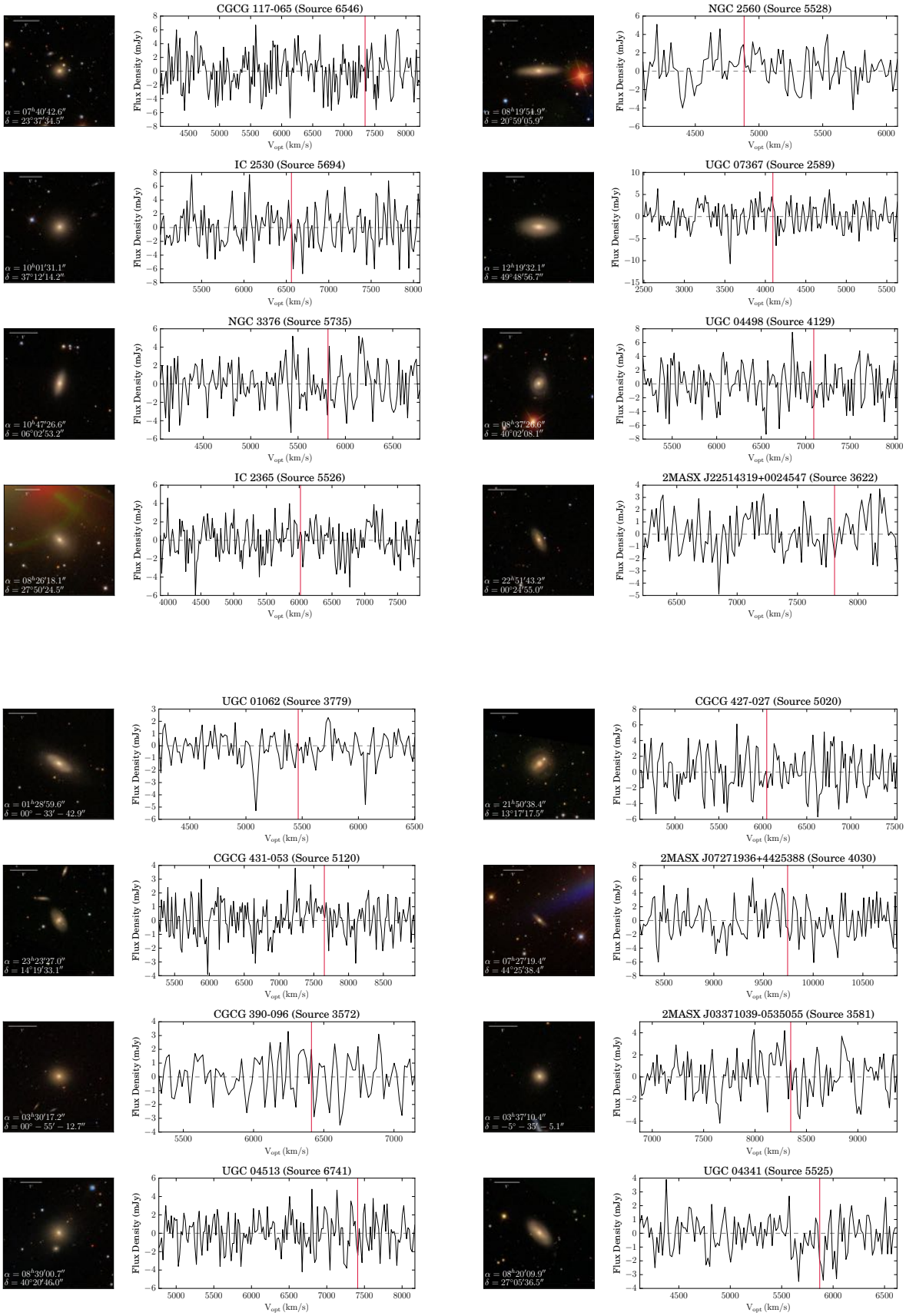


Figure C.1: *(continued) CRUMBS Stacking Sample*

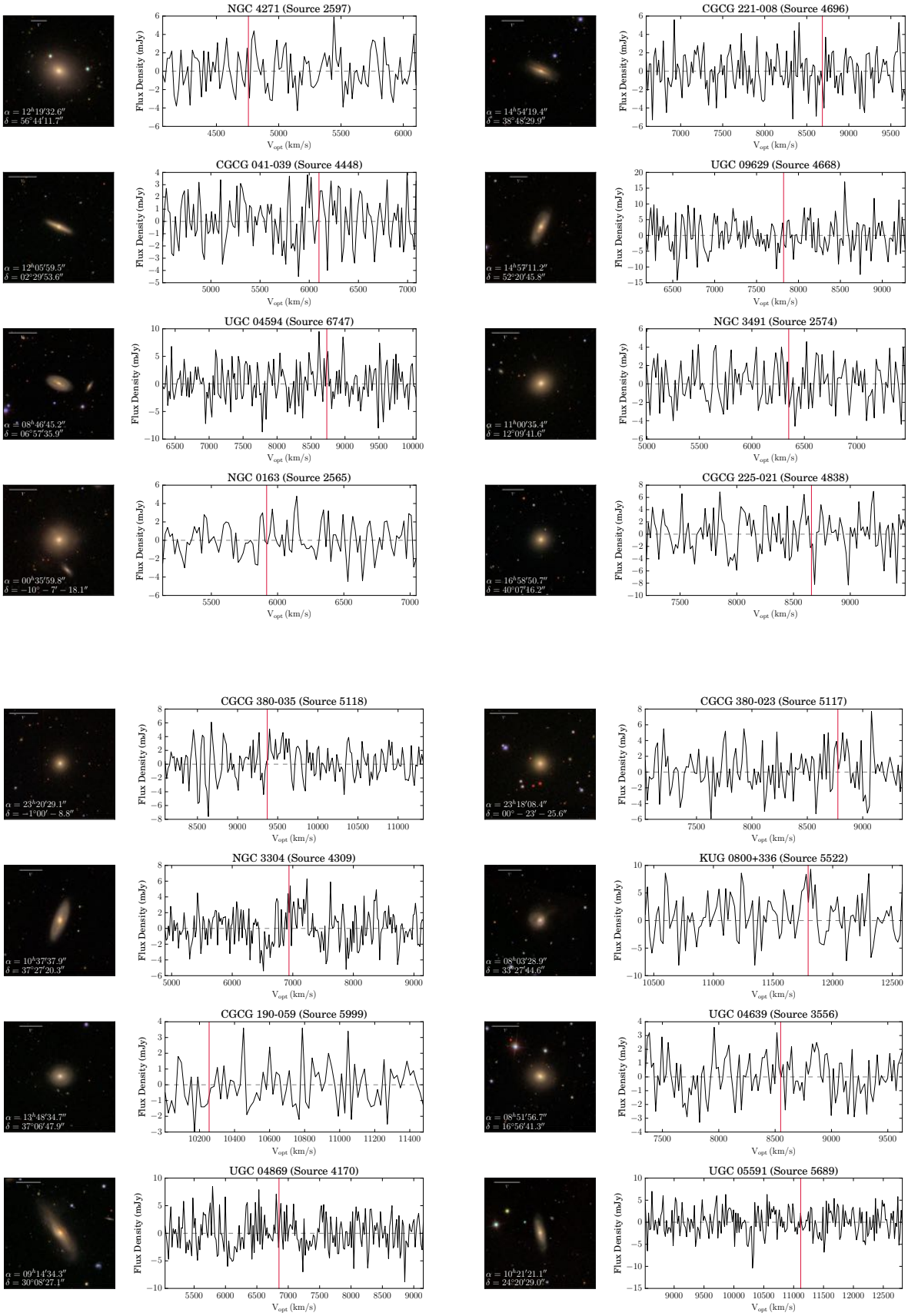


Figure C.1: *(continued) CRUMBS Stacking Sample*

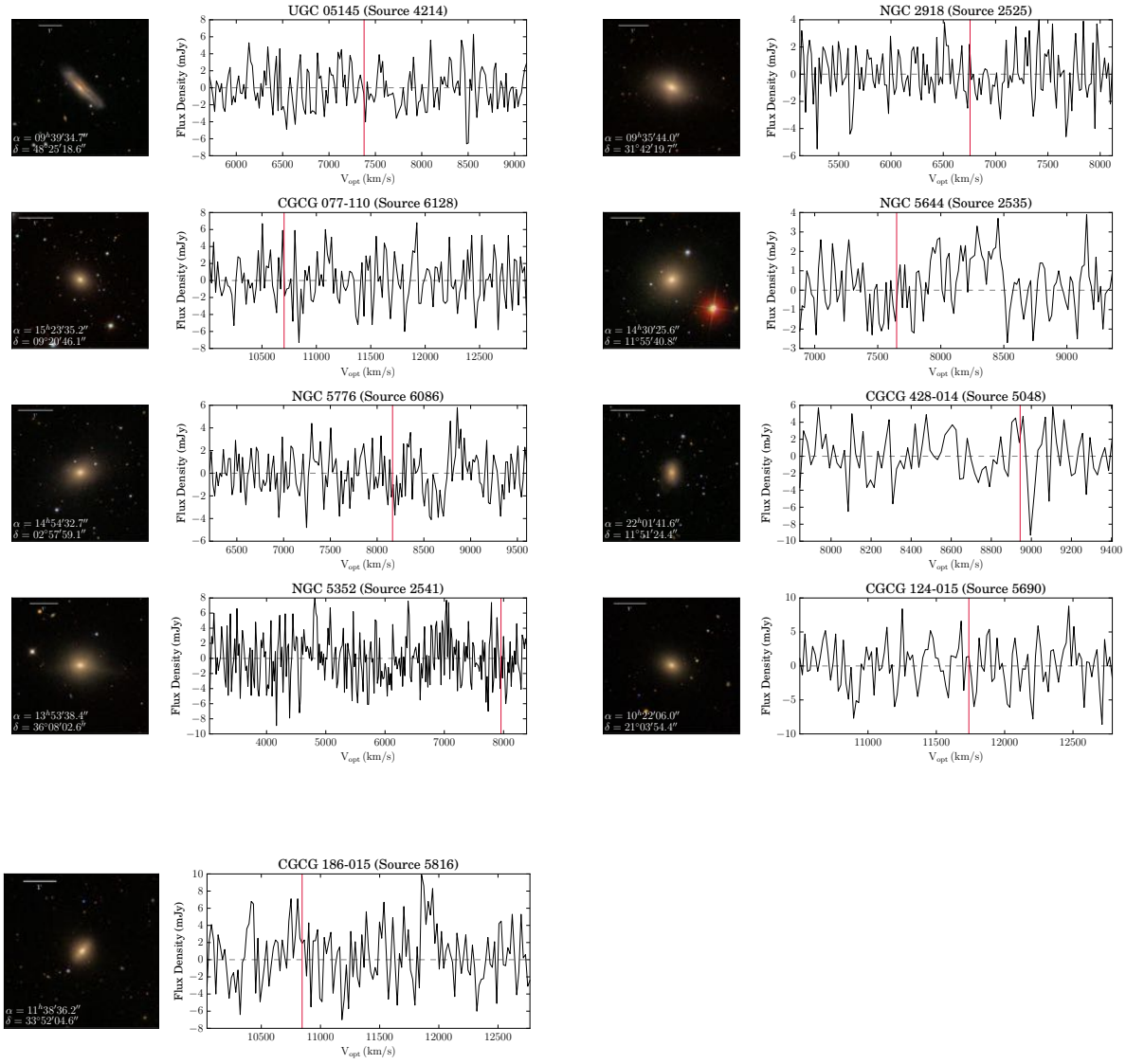


Figure C.1: *(continued) CRUMBS Stacking Sample*

Appendix D

Stacked Spectra

D.1 Stacking non-detections

D.1.1 Stacking in bins of colour

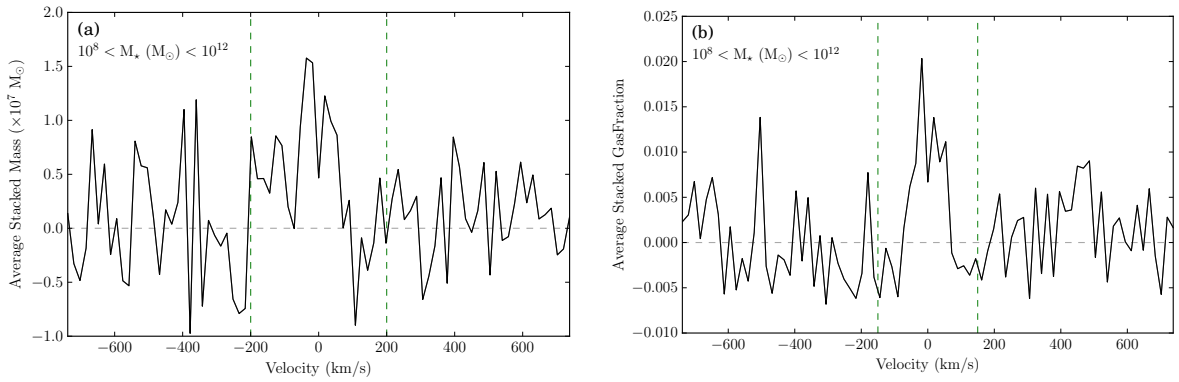


Figure D.1: *Stacked spectra for all non-detections with a stellar mass. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel.*

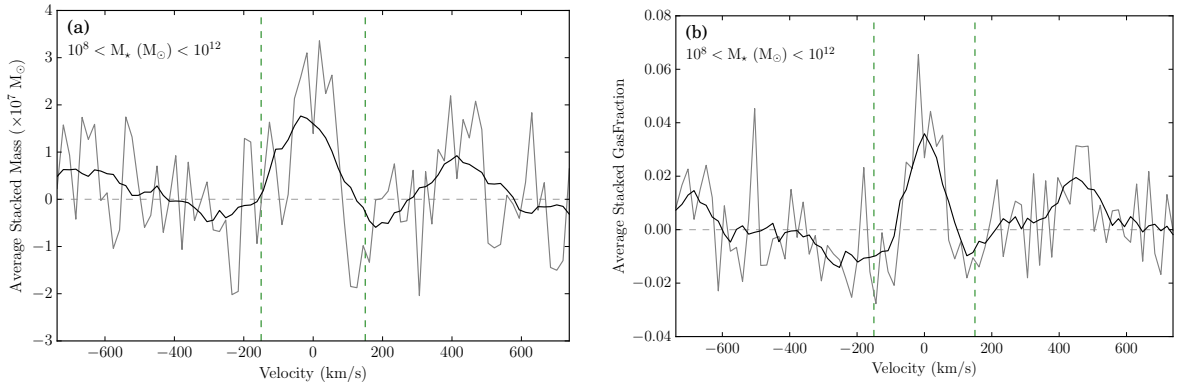


Figure D.2: *Stacked spectra for all blue non-detections with a stellar mass. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI} \rangle$ and $\langle f_{\rm HI} \rangle$ are calculated.*

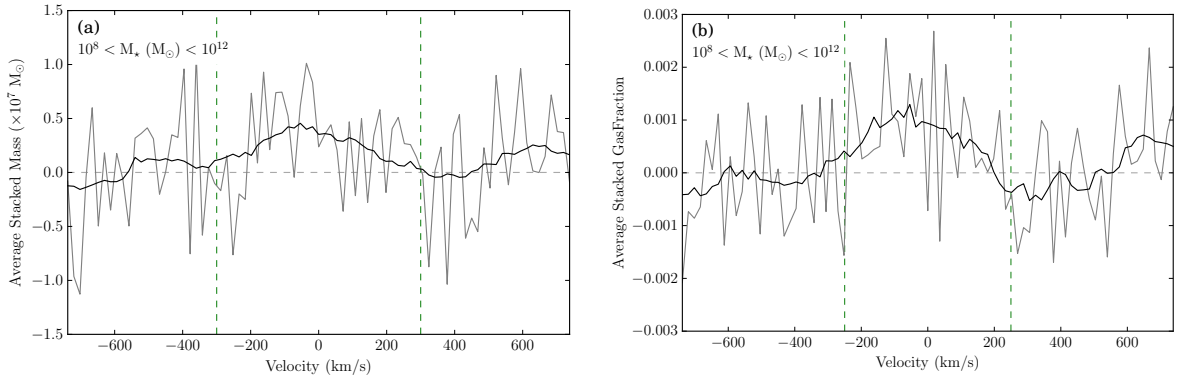


Figure D.3: *Stacked spectra for all red non-detections with a stellar mass. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI} \rangle$ and $\langle f_{\rm HI} \rangle$ are calculated.*

D.1.2 Stacking in bins of colour and stellar mass

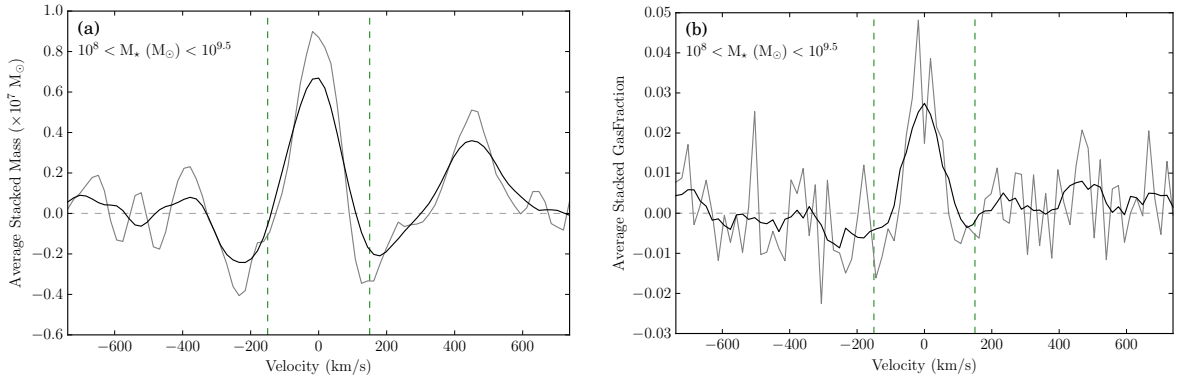


Figure D.4: *Stacked spectra for all non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI} \rangle$ and $\langle f_{\rm HI} \rangle$ are calculated.*

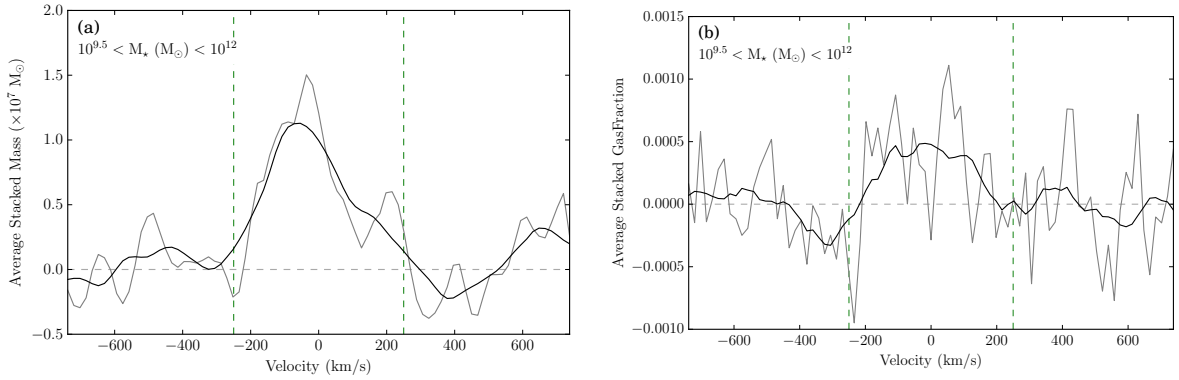


Figure D.5: *Stacked spectra for all non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI} \rangle$ and $\langle f_{\rm HI} \rangle$ are calculated.*

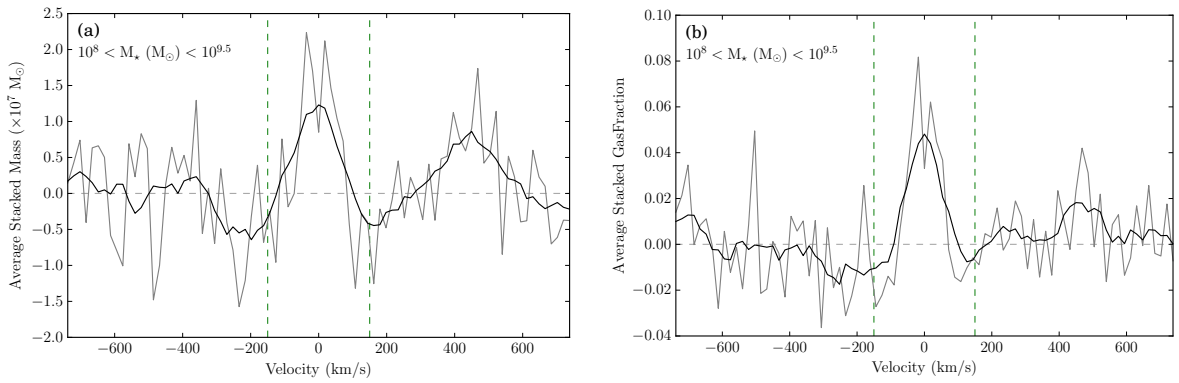


Figure D.6: *Stacked spectra for blue non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

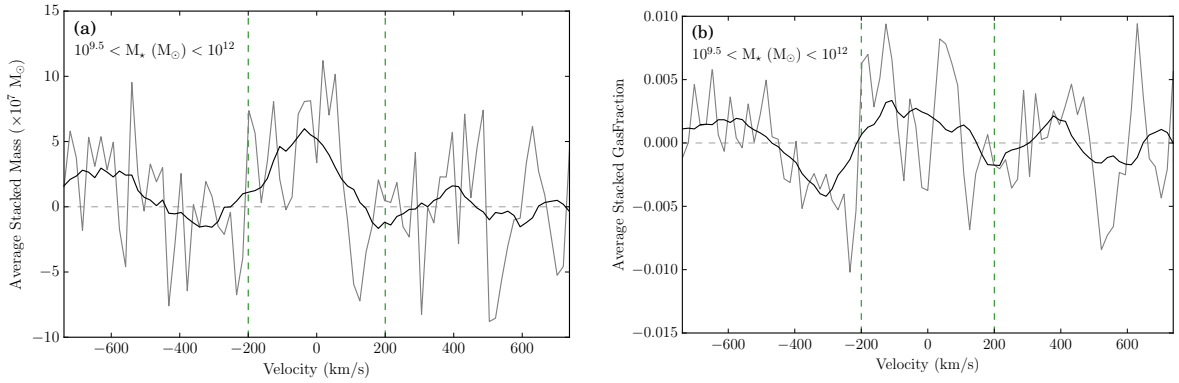


Figure D.7: *Stacked spectra for blue non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

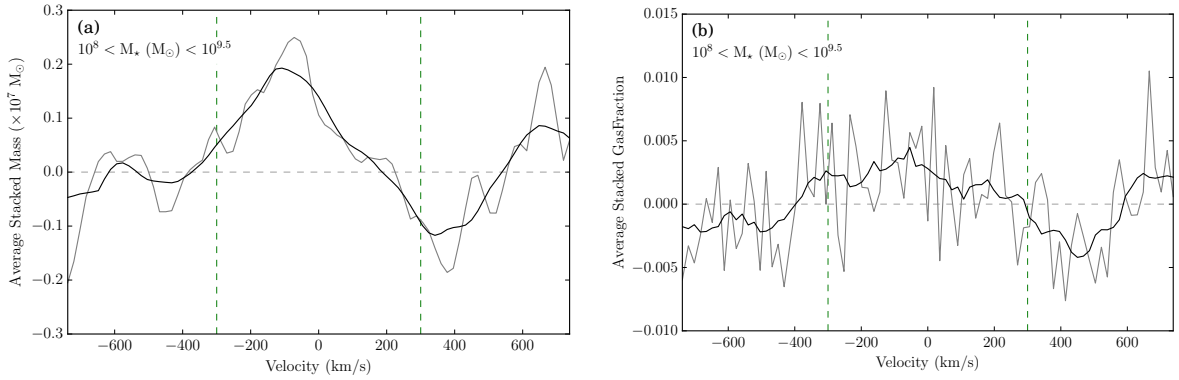


Figure D.8: *Stacked spectra for red non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

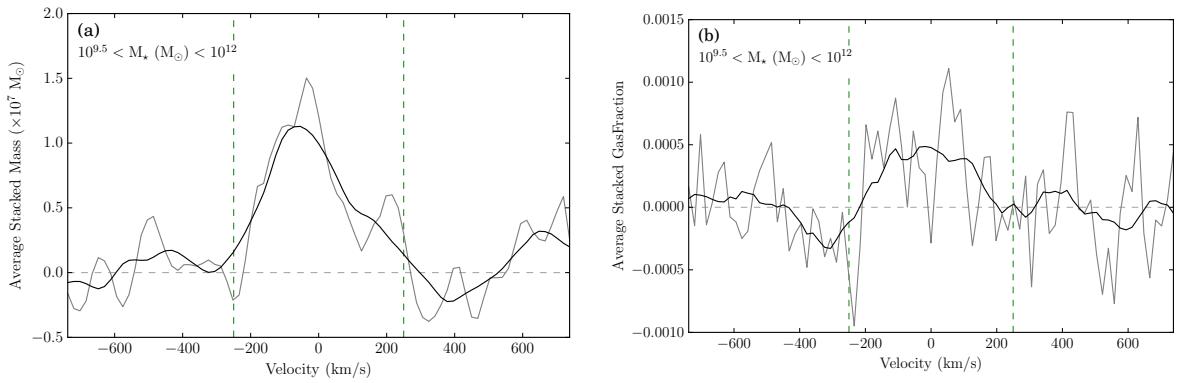


Figure D.9: *Stacked spectra for all non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The grey spectrum shows the unsmoothed stacked spectra, the black spectra represent the smoothed spectra from which $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

D.2 Stacking detections

D.2.1 Stacking in bins of colour

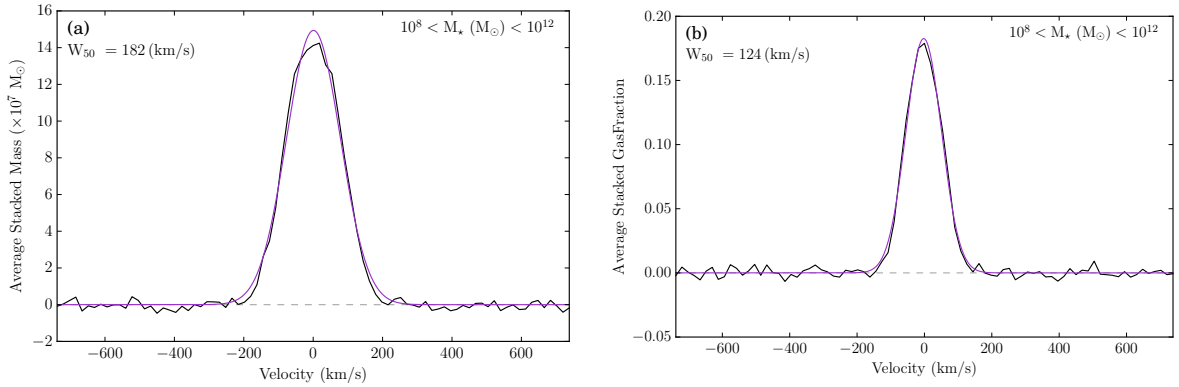


Figure D.10: *Stacked spectra for all blue detections with a stellar mass. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The purple line represents the single Gaussian fitted to the stacked spectrum. From this single Gaussian fit the width of the profile is calculated.*

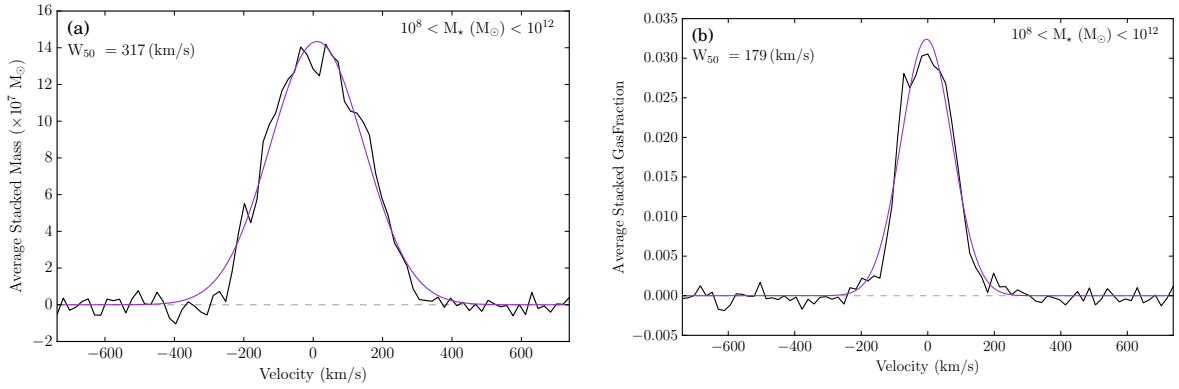


Figure D.11: *Stacked spectra for all red detections with a stellar mass. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The purple line represents the single Gaussian fitted to the stacked spectrum. From this single Gaussian fit the width of the profile is calculated.*

D.2.2 Stacking in bins of colour and stellar mass

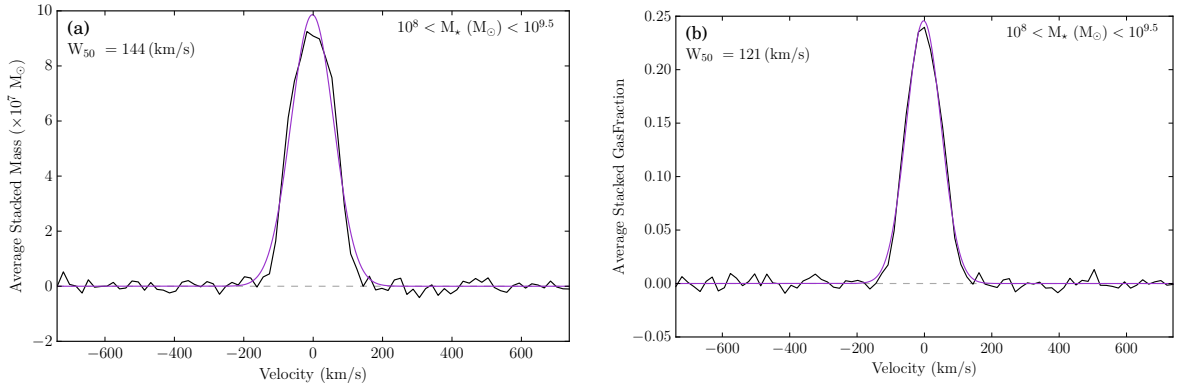


Figure D.12: *Stacked spectra for blue non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. The purple line represents the single Gaussian fitted to the stacked spectrum. From this single Gaussian fit the width of the profile is calculated.*

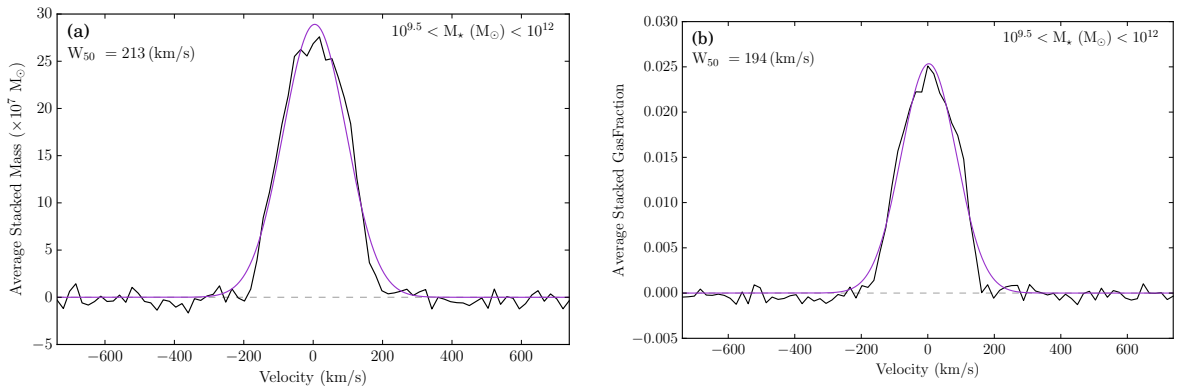


Figure D.13: *Stacked spectra for blue detections with a stellar mass in the range $10^{9.5}\,M_\odot < M_\star < 10^{12}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The purple line represents the single Gaussian fitted to the stacked spectrum. From this single Gaussian fit the width of the profile is calculated.*

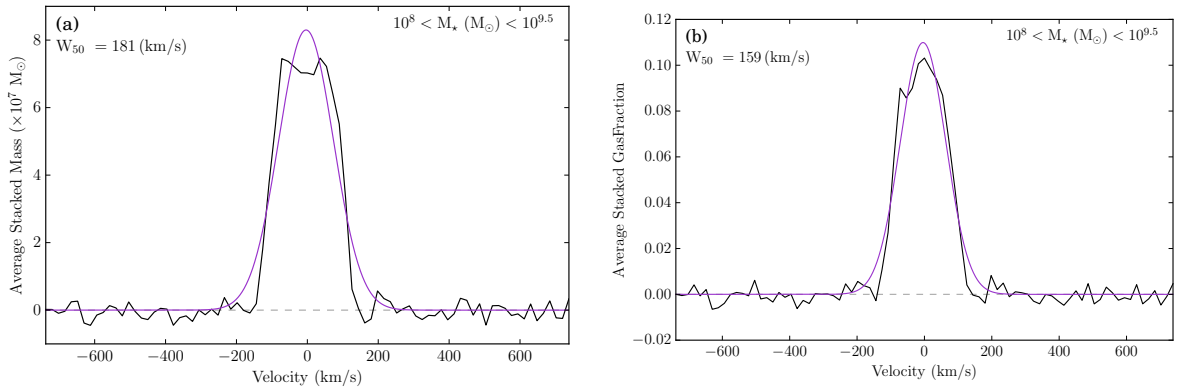


Figure D.14: *Stacked spectra for red detections with a stellar mass in the range $10^{8}\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The purple line represents the single Gaussian fitted to the stacked spectrum. From this single Gaussian fit the width of the profile is calculated.*

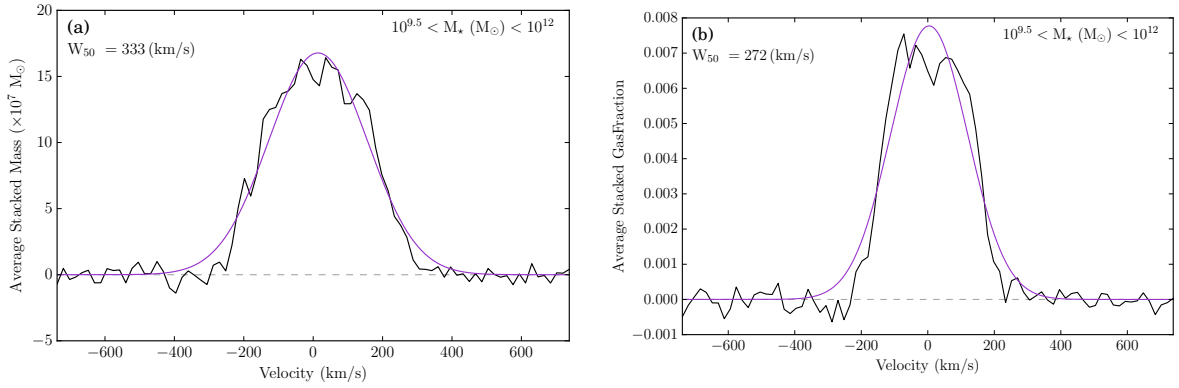


Figure D.15: *Stacked spectra for red detections with a stellar mass in the range $10^{9.5}\,M_\odot < M_\star < 10^{12}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The purple line represents the single Gaussian fitted to the stacked spectrum. From this single Gaussian fit the width of the profile is calculated.*

D.3 Stacking detections and non-detections

D.3.1 Stacking in bins of colour and stellar mass

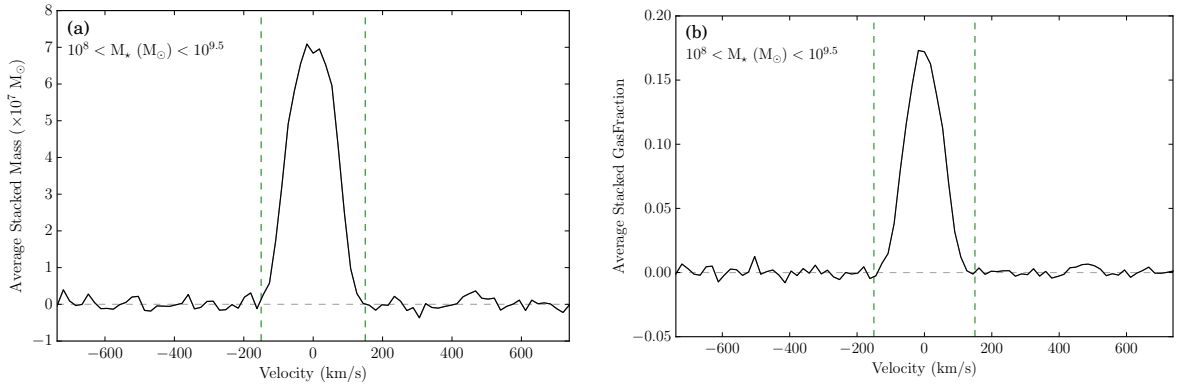


Figure D.16: *Stacked spectra for all detections and non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The dashed vertical lines denote the integration range used to calculate $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

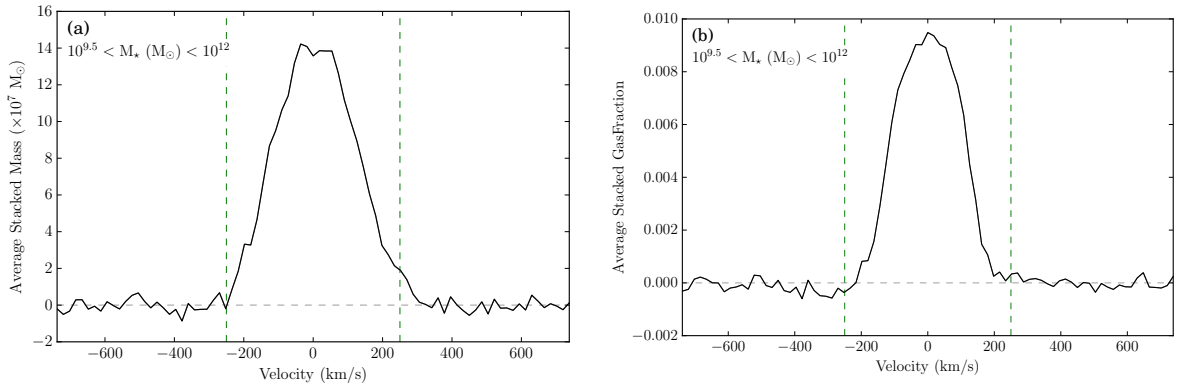


Figure D.17: *Stacked spectra for all detections and non-detections with a stellar mass in the range $10^{9.5}\,M_\odot < M_\star < 10^{12}\,M_\odot$. Panel* **a** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* **b** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The dashed vertical lines denote the integration range used to calculate $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

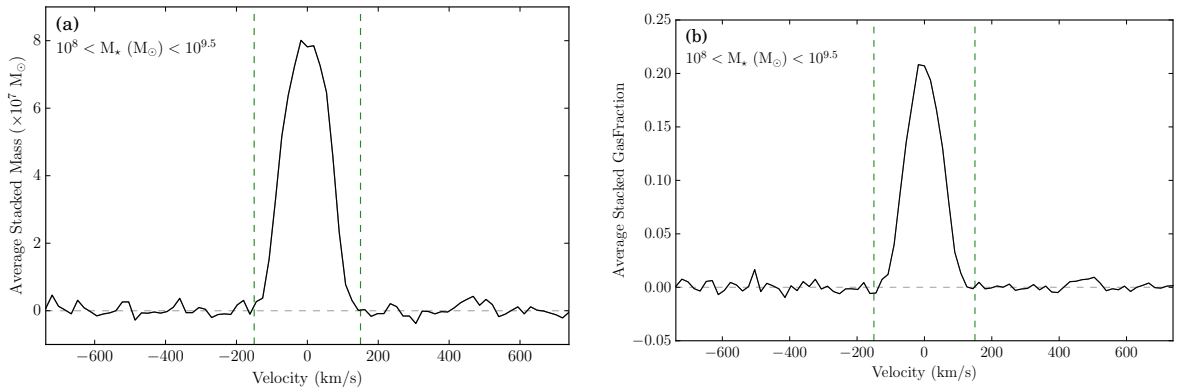


Figure D.18: *Stacked spectra for blue detections and non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* ***a*** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* ***b*** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The dashed vertical lines denote the integration range used to calculate $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

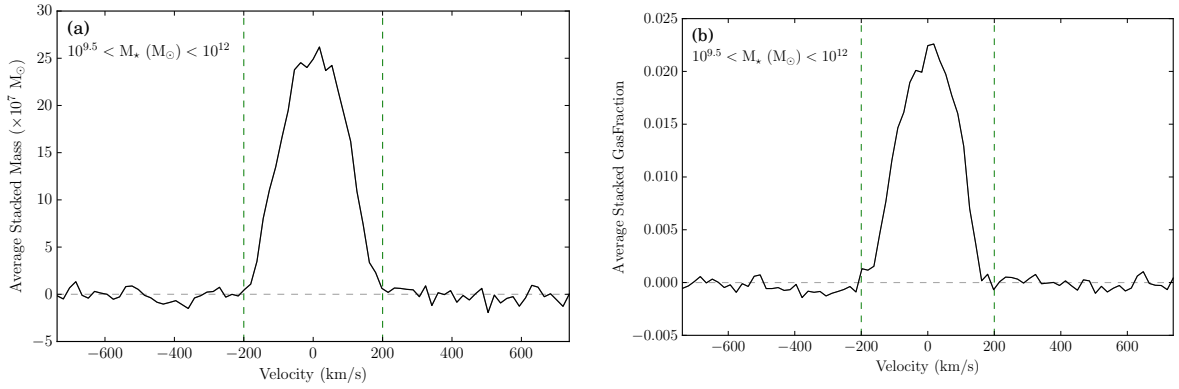


Figure D.19: *Stacked spectra for blue detections and non-detections with a stellar mass in the range $10^{9.5}\,M_\odot < M_\star < 10^{12}\,M_\odot$. Panel* ***a*** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* ***b*** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The dashed vertical lines denote the integration range used to calculate $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

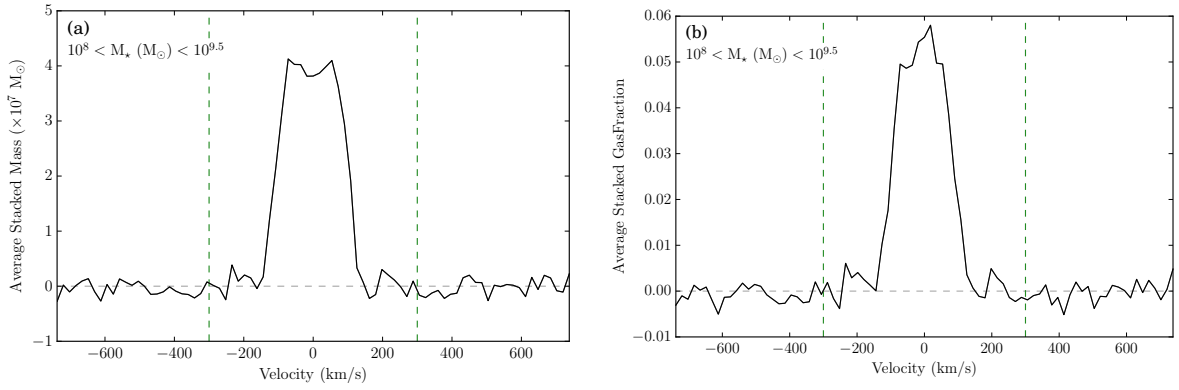


Figure D.20: *Stacked spectra for red detections and non-detections with a stellar mass in the range $10^8\,M_\odot < M_\star < 10^{9.5}\,M_\odot$. Panel* ***a*** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* ***b*** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The dashed vertical lines denote the integration range used to calculate $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$ are calculated.*

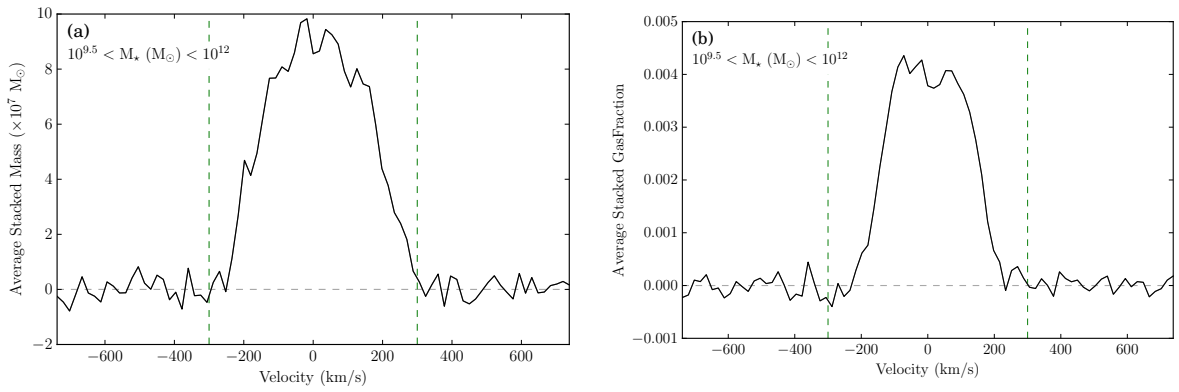


Figure D.21: *Stacked spectra for red detections and non-detections with a stellar mass in the range $10^{9.5}\,M_\odot < M_\star < 10^{12}\,M_\odot$. Panel* ***a*** *shows the average stacked spectrum in units of $M_{\rm HI}$ per channel. Panel* ***b*** *shows the average stacked spectrum in units of $f_{\rm HI}$ per channel. The dashed vertical lines denote the integration range used to calculate $\langle M_{\rm HI}\rangle$ and $\langle f_{\rm HI}\rangle$.*

D.3.2 Stacking in bins of stellar mass

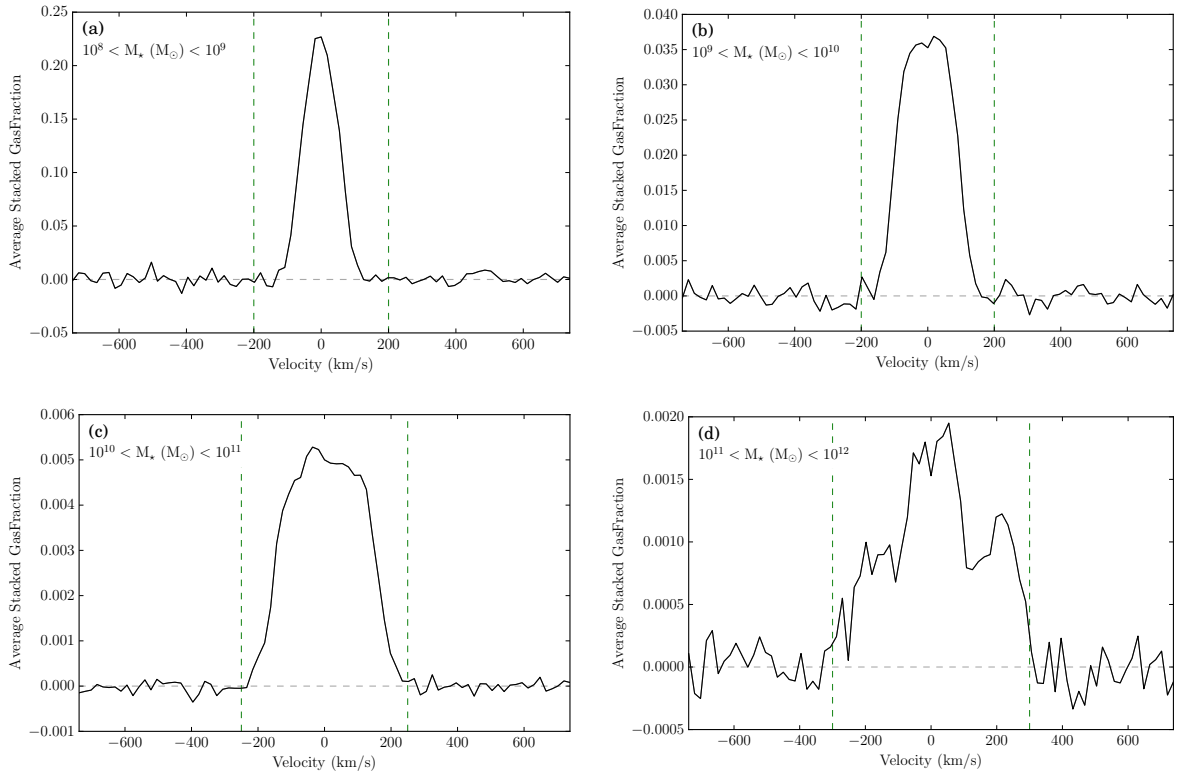


Figure D.22: *All detections and non-detections stacked in bins of stellar mass. The bins increase in stellar mass from panel (a) to (d). The dashed vertical lines denote the range over which the spectra are integrated.*

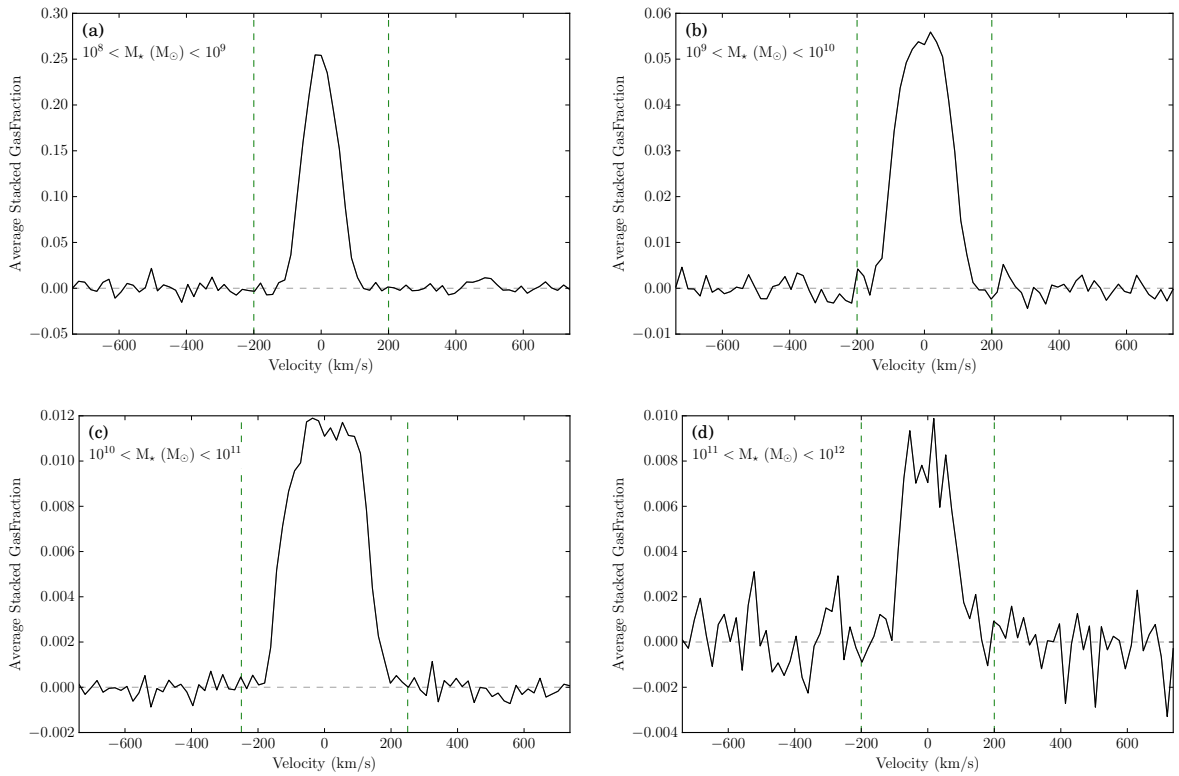


Figure D.23: *Blue detections and non-detections stacked in bins of stellar mass. The bins increase in stellar mass from panel (a) to (d). The dashed vertical lines denote the range over which the spectra are integrated.*

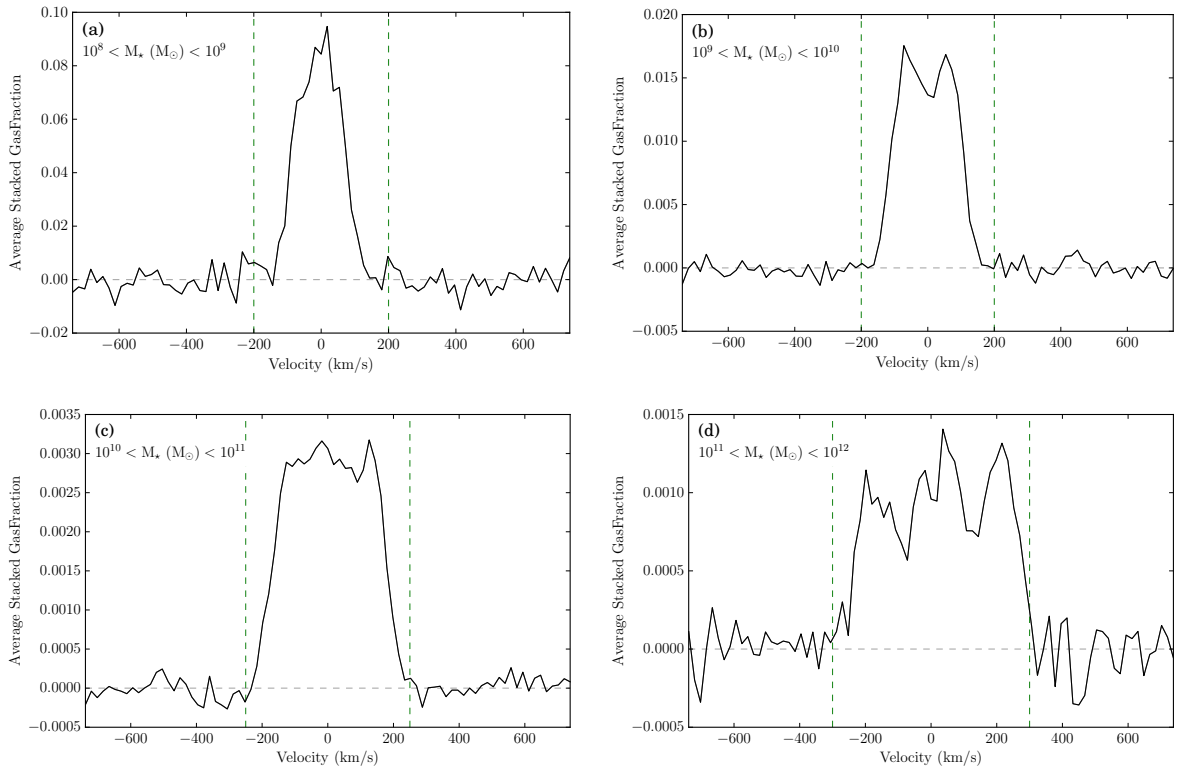


Figure D.24: *Red detections and non-detections stacked in bins of stellar mass. The bins increase in stellar mass from panel (a) to (d). The dashed vertical lines denote the range over which the spectra are integrated.*

D.3.3 Stacking in bins of $NUV - r$ colour

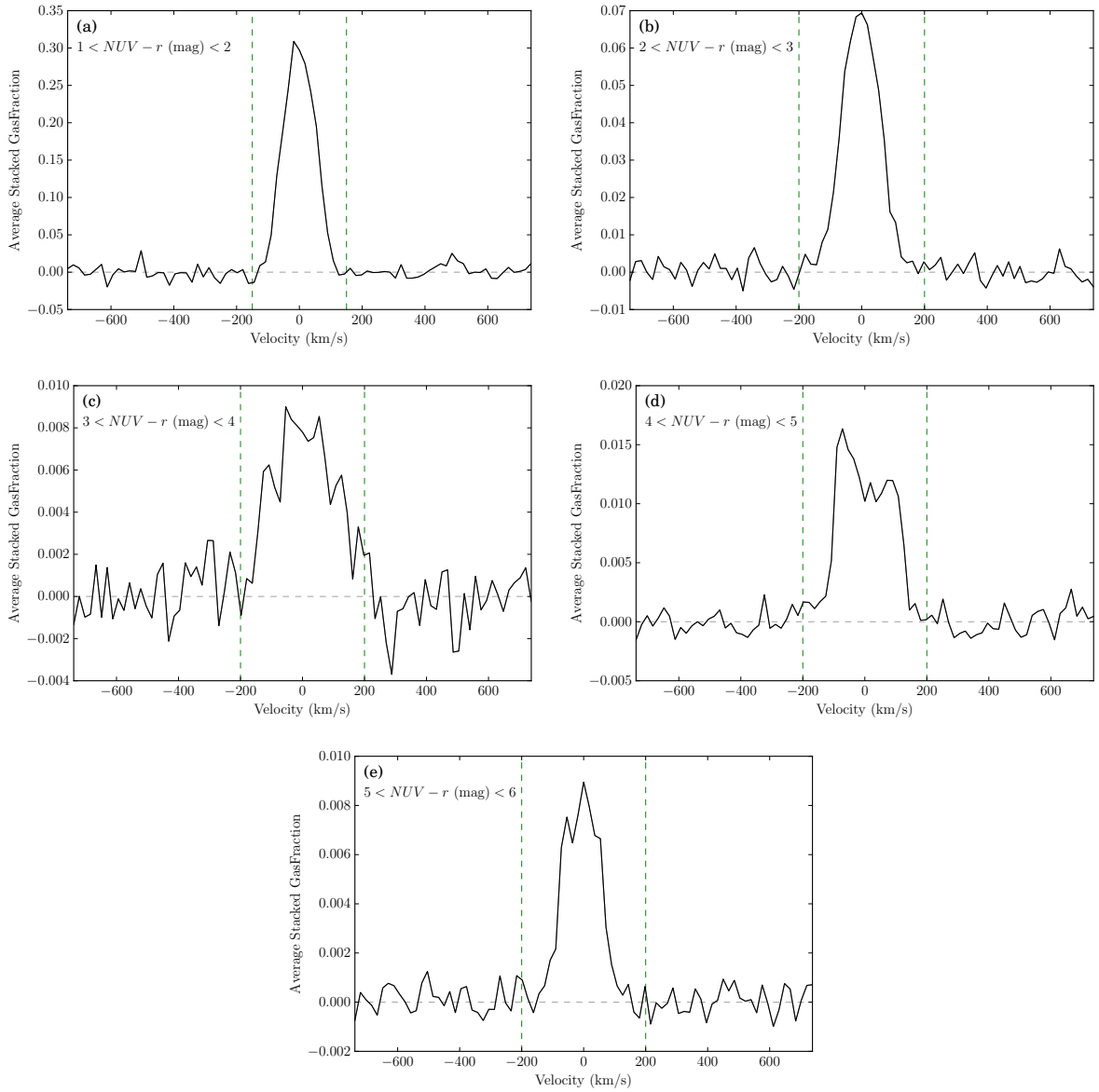


Figure D.25: *All detections and non-detections stacked in bins of $NUV - r$ colour. The bins increase in $NUV - r$ colour from panel (a) to (e). The dashed vertical lines denote the range over which the spectra are integrated.*