Investigating High Velocity HI Clouds in the Milky Way Halo

By

Natasha Marshall

A thesis submitted to Macquarie University

for the degree of Master of Research

Department of Physics and Astronomy

December 2019

MACQUARIE
University
SYDNEY·AUSTRALIA

Examiner's Copy

© Natasha Marshall, 2019.

Typeset in LaTeX 2_ε.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Natasha Marshall

Acknowledgements

To the two best supervisors Dr Joanne Dawson and Dr Vanessa Moss, thank you from the bottom of my heart. Your encouragement, support and guidance has renewed my love for astronomy and created a passion for research. Thank you to Anita for providing me with an abundance of statistical knowledge as well as unwavering parental support as we battle academia and motherhood. Thank you to Dad and Gemma for providing me with the best start in my life as well as supporting me and Finn. Mum, Daz, Lex and Vick for always being able to reduce my research to a punch line. To my best friends Alana, Micci and Simone for listening to me complain and celebrate every milestone. You too Cody and Kelly, you get a shout-out for coming to Outback. But most of all, Arden and Finn for keeping me sane for the past year.

Abstract

Understanding the gas exchange dynamics in the Milky Way galaxy is important in understanding Galactic evolution and the ongoing fuelling of star formation. Determining the structure of neutral hydrogen high velocity clouds in the halo can provide insight into how the gas within the Halo feeds star formation in the Galaxy. We examine the 21cm HI emission spectra from 84 sightlines in the Milky Way halo using data from both the Southern Weak Intensity Survey of HI (SWISH) Moss and McClure-Griffiths [2017] and the Galactic All-Sky Survey (GASS) McClure-Griffiths et al. [2009]. This thesis presents a semi automatic calibration method for analysis of SWISH sightlines, with the inclusion of Bayesian Gaussian modelling. We find a population of low column density gas previously undetected by GASS and discuss these preliminary results in the context of the overall structure of the neutral gas in the Galactic Halo.

Contents

Science, for me, gives a partial explanation for life. In so far as it goes, it is based on fact, experience and experiment.

Dr Rosalind Franklin

1 Introduction

1.1 Anatomy of the Gaseous Milky Way

The Milky Way Galaxy is a barred spiral galaxy with an age of approximately 13 billion years [Schlaufman et al., 2018]. The disk of the Galaxy contains most of the stars, with the star-forming gas and newly-formed stars mainly concentrated within spiral arms extending around the circumference of the disk. The rarefied matter between stars is known as the Interstellar Medium (ISM) and is comprised mostly of gas (99% by mass) and dust (1% by mass) which exists in various states primarily characterised by density and temperature [Boulanger et al., 2000]. Beyond the disk of the Galaxy lies the Halo which extends many kilo-parsecs from the Galactic centre [Lockman, 1984] and is made up of both stars and ISM. This thesis focuses on the ISM component of the Halo. The location and extent of the Halo can be conveniently visualised via its stellar component, which can be seen in Figure 1.1.

Within the Galaxy, matter and energy are exchanged on large scales which can be thought of as one aspect of a "Galactic ecosystem". The Halo of our Galaxy contains material that

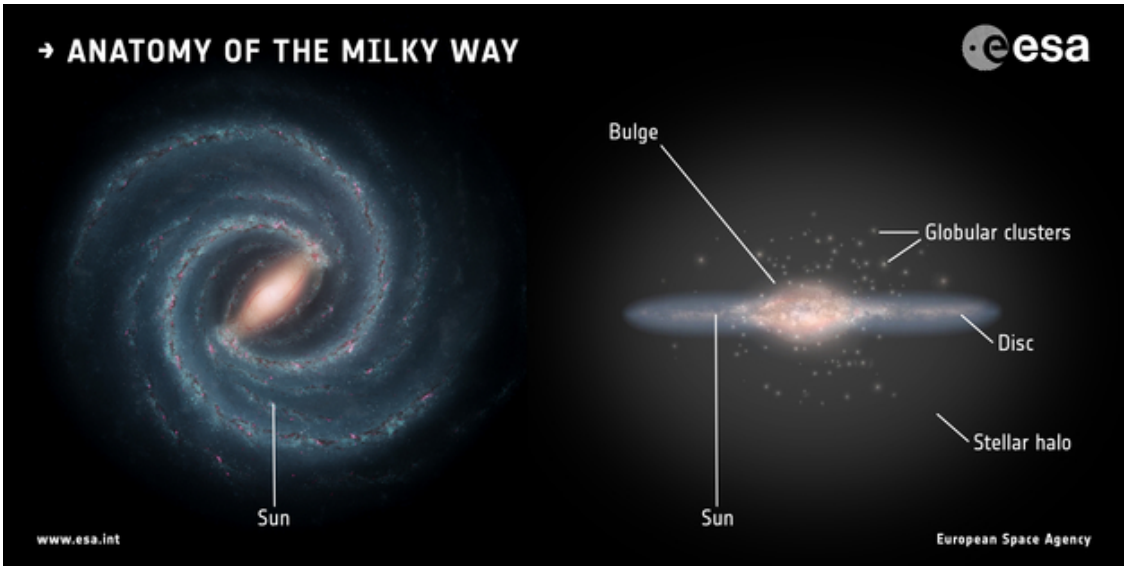


Figure 1.1: This illustration shows an artists impression of the structure of the Milky Way galaxy showing both the face on and edge on view. The stellar component of the halo of the Galaxy can be seen as a faint "glow" in a spherical shape encompassing the disc. Illustration credit: Left: NASA/JPL-Caltech; right: ESA; layout: ESA/ATG medialab.

has been accreted from both within the Galaxy and from extragalactic sources [Freeman and Bland-Hawthorn, 2002]. HI structures, such as those investigated in McClure-Griffiths et al. [2006] and Westmeier et al. [2013], provide a snapshot in time of the mechanisms of this Galactic ecosystem, as material moves between different structures of the Galaxy. The evolution of the Galaxy is underpinned by the creation of new stars – which form from gas reservoirs – so an understanding of ongoing gas supply (from the Halo to the star-forming disk) is important to explain the continued star formation in our Galaxy. Therefore it is of great interest to track the gas in-flowing to and out-flowing from the Halo at any given time.

Within the Halo there exists gas with observed line-of-sight velocity (as inferred from the Doppler shifting of spectral lines) that deviates from the values permitted by Galactic rotation, indicating material that does not form part of the normally-rotating Disk. This was was first observed by Muller et al. [1963] as *hydorgène neutre dans la couronne galactique*, which translates to "Neutral Hydrogen in the Galactic Halo", and proposed origins for this gas included hypotheses such as galactic mergers and accreted ejected material from the disk of the Galaxy [Oort, 1966]. This gas is now commonly referred to as "anomalous gas" and it may have a role in supplying gas reservoirs for continued star formation. The origins of anomalous gas are the subject of ongoing research.

1.2 Halo Clouds and their Origins

When a region of gas is denser than the ISM that surrounds it, it is often referred to as an interstellar cloud. Anomalous gas can be found in such isolated cloud-like structures which are categorised depending on their velocities relative to the local standard frame of rest (LSR). The three main categories are low velocity clouds (LVCs), intermediate velocity clouds (IVCs) and high velocity clouds. There are no strict separations between these categories, however for the purpose of this thesis, HVCs are categorised as clouds with $|v_{LSR}| \geq 90$ km s^{-1} (for observations off of the Galactic Plane, so their velocities can not be explained by normal Galactic rotation, as previously mentioned). This aligns with previous works on HVCs such as Wakker and van Woerden [1991] and Putman et al. [2002].

The first review of HVCs and their potential origins was undertaken by Oort [1966] after the discovery of anomalous gas. During this review, origins were discussed for all categories of anomalous gas, which have since been thoroughly investigated through different research efforts. By 1991 it was clear that specific HVCs represented the outflow of material from the Galactic nucleus [see review by Vogelaar et al. [1991]]. One of the initial origin theories suggested that supernovae could be a possible source of HVCs. Recent investigations [Ford et al., 2008, de Gouveia Dal Pino et al., 2009, Ford et al., 2010] revealed that only IVCs could be attributed to supernovae where as McClure-Griffiths et al. [2013] identified 86 HVCs possibly originating from flows from the AGN, but this origin could not explain the large population of HVCs found in other parts of the Halo. Currently, one of the most widely accepted origins for the majority of HVCs lies in accretion from extragalactic sources. Studies such as van Woerden and Wakker [2004] have shown the great influence that the Magellanic system has on the Halo. In addition, the low metallicities of HVC complexes support extragalactic accretion as a major source of HVCs in the Halo. Although the origins of HVCs are still debated, understanding how they are distributed in the Halo and what forms they take can provide more evidence to inform the ongoing debate.

1.3 Mapping the Halo

The first all-sky survey of HVCs was conducted in the northern hemisphere, and culminated in an extensive catalogue of clouds and their properties [Hulsbosch and Wakker, 1988]. By the late 1990s, the most complete catalogue of Northern HVCs was released by Wakker and van

Woerden [1991], which was complete for column densities of $N_{HI} > 3 \times 10^{18}$ cm^{-2}, and used data from Hulsbosch and Wakker [1988] and Bajaja et al. [1989]. The first all-sky HI survey that was conducted in the southern hemisphere with the velocity coverage to detect HVCs was the HI Parkes All Sky Survey (HIPASS) [Barnes et al., 2001]. From this survey, an HVC catalogue was presented by Putman et al. [2002] which used a search algorithm (explained further in de Heij et al. [2002]) to identify HVCs in the data set. Conclusions of this paper suggested that the catalogue was complete for column densities of $N_{HI} > 10^{19}$ cm^{-2}.

1.3.1 Discovering Diffuse Gas

Until recent years, a long-standing problem in this field was the "missing gas" within the Halo. While present day star formation rates are estimated at $\sim$ 1–3 M$_\odot$ yr^{-1} [Robitaille and Whitney, 2010], the maximum accretion rate that could be accounted for by known Halo flows was calculated between 0.1−0.5M$_\odot$ yr^{-1} [Putman et al., 2012] (although this was calculated with unrealistic assumptions about accretion efficiency). Although radio surveys down to $N_H > 10^{19}$ cm^{-2} were complete, a potential answer was found to lie in ultra-violet (UV) surveys such as Savage et al. [1993] and Bowen and Blades [1993]. Bekhti et al. [2010] details how lower density gas within the Halo can be probed through UV studies using chemical tracers such as CaII and NaI. These absorption lines are detected when observations are conducted towards quasars, and are measured as a column density. The conversion of these metal column densities to a hydrogen column density (as performed in Wakker and Mathis 2000) reveals diffuse gas present in the Halo at column densities considerably below 10^{19} cm^{-2}. The ubiquitous nature of this detected gas implies a potentially a significant fraction of mass that had been previously undetected by the less sensitive HI surveys; e.g. Fox et al. [2006] find high velocity gas present in $\sim$ 80% of sightlines across the sky. Follow up UV work by Collins et al. [2009] concluded that a mass infall rate of $\sim$ 1 M$_\odot$ yr^{-1} from this diffuse combination neutral and ionised hydrogen could be calculated.

Although UV observations have confirmed the existence of a diffuse gas population at low column densities, how it relates to known catalogued HVCs remains a mystery. For this reason, further surveys, such as GASS [McClure-Griffiths et al., 2009], have been conducted with higher sensitivities (compared to surveys such as HIPASS) to better characterise HVCs, and create a clearer map of the HI in the Halo. It is hoped that with a better understanding of the distribution of gas in the Halo will help provide evidence to complete the picture of

Galactic evolution.

1.4 Tracing Halo Clouds with Neutral Hydrogen

As mentioned, HVCs are composed of HI, which takes the form of a single proton and a single electron, each of which can exist in either a spin-up or spin-down state. When the system undergoes a spin-flip transition from the higher energy parallel state to the lower energy anti-parallel state, a photon is released at a frequency of 1420.405 MHz. This transition is extremely rare, with an average occurrence of approximately once every 10 million years. However the overwhelming amount of atomic hydrogen in the Galaxy means that this transition is easily detectable.

1.4.1 Radiative Transfer

The process which describes how electromagnetic radiation propagates through a medium is known as radiative transfer, and may be formulated as shown in Equation 1.1.

$$dI_\nu = \epsilon_\nu ds - \kappa_\nu I_\nu ds. \tag{1.1}$$

This describes the change in specific intensity (dI_ν), along a infinitesimal path ds as it passes through a medium. The emission coefficient (ϵ_ν) is defined as the power radiated per unit frequency per unit solid angle per unit volume for randomly oriented emitters. The absorption coefficient (κ_ν) is proportional to the true absorption minus the stimulated emission: the net absorption. (At radio wavelengths, the effects of scattering can be ignored.) By defining the optical depth as $d\tau_\nu = -\kappa_\nu I_\nu ds$ and the source function as $S_\nu = \epsilon_\nu/\kappa_\nu$, Equation 1.1 can be rearranged as:

$$\frac{dI_\nu}{d\tau_\nu} = I_\nu - S_\nu. \tag{1.2}$$

Although the formal solution to Equation 1.2 is given as:

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + \int_0^{\tau_\nu} S_\nu(\tau'_\nu)e^{-(\tau_\nu - \tau'_\nu)}d\tau'_\nu, \tag{1.3}$$

a simple solution exists in the case when the medium is approximated to be isothermal and homogeneous:

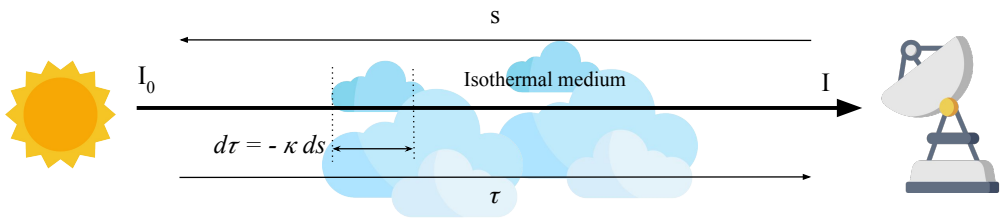


Figure 1.2: This is an illustration of the intensity of radiation through an isothermal medium where the change in intensity over a small interval $d\tau$ is given by Equation 1.2

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + S_\nu(1 - e^{-\tau_\nu}). \tag{1.4}$$

The total intensity of radiation $I_\nu(\tau_\nu)$ is therefore given by the sum of the radiation from the background source transmitted through the medium $I_\nu(0)e^{-\tau_\nu}$ and the radiation emitted by the medium $S_\nu(1 - e^{-\tau_\nu})$.

The Source Function

The source function, S_ν, is defined as the ratio between the absorption and emission coefficients, which may be expressed in terms of the Einstein coefficients. The Einstein coefficients describe the rates at which transitions occur within a system: spontaneous photon emission denoted by A_{ul}, photon absorption denoted by B_{lu}, and finally the stimulated photon emission denoted by B_{ul}. The Einstein coefficients are related as:

$$g_u B_{ul} = g_l B_{lu} \quad \text{and} \quad B_{ul} = \frac{c^2}{2h\nu^3} A_{ul}. \tag{1.5}$$

The coefficients of absorption and emission can be expressed with the Einstein coefficients as:

$$\kappa_\nu = \frac{h\nu_{ul}}{4\pi}(n_l B_{lu} - n_u B_{ul})\phi(\nu), \tag{1.6}$$

$$\epsilon_\nu = \frac{h\nu_{ul}}{4\pi} n_u A_{ul} \phi(\nu). \tag{1.7}$$

In both Equations 1.6 and 1.7, $\phi(\nu)$ is the line profile ($\int \phi_\nu d\nu = 1$) which defines the distribution of absorption and emission across frequency, which in the case of a single

transition will be determined by the velocity distribution of the HI in the cloud.

The populations of a pair of upper and lower levels, separated by energy ΔE, given the available energy $k_B T_{ex}$, may be described as:

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} e^{\Delta E / k_B T_{\rm ex}}, \tag{1.8}$$

This expression is based on the Boltzmann distribution, which relates temperature of a system to the number of particles in the upper or lower states. Here, n_u and n_l are the number densities of particles in those levels, and g_u and g_l are the level degeneracies. Generally, the excitation temperature, $T_{\rm ex}$ is a reparameterisation of the ratio of the level populations.

By re-writing the source function using Equations 1.5–1.8, we obtain the Planck function, where T is given by the excitation temperature between the levels of the atom:

$$S_\nu = \left(\frac{2h\nu^3}{c^2} \right) \frac{1}{e^{h\nu / k_B T_{ex}} - 1}. \tag{1.9}$$

Brightness Temperature

Given observed radiation of specific intensity I_ν, the brightness temperature is defined as the temperature of an equivalent black body emitter that would produce that radiation. The Planck function (*cf* Equation 1.9) is thus used to convert between an observed specific intensity and a "temperature".

At radio frequencies the Planck function can be simplified to the Rayleigh-Jeans limit. This occurs because in the radio regime, only the first term of the Taylor expansion of the exponential is used. Therefore it becomes:

$$B_\nu(T) \approx \frac{2\nu^2 k_B T}{c^2}, \tag{1.10}$$

and the definition of brightness temperature in the radio regime is given by:

$$T_b = \frac{I_\nu c^2}{2\nu^2 k_B} \tag{1.11}$$

This relates directly to the measured antenna temperature of the telescope, and for this reason, the brightness temperature may be thought of as a directly measurable quantity in radio astronomy.

By using the Rayleigh Jeans approximation and expressing the intensities I_ν and source function S_ν in Equation 1.4 in terms of temperature, we may therefore obtain:

$$T_b(\tau_\nu) = (T_{\rm ex} - T_c)(1 - e^{-\tau_\nu}), \tag{1.12}$$

where T_c is the brightness temperature of the continuum radiation incident on the back of the cloud, and T_b is the brightness temperature of the observed (continuum-subtracted) line radiation. In the observations used throughout this thesis, the emission line is detected against a weak background source (specifically the Cosmic Microwave Background), and the excitation temperature of the HI line (also known as the "spin temperature", $T_{\rm s}$) is large enough such that $T_c << T_{\rm s}$. We may therefore generally ignore the background radiation term and assume:

$$T_b(\tau_\nu) = T_{\rm s}(1 - e^{-\tau_\nu}), \tag{1.13}$$

Column Density

The column density is a measure of all atoms along the line of sight to the source within a cross-sectional area of 1 $\rm cm^2$, giving the units of column density $\rm cm^{-2}$.

Knowing the definition of optical depth and using Equation 1.6:

$$\tau_\nu = \frac{h\nu_{ul}}{4\pi}(n_l B_{lu} - n_u B_{ul})\phi(\nu) \int n_l ds. \tag{1.14}$$

The relevant substitutions can be made according to Equation 1.5 which gives:

$$\tau_\nu = \frac{c^2}{8\pi\nu^2}\frac{g_u}{g_l}A_{ul}(1 - e^{-\frac{h\nu}{k_B T_{\rm s}}})\phi(\nu) \int n_l ds. \tag{1.15}$$

The column density is the integration of n_l (number density) along the path s. Therefore $N_l = \int n_l ds$.

For the case of HI 1420 MHz transition, $h\nu << k_B T_{\rm s}$, meaning that we may retain only the first term in the Taylor expansion of the exponential. Using this fact and substituting all relevant constants then gives:

$$N_{\rm HI} = 1.8224 \times 10^{18}\, T_{\rm s} \int \tau_\nu \, dv, \tag{1.16}$$

where $N_{\rm HI}$ is given in units of cm^{-}2 and we have also converted from frequency space to velocity space. Finally, provided that τ_v is small, Equation 1.13 may be approximated as $T_b(\tau_v) \approx T_{\rm s}\ \tau_v$, giving

$$N_{\rm HI} = 1.8224 \times 10^{18} \int T_b\, dv. \tag{1.17}$$

When calculating the column density for a line profile using Gaussian fit parameters, Equation 1.17 can be rewritten as:

$$N_{\rm HI} = 1.8224 \times 10^{18}\, \sqrt{2\pi}\, T_b\, \sigma. \tag{1.18}$$

where T_b is the peak brightness temperature of the from the model and σ represents the square rooted variance which is a measure of the spread of the function.

1.5 Thesis Overview

Understanding the complicated structure of the Milky Way halo can unlock vital information required to answer long-standing questions about the evolutionary history of our Galaxy. This thesis presents an observational investigation into neutral hydrogen (HI) High Velocity Clouds (HVCs) which reside in the Galactic Halo, and includes comparisons with previous surveys and an exploration of the implications of survey sensitivity in determining the structure of the Halo. Data collected by the Survey of Weak Intensity Southern HI (SWISH) [Moss and McClure-Griffiths, 2017] and the Galactic All Sky Survey (GASS) [McClure-Griffiths et al., 2009] using the Parkes Radio Telescope were used in this study, as well as data from the Lockman Survey performed at the Green Bank Observatory [Lockman et al., 2002, hereafter L02]. The aims of this project were to:

- *Produce a semi-automatic pipeline to reduce SWISH data.* Building on calibration code previously used in similar surveys it was possible to construct a data reduction pipeline suitable for the position switching technique used in SWISH data collection. This pipeline relies on the traditional assumption that on-off spectra can be used to calibrate HI emission line data, the validity of which (when applied to SWISH data) is explored in this thesis.

- *Perform a comparative analysis of a subset of common sightlines using data collected from SWISH, GASS and L02.* This goal served to ensure that the data from SWISH was consistent with GASS and L02. A Gaussian fitting model was applied to both the SWISH and GASS data sets to determine cloud properties. The catalogue of cloud information from L02 was used to make a second comparison against SWISH results.

- *Determine if SWISH produces results in agreement with Moss et al. [2017], which compares the results from two HVC surveys.* By looking at any systematic differences along common sightlines as well as comparing Gaussian fits, it was possible to explore whether SWISH (with its higher sensitivity) produced any differences in the measured properties of a given cloud.

- *Discuss the results of SWISH and assess the possible implications of these results in the wider context of Galactic evolution.* The sensitivity of SWISH should result in better characterisation of diffuse gas compared to GASS, allowing for a better understanding of the neutral hydrogen distribution in the Halo. I considered the results of the analysis presented here in relation to the quest to better understand the diffuse HI population.

- *Determine the future directions of this research from the results of the analysis.* Based on the investigation of SWISH data I identify here the future directions for this work in terms of understanding the structure of the Galactic halo and limitations of the current calibration approach.

Through the analysis of the first 100 hours of SWISH data, this thesis represents the beginning of a greater study of the HI in the Halo of our Galaxy. By understanding the results of SWISH and the limitations of its data, this project will be used as a starting platform to inform future studies of diffuse HI in the Halo of our Galaxy. This thesis will primarily present initial results from SWISH, the latest survey conducted on the Southern HI Halo, and discuss these results within the context of the field. Chapter 2 details the motivations and survey techniques of SWISH. Chapter 3 presents the methodology behind the pipeline as well as a discussion of how the collection of data presented unique challenges for the reduction. Initial results from the data reduction are presented in Chapter 4, with an analysis of how these findings compare with the GASS and L02 surveys. The final chapter discusses future directions for this project as well as a summary of the outcomes of the thesis.

Science is actually, I think, a very creative venture. It requires thinking outside of the box. It requires an ability to be open to new experiences and an ability to change course in the middle, try a different path, or go about things in a new way.

Cara Santa Maria

2

Survey of Weak Intensity Southern HI (SWISH)

2.1 Previous Surveys of Interest

2.1.1 The Galactic All Sky Survey

The lack of sensitive HI surveys with high velocity resolution with wide velocity coverage saw the need for a survey such as the Parkes Galactic All Sky Survey (GASS). With the aim of building a more detailed map of the distribution of HI, and more specifically, HVCs in the Halo, GASS conducted a survey of neutral hydrogen for the southern sky below $\delta = +1°$. This is the most sensitive all-sky map of Southern Hemisphere HI to date, and covers all Milky Way velocities between LSR -400 km^{-1} and 500 km^{-1} [McClure-Griffiths et al., 2009]. GASS has an effective angular resolution of $\sim 16'$ and a velocity resolution of 1.0 $km\,s^{-1}$. The root mean square (rms) brightness temperature noise for GASS was 57 mK per channel.

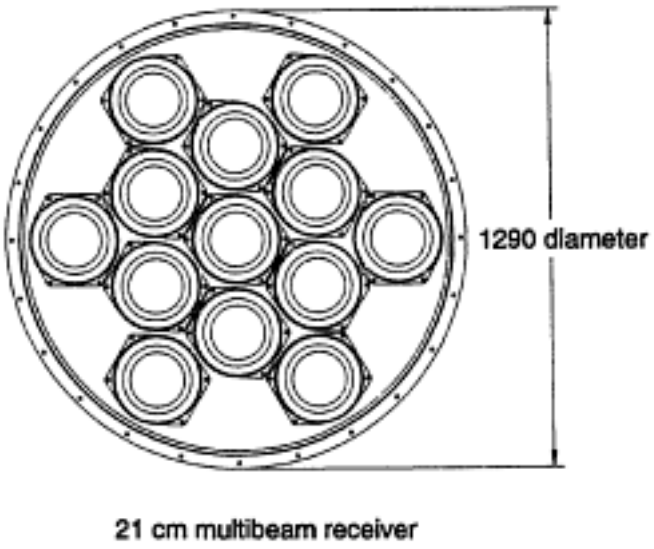


Figure 2.1: Diagrammatic view of the feed array from the 21cm Multibeam Receiver which was used on the Parkes telescope during the GASS observations [Staveley-Smith et al., 1996]

The GASS HVC catalogue had a 4σ detection limit which corresponded to a column density detection limit of $\sim 5 \times 10^{18}$ cm^{-2} [Moss et al., 2013]. While this is not low enough to probe in detail the more diffuse gas in the Halo, GASS data may provide a valuable scale against which to calibrate deeper surveys with lower detection limits.

GASS observations were conducted on the Parkes 64 m Radio Telescope with use of the 21 cm multibeam receiver. This instrument has 13 beams stacked in a hexagonal pattern, as shown in Figure 2.1 [Staveley-Smith et al., 1996]. Data was collected by frequency switching, which involved recording spectra at two closely spaced frequencies centred at 1418.8435 and 1421.9685 MHz and using the offset frequency observations to characterise the bandpass and correct for instrumental effects. The first release of the GASS data set was completed in 2009 and can be seen visualised in Figure 2.2 [McClure-Griffiths et al., 2009]. Soon after, an HVC catalogue was created which used an automated source-finding technique to find HVCs with peak brightness temperatures of 230 mK and above [Moss et al., 2013]. This catalogue will be referred to in subsequent chapters as the "GASS catalogue".

2.1.2 The Lockman Survey

Unlike GASS, the Lockman Survey (L02) consisted of targeted, sensitive observations of weak HVCs at $\delta \geq$ -43° performed on the 43 m Telescope of the NRAO in Green Bank

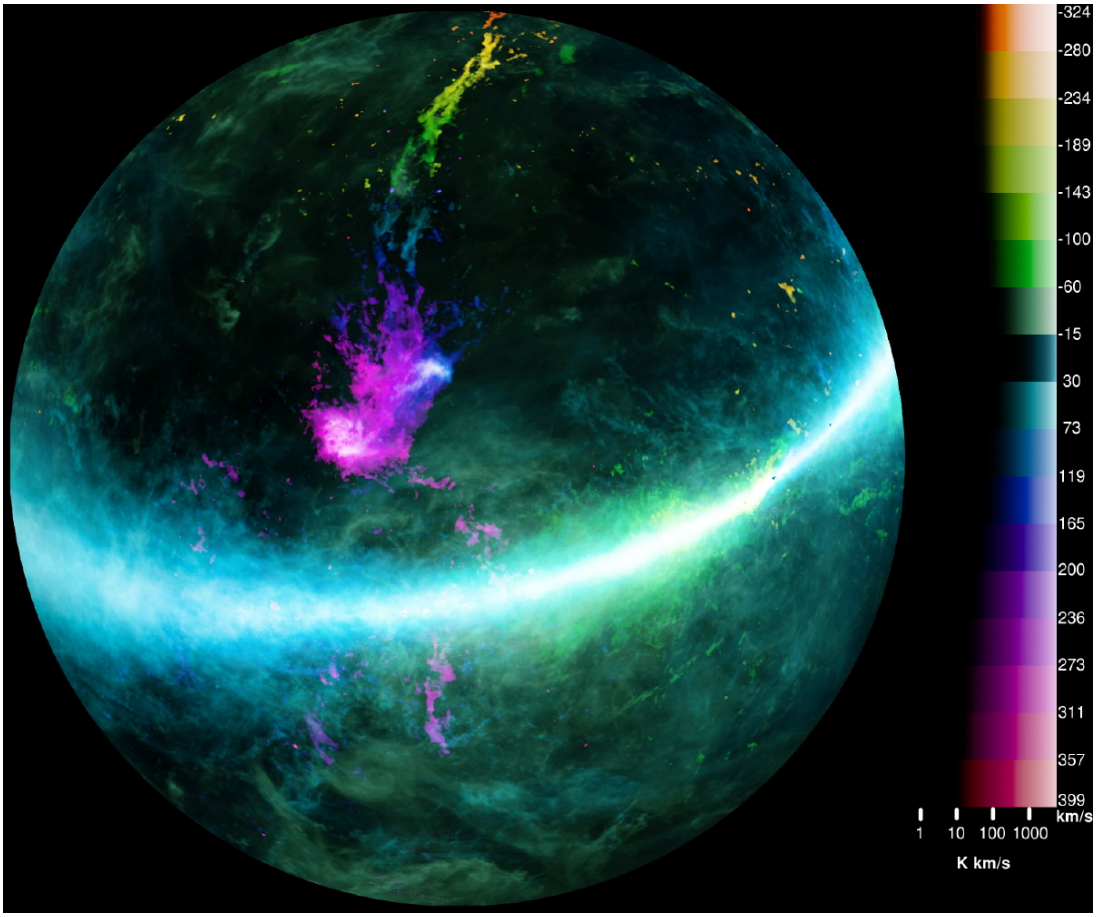


Figure 2.2: The GASS I data release visualised with a ZEA projection where the centre of the image is located on the South Celestial Pole. 0 hr right ascension (RA) is at the top of the figure with RA increasing counter-clockwise. The colour bar represents velocity integrations of ~ 40 km s^{-1} with intensity of the brightness temperature given in a logarithmic scale for each velocity chunk. Credit: S. Janowiecki

Observatory [Lockman et al., 2002]. The angular resolution for this telescope at 1420 MHz was $21'$ and the observations had a velocity resolution of ~ 5 km s^{-1}. The median rms per-channel brightness temperature for this survey is 3.4 mK which results in a column density detection limit around $\sim 8 \times 10^{17}$ cm^{-2}, which is significantly better than GASS. For this reason, L02 saw the detection of much weaker HVCs which were not detected in GASS due to the lower sensitivity of the observations. Consequently, L02 reveals that there is likely a missing population of faint high velocity gas in the GASS catalogue due to the sensitivity limitations of an all-sky survey.

2.2 Scientific Motivations of SWISH

The Survey of Weak Intensity Southern HI was designed to investigate results found in Moss et al. [2017]. In Moss et al. [2017], the two HVC populations observed in GASS and L02 were run through a machine learning algorithm to classify each HVC as "dense" or "diffuse". The learning set consisted of the cores of HVCs detected in GASS (dense) and clouds observed in L02 (diffuse). From this set, the algorithm classified HVCs from both surveys and determined some fundamental properties which separated the diffuse and dense components. Within the GASS data, the dense components outnumbered the diffuse components with a ratio of 9:1. Interestingly, in L02, the dense components were outnumbered 1:3. Using these ratios, Moss et al. [2017] concluded that diffuse gas would be 2–3 times the mass of the dense population. The motivation behind SWISH was to conduct a deep HI survey similar to L02 of the southern sky, which aimed to reproduce the results seen when comparing GASS/L02 in the North. Note that much of the content of this chapter is based on the original project proposal for SWISH (Project number P953 Moss and McClure-Griffiths [2017]).

2.3 Survey Details

The first data collection for SWISH took place from April to September 2017. A total of 115 hours were allocated during this semester which were used to investigate both the optimal observing mode, as well as carry out the observations. Two possible observing methods were available to this project which included the DFB4 (referring to the standard digital filterbank) mode or the newer, HIPSR backend. For the DFB4 mode, only the central beam of the multibeam receiver could be used which meant the more common method of frequency switching was available. However, the HIPSR backend was chosen which allowed for all 13 beams to simultaneously record data. The motivation to use HIPSR rested with the ability to maximise the on-source time, as well with the alternative option presenting systematic issues during the testing phase. Since frequency-switching was not possible with HIPSR, finding a way to calibrate and reduce the data presented complications which this project investigates.

2.3.1 Position Switching

During the SWISH observations, the multibeam receiver was directed towards a target with the central beam of the receiver initially centred on the source position. The remaining 12 beams recorded information from positions surrounding the source, with the assumption that these beams would be located "off source". This resulted in an efficiency of 39 minutes per hour on-source, or 72% on-source time. The assumption of the off-source beams containing no emission is discussed in the next chapter when reducing the data. For each sightline, all 13 beams in the receiver were placed in turn on the source position, which resulted in 13 on-source spectra and 12×13 off-source spectra for each source. This had the additional benefit of reducing the noise in the final spectrum for an observation.

2.3.2 Sightlines

Since SWISH was created to investigate conclusions made by Moss et al. [2017] which targeted source in both GASS and L02, a large portion of the observing time was dedicated to re-observing a combination of these sightlines. A total of 129 sightlines were observed; 27 calibration sources, 34 L02 sightlines, 15 GASS diffuse clouds and 53 blind sightlines. The calibration sources were chosen to ensure that both flux-scaling and HI objects were observed to provide the most flexible opportunities when calibration of the data was performed at a later time. Initially, SWISH observed sightlines were constrained to altitudes of $60^\circ < \theta < 85^\circ$. The former constraint was to maximise altitude (Parkes can only observe > 30° at best) and the latter is due to difficulties for the large dish tracking near zenith. The 60° constraint was later relaxed to 40° to allow more observability of sources. The L02 re-observations were biased towards fainter clouds to test the sensitivity of SWISH and consistency of measurements. The re-observations of GASS sightlines, on the other hand, were not targeting the centre of catalogued HVCs but their surrounding diffuse regions (as found using the machine learning classification in Moss et al. [2017]) to better understand the structure an HVC takes. The SWISH "blind" targets were chosen based on GASS data (masked where there are detections of statistically significant HI column density) indicating a "blank" region, in order to provide quantitative insight into the ability of GASS to detect the diffuse HI HVCs. All observations are shown in Figure 2.3.

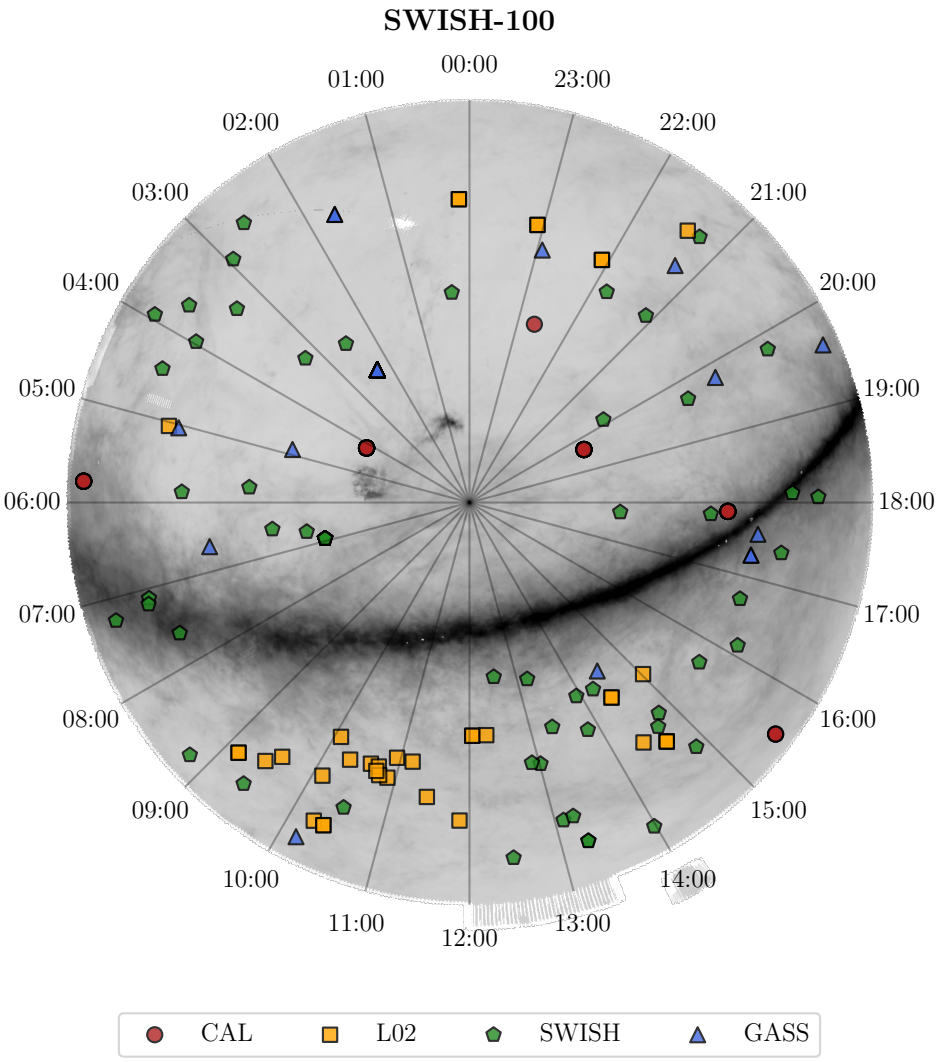


Figure 2.3: Target positions observed during the first semester of SWISH over-plotted on an integrated intensity map of HI in the Southern hemisphere from GASS data. Each different target category is shown in the legend. RA is indicated around the image to identify the orientation of the image.

I was taught that the way of progress was neither swift nor easy.

Dr Marie Curie

3

Data Reduction and Analysis

3.1 Data Reduction

Data collected through SWISH was reduced using a semi automatic pipeline, partially developed in this project. The code that was initially provided at the start of the project used existing gain calibration code to calculate system temperature from a calibration source observation [Price, 2012]. It also performed a simple bandpass calibration from the combined on- and off-source data for each beam in the multibeam receiver. However, beyond that its functionality was limited.

An observational dataset contained spectra in each of two orthogonal linear polarisations for all 13 beams when they were on and off source. The sum of these two polarisations gives the total intensity. The dataset for each beam thus contained 24 off-source spectra and two on-source spectra which are combined to produce the raw bandpass-calibrated spectrum. The bandpass calibration is performed separately for each polarisation, as shown in Equation 3.1, where $P_{\mathrm{on},X}$ represents the spectrum in polarisation X when a beam is on source and

$P_{\mathrm{off},X}$ when the beam is off source. $T_{\mathrm{sys},X}$ is the system temperature which is calculated as mentioned earlier.

$$P_{\mathrm{on/off},X} = \frac{P_{\mathrm{on},X} - P_{\mathrm{off},X}}{P_{\mathrm{off},X}} T_{\mathrm{sys},X}. \tag{3.1}$$

An important consideration for the bandpass calibration calculation is to ensure that the off-source spectrum is free from emission. "Contaminated" off positions contained HVC emission due to non homogeneous cloud distribution in the sky. Throughout this chapter, these will be referred to as "bad" data or positions. During the observations, the assumption was that only one beam fell on source and the other 12 were on "blank" sky. However, if a beam's off position collected any emission from the target source, this would reduce the measured peak brightness temperature. Since this survey aims to conduct a sensitive survey of weak HI, it is important that these contaminated positions are removed to ensure the final bandpass-calibrated data is not compromised. In addition to contaminated off positions, system issues and radio frequency interference (RFI) contributed to a small amount of data being impossible to analyse.

3.1.1 Determining Bad Off Positions

A preliminary visualisation was conducted on known HVCs from the GASS catalogue to better understand the environments around source locations. By superimposing the location of the beams during the SWISH observation onto existing GASS III data, I was able to develop an understanding of how the multibeam receiver collected data over the cloud structure, and how the extended structure of an HVC could introduce contaminated emission. Two representative plots can be found in Figures 3.1 and 3.2. Figure 3.1 shows a GASS catalogued HVC at 119.55 km/s. This particular sightline is a clear example of an HVC which extends over "off" beams. Using the position switching technique for this cloud would be difficult and would require a well defined list of contaminated positions for removal. Figure 3.2 on the other hand shows a relatively compact source which does not spill into the off beams. For this source, all off-source positions could likely be used when performing the bandpass-calibration.

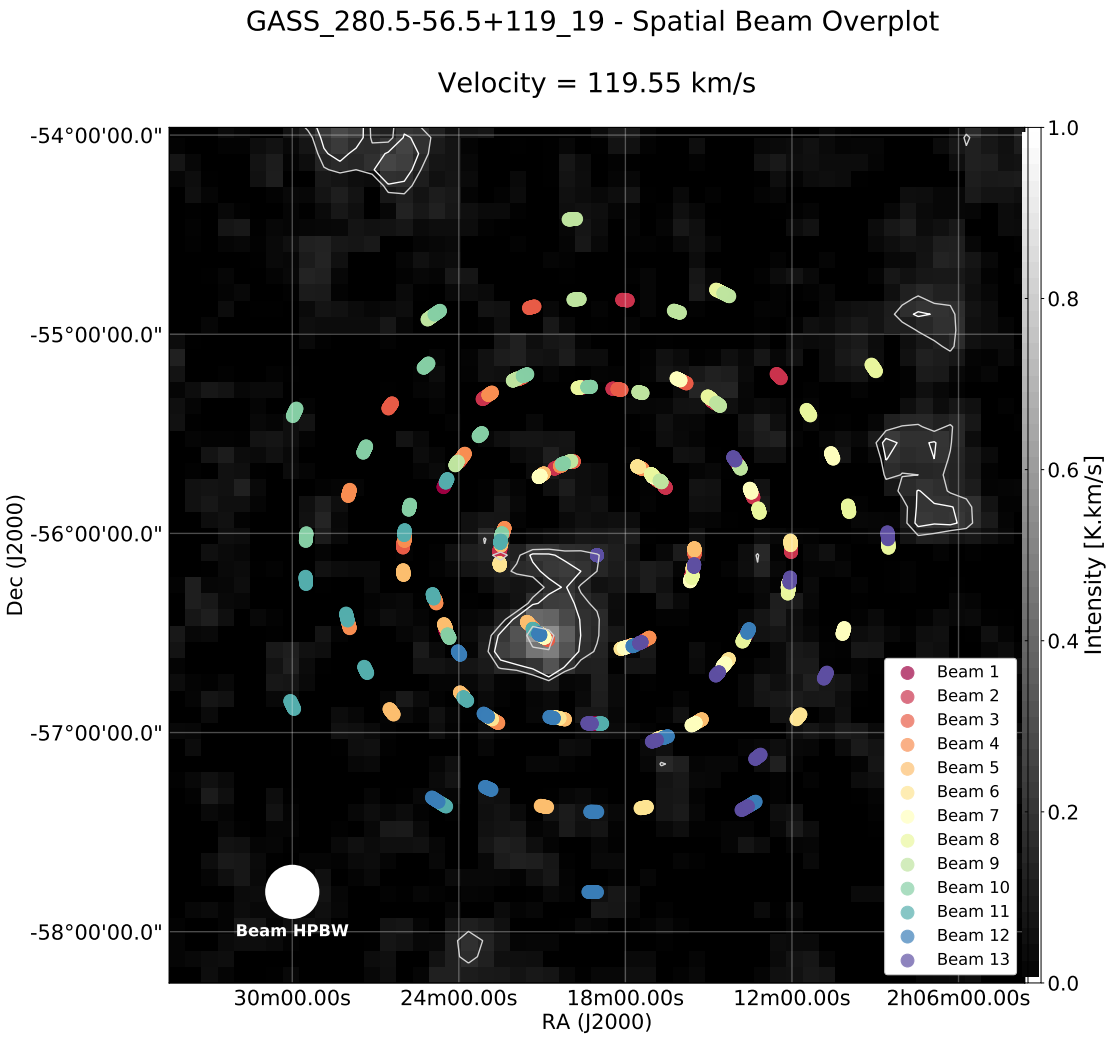


Figure 3.1: The environment around a catalogued HVC – GASS-280.5−56.5+119_19 from GASS data. This image shows a single spectral channel with greyscale indicating the velocity-integrated intensity within this channel. The contours mark integrated intensity levels of 0.15, 0.2, 0.5 and 1.0 K km s^{-1}. SWISH observations of this sightline are shown by circles, where each beam on the multibeam receiver is coloured coded according to the legend. The central circle of this pattern indicates the target and "on" source location (Since beam 13 was the last to be on source, the final colour on this plot is the purple 13th beam). A representation of a beam's half power beam width (HPBW) is given in the lower left corner. This particular example shows a case where the HVC is extended over a large portion of the total observing region, meaning that "off" beams collect HI emission, resulting in highly contaminated off-source spectra.

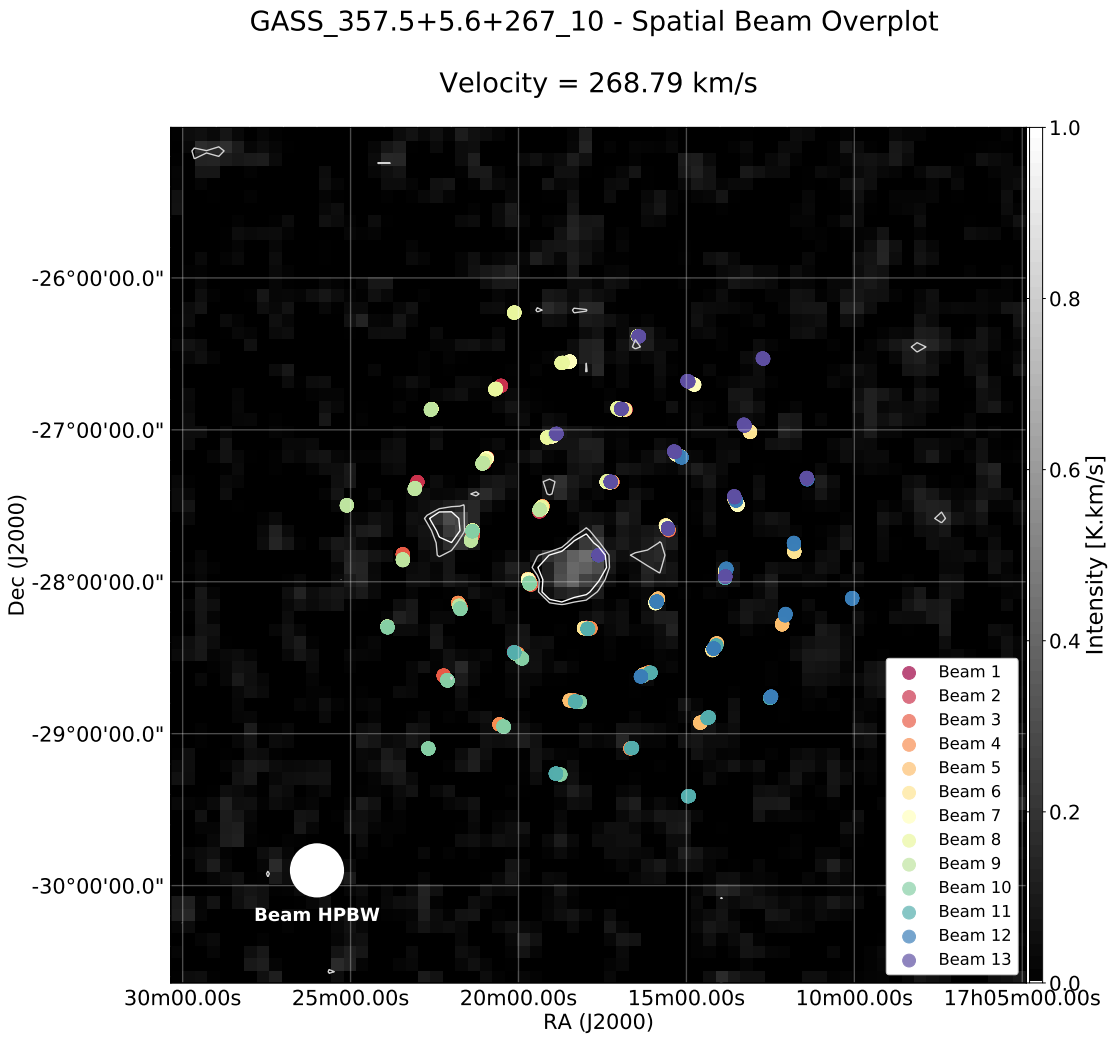


Figure 3.2: The environment around a catalogued HVC – GASS-357.5+5.6+267_10 from GASS data. This image was generated using the same parameters as Figure 3.1. In this example, the HVC is small and centred on the "on" beam location. There are small regions which have HI emission but will only effect a few off beams. This sightline shows an observation that will contain a very small number of contaminated off positions.

The first approach taken was to develop an automatic analysis of on/off spectra generated separately for each off-source position and each beam. For each beam, 12 bandpass-calibrated spectra, generated from each combination of the on-source measurement and the 12 off-source measurements, were plotted. By masking the Galactic velocity range between -90km/s and +90km/s, theoretically, the remaining spectra could be scanned to find any peaks indicating an HVC detection. The spectrum with the highest peak brightness temperature was recorded with the assumption that this detection had the least contamination in its off beam position.

Using this same logic, any other spectra taken by this beam that fell below 30% of this value were assumed to have significant contamination present in their off positions, and so these off positions were discarded. The value of 30% was chosen to maximise inclusion of spectra to reduce the overall noise of the final spectrum, but to discard spectra which would reduce the detection. It is important to note that this value was chosen on a qualitative basis to ensure a significant portion of the data could be included to reduce the noise value of the final spectrum. It was understood that this could in some cases result in a underestimation of the final cloud properties, to a maximum level of 30% (and in most cases much less).

In an ideal sightline where the baseline level of all beams remained close to zero, this automatic technique was able to identify and remove any contaminated off positions. In practice, this survey is locating clouds with a weak brightness temperature and the residual baseline structure often dominates spectra, therefore detecting a cloud using this simple automatic method was challenging in most sightlines. Since this method involved identifying the peak of an HVC before a baseline correction was performed, it was concluded that manual observation of each bandpass-calibrated spectrum was needed as a by-eye detection could be made even if the residual baseline dominated in an observation. The reason a baseline correction could not be performed before the identification of the peak was due to the possibility of a weak detection being fitted during the correction process and eliminated from the spectrum. Since there was a strong possibility of weak detections (due to SWISH being such a sensitive survey) it was concluded baseline correction should only occur after an HVC had been identified and the relevant velocity range excluded from the baseline fit. During this project a by-eye method was feasible due to the number of sightlines being relatively limited.

As in the automatic method, a rudimentary plot was shown of a sightline with the bandpass-calibrated spectra for all beams averaged together. This was used to make an initial determination if an HVC was detected in this sightline. An example of this plot can be found in Figure 3.3 and shows a detection of an HVC. If an HVC was identified, the classification of any bad spectra in each beam was performed using the criteria bellow. It is important to note that some sightlines had more than one HVC detection which were separated and and run through the reduction independently.

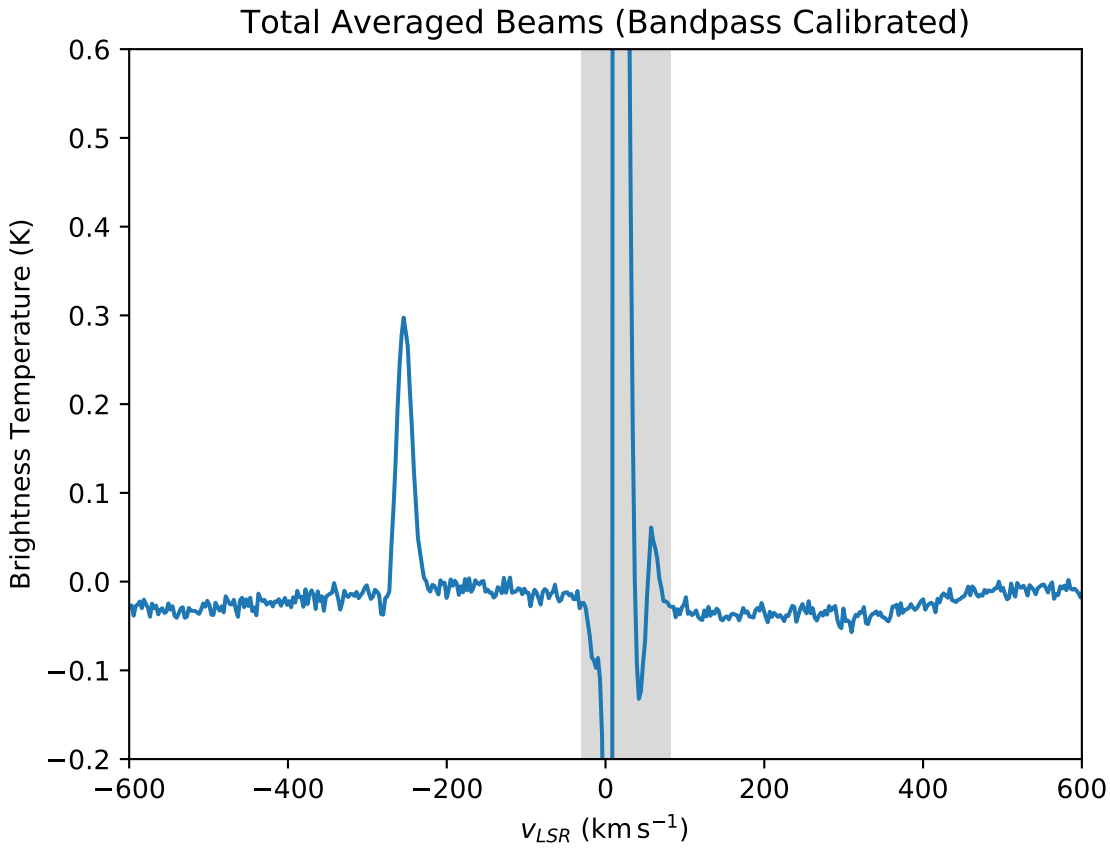


Figure 3.3: A plot showing the initial bandpass calibration of a known catalogued HVC – GASS-357.5+5.6+267_10 from the SWISH observation run. Here all beams have been bandpass calibrated and averaged together showing the HVC detection at approximately 251 km s^{-1}. At this stage in the data reduction, no bad data has been removed. The grey shaded region indicates the velocity range of the Galactic Disk and is not considered when seeking HVCs.

Characteristics of a Bad Bandpass-Calibrated Spectrum

An important part of this method was to use a consistent classification for determining a bad off position spectrum for a beam. Since all combinations of on-off spectra for a given beam were analysed by eye, the following criteria were used as a set of guidelines when removing bad bandpass-calibrated spectra.

- The feature produced by using that off position in a bandpass-calibration was presenting as an absorption feature. This would indicate that the off spectrum had a higher brightness temperature than the on spectrum, resulting in a negative brightness temperature when the two were combined. This would be very common in an HVC with

extended structure over the entire 13 beam projection.

- If a feature from an off position combination had absorption wings around the HVC of interest. This would occur if the HVC was extended as well as having a range of velocities within the cloud. Although this characteristic was not as common, it was important in ensuring that any HVC with multiple components would not have one of those components lost in the final averaged mean of the beams.

- The last and more subjective criterion was to determine the highest peak brightness temperature of the HVC and compare this with the other bandpass calibrated spectra from the same beam. Since all bandpass-calibrated spectra had varying baselines (the reason a manual by-eye method was used), a qualitative assessment of the off beams was done to determine which combinations produced the feature with the highest brightness temperature. As discussed earlier, this process was done knowing that a reduction in the feature's intensity would occur if spectra with lower peak brightness temperatures were included, but would also reduce the noise in the final spectrum. Depending on the intensity of the HVC, a certain level of flexibility and judgement was used. This often manifested in removing various different beams in different combinations to produce the highest detection.

Since a manual classification of bad bandpass-calibrated spectra was used, it was important that the limitations of this method were investigated and understood. During the data reduction development, sightlines were often reduced in slightly different ways which produced a range of results. Since there was a subjective element to the classification, it is expected that there will be small variations if this method is carried out by different people. However it is important to note that the results will (if anything) tend to underestimate the brightness temperature of the feature rather than overestimate it.

3.1.2 Applying a Baseline Correction

Once all bad off positions were removed, a mean average of all remaining bandpass-calibrated spectra for a given beam was plotted, as shown in Figure 3.4. This final spectrum is left with residual baseline structure, including a non-zero offset. The user inputs the velocity range of the Galactic Disk emission, as well as any HVC detections on the spectrum (in some cases, a sightline had more than one detection). With the main HVC feature, Galactic emission and

any other HVCs defined, a mask was applied to these regions. Using the polyfit function in the NUMPY package, a least squares fitting method was used to fit each beam's residual bandpass ripple. The best polynomial fit was then be subtracted from the original spectrum to produce the baseline corrected spectrum for a given beam. All beams were then mean averaged to produce the final spectrum for the sightline, which was used in the analysis as discussed in the following chapter.

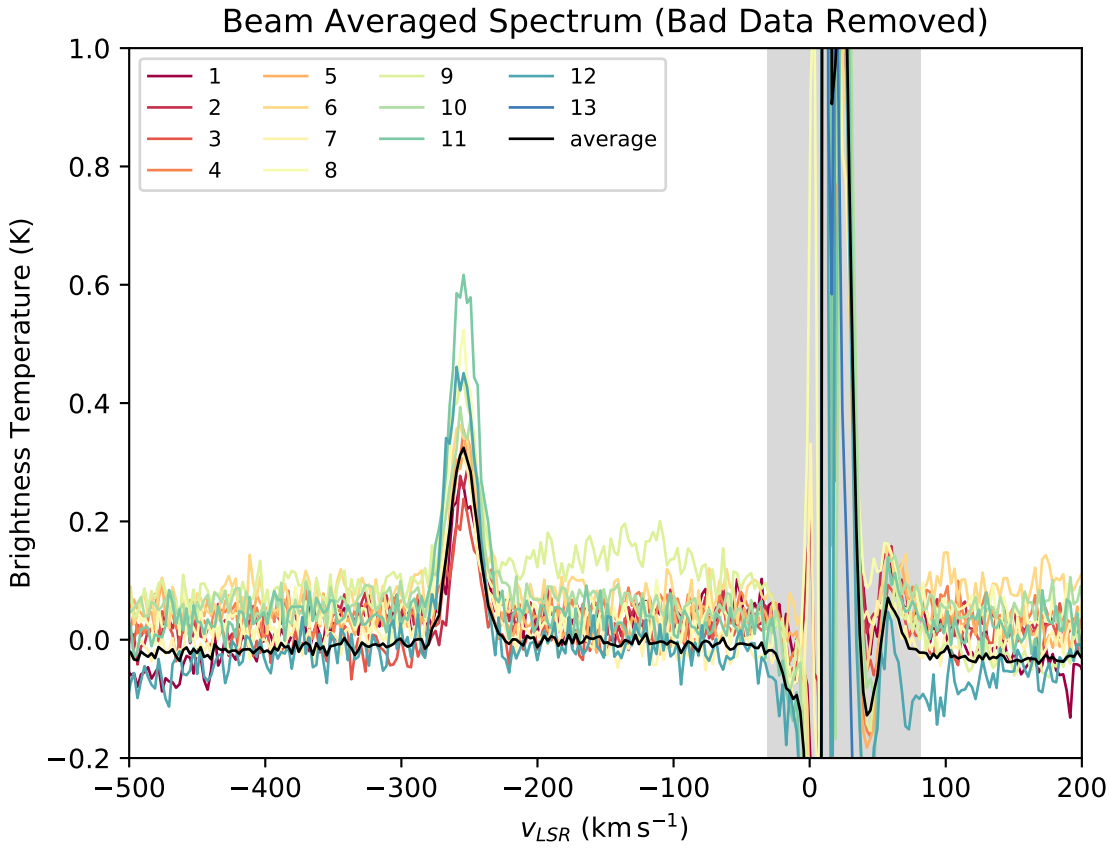


Figure 3.4: Averaged bandpass-calibrated spectra for all individual beams to highlight the need for baseline correction for individual beams. The black "averaged" line shows the average between all positions which results in a small offset from the baseline, which will need to be corrected for. Beam 9 shows an extreme example of the baseline correction needed in the negative velocities. The HVC shown is a the same sightline as above: HVC – GASS-357.5+5.6+267_10). The grey shaded region indicates the velocity range of the Galactic Disk

3.1.3 Gaussian Fitting

Once the final spectrum was generated, the physical properties of the resulting feature could be investigated by "fitting" Gaussian models to the data [Moss et al., 2017]. Several approaches are commonly used to produce best-fit models, and several standard least-squares fitters such as MPFIT were tested. However in this project, model selection was carried out with the EMCEE package [Foreman-Mackey et al., 2013] in Python, which uses the Markov-Chain Monte Carlo (MCMC) technique. Unlike traditional fitting software, MCMC samples the posterior probability distribution which then infers the positions of the best parameters. This method has been used in similar Gaussian decomposition projects in similar fields of radio astronomy [Hogg and Foreman-Mackey, 2018, Liu and Shaw, 2019, Petzler, 2018]. Both a single and double Gaussian model were generated for each spectrum to determine if a second cloud component was present.

The best model was determined by using the Bayesian Information Criterion (BIC), which is given in Equation 3.2. In this equation, χ^2 is a statistical measure of how good a fit is to the data and the second term will penalise a model for increasing the complexity of said model.

$$BIC = \chi^2 + d \ln\left(\frac{N}{2\pi}\right) \tag{3.2}$$

If the BIC value of the double Gaussian model is at least 10 less than the single Gaussian model, the double Gaussian fit is strongly preferred Kass and Raftery [1995]. By using the BIC in model comparison, human judgement was eliminated when judging which model would best represent a cloud. Using the BIC to determine model selection is common in astronomy where model selection is important in determine physical structures of the model being used [Owers et al., 2019].

Errors

The Markov Chain Monte Carlo technique employs a group of ‘walkers’ that sample the posterior probability distribution in a biased random walk through the available parameter space. Each step the walkers take is random, but the step is only accepted if the new local value of the posterior probability function satisfies an acceptance algorithm. Once the walkers converge, they will tend to spend proportionally more time in areas with higher posterior probability density. If the posterior probability density over parameter space is

Gaussian in nature, the distribution of the walkers will also be Gaussian and therefore the median position of the walkers corresponds to the peak of the probability density function. The 16^{th} and 84^{th} quantiles of the walkers' positions in each dimension of parameter space correspond to the standard deviation of that parameter. This method of calculating the standard deviation for the best model is detailed in Hogg and Foreman-Mackey [2018].

3.1.4 GASS Data

In order to compare the results of the two surveys and ensure no systematic differences were present, a subsection of sightlines which had been successfully run through the data reduction pipeline were compared against data from the GASS III release [McClure-Griffiths et al., 2009, Kalberla et al., 2010, Kalberla and Haud, 2015]. This was done by extracting the spectrum from the closest pixel to the position of the SWISH on source location, and then applying the same Gaussian fitting method using the same velocity range as the SWISH HVC. Since GASS had an average characteristic per-channel noise of 45 mK, compared to SWISH at which ranged between 2 mK and 6 mK, it was expected that there may be cases where a Gaussian may not be the best fit and a null-model preferred. In the case that a Gaussian was fitted to GASS data and the null-model was preferred, the interpretation was that there was no detection by GASS in the sightline. The results of this Gaussian fitting will be discussed in the next chapter.

Science is about the process; it's not about the conclusion.

Dr Steven Novella

4

Results and Discussion

4.1 HVC Detections

Of the 72 unique sightlines analysed from the first set of observations taken by SWISH, 35 had no detection (null model preferred), and 14 had detections which occurred primarily in the off beams and thus were excluded from the present analysis. This left 23 sightlines with clean detections, 3 of which contained more than one HVC feature in their spectra. Of the 26 total HVC detections, 24 were best fit by a single Gaussian, and 2 were best fit by double Gaussian functions.

Of the 72 sightlines, 28 were re-observations of known sources and 44 were new observations with no previous detection recorded. The re-observations had a 89% detection rate (25/28). Of these, all previously catalogued GASS clouds were re-detected, whereas three weaker L02 clouds (all with peak $T_b < 12$ mK; Lockman et al. 2002) were missed by SWISH. The new detection rate was 25% (11/44). Note that here we have included both clean and contaminated detections in the statistics, but that only clean detections were carried forward

for further analysis.

4.1.1 Single Component Clouds

Table 4.1 shows the results for the single Gaussian fits from both the SWISH and GASS data. In these results, the preferred model according to the BIC was a single Gaussian function, for which the model parameters are recorded. Of the clouds detected in SWISH data, 3 had no detection in GASS (null model preferred), and 6 preferred a single Gaussian model but with high errors, suggesting the confidence for these values is very low.

4.1.2 Double Component Clouds

Two HVCs which were re observations of GASS-catalogued clouds (GASS-357.5−5.6+267_10 and GASS-37.2−10.3+256_0) were found to have double Gaussian components which were not found in the corresponding GASS data. Both clouds are shown in Figure 4.1 and 4.2 and show the spectra with the preferred model over-plotted. The double Gaussian model indicates the presence of a narrow (possibly dense) and a broad (possibly diffuse) component for the HVC which was not detected in the GASS data, due to the higher noise level masking the diffuse component. The parameters for these clouds can be found in Table 4.2.

4.2 Systematic Differences in SWISH and GASS Data

4.2.1 Brightness Temperature

The first thing that was noticed in these results was an apparent increase in brightness temperature in the SWISH data compared to GASS. A comparison plot of peak brightness temperature of the SWISH and GASS single-component fits is shown in Figure 4.3. Using an orthogonal distance regression (ODR) fitting package that forms part of the SCIPY package [Jones et al., 2001], a linear fit was made to the data, resulting in a best fit line of $y = (1.17 \pm 0.073)\,x + (0.019 \pm 0.023)$. In this equation, x and y represent the brightness temperature of the GASS and SWISH fits, respectively. ODR fitting ensured that no preference was given to the orientation of the axes, and the errors for each data set were included to down-weight points with significant uncertainties.

Source Name	SWISH Data				GASS Data			
	v_{LSR} (km s^{-1})	FWHM (km s^{-1})	$T_{b,fit}$ (K)	N_{HI} ($\times 10^{18}$ cm^{-2})	v_{LSR} (km s^{-1})	FWHM (km s^{-1})	$T_{b,fit}$ (K)	N_{HI} ($\times 10^{19}$ cm^{-2})
GASS_10.3-25.0-271_41	-254.39 ± 0.072	23.46 ± 0.17	0.324 ± 0.002	1.47 ± 0.04	-258.13 ± 0.65	24.46 ± 1.39	0.26 ± 0.01	1.23 ± 0.24
GASS_180.6-77.4-116_12	-90.58 ± 0.234	23.61 ± 0.23	0.33 ± 0.006	1.53 ± 0.08	-93.95 ± 0.66	25.87 ± 1.8	0.23 ± 0.01	1.14 ± 0.26
GASS_224.1-34.4+133_44	128.4 ± 0.056	29.52 ± 0.132	0.22 ± 0.001	1.26 ± 0.02	213.9 ± 1.17	36.65 ± 2.69	0.18 ± 0.01	1.28 ± 0.33
GASS_246.2+39.9+312_0	311.31 ± 0.099	36.93 ± 0.25	0.22 ± 0.001	1.55 ± 0.04	309.11 ± 0.58	41.02 ± 1.38	0.28 ± 0.01	2.20 ± 0.31
GASS_28.2-41.1-221_19	-204.68 ± 0.075	24.96 ± 0.18	0.285 ± 0.002	1.38 ± 0.04	-207.19 ± 0.79	23.36 ± 1.92	0.2 ± 0.01	0.91 ± 0.24
GASS_280.5-56.5+119_19	108.56 ± 0.07	25.77 ± 0.16	0.27 ± 0.001	1.36 ± 0.03	106.35 ± 0.55	22.13 ± 1.22	0.21 ± 0.01	0.90 ± 0.19
GASS_280.5-56.5+119_20	196.22 ± 0.06	24.33 ± 0.13	0.24 ± 0.001	1.15 ± 0.02	193.71 ± 0.75	24.98 ± 1.74	0.18 ± 0.01	0.89 ± 0.22
M16.11+	-151.09 ± 0.11	18.11 ± 0.28	0.28 ± 0.004	0.98 ± 0.06	-153.85 ± 0.82	15.33 ± 2.18	0.2 ± 0.02	0.60 ± 0.29
PKS_1009-321-	175.35 ± 0.09	18.63 ± 0.2	0.23 ± 0.002	5.03 ± 0.08	173.14 ± 1.28	24.46 ± 2.65	0.15 ± 0.02	5.35 ± 0.34
PKS_1009-321-	214.71 ± 0.03	25.94 ± 0.08	0.83 ± 0.002	4.18 ± 0.05	212.71 ± 0.2	26.77 ± 0.6	0.89 ± 0.01	4.62 ± 0.31
PKS_1011-282+	125.14 ± 0.01	22.92 ± 0.02	1.32 ± 0.001	5.85 ± 0.02	123.17 ± 0.13	23.3 ± 0.29	1.02 ± 0.01	4.59 ± 0.21
PKS_1020-103–	86.42 ± 0.04	19.33 ± 0.11	0.23 ± 0.001	0.86 ± 0.02	82.97 ± 0.25	23.1 ± 0.6	0.16 ± 0.003	0.70 ± 0.06
PKS_1034-29+	117.76 ± 0.1	24.1 ± 0.24	0.18 ± 0.002	1.40 ± 0.06	*117.86 ± 2.68*	*19.9 ± 36.84*	*0.13 ± 0.03*	*0.50 ± 2.09*
PKS_1034-29+	184.33 ± 0.18	27.07 ± 0.42	0.1 ± 0.001	0.53 ± 0.03	*184.1 ± 18.35*	*11.31 ± 9.9*	*0.08 ± 0.1*	*0.18 ± 0.75*
PKS_1101-325+	94.33 ± 0.07	25.38 ± 0.17	0.33 ± 0.002	1.63 ± 0.04	95.24 ± 0.98	45.34 ± 2.73	0.23 ± 0.01	2.02 ± 0.42
PKS_1117-248+	149.8 ± 0.313	20.89 ± 0.81	0.074 ± 0.002	0.30 ± 0.04	0 ± 0	0 ± 0	0 ± 0	0.00 ± 0.00
PKS_1207-399-	121.08 ± 0.12	24.32 ± 0.29	0.14 ± 0.001	0.66 ± 0.03	115.05 ± 1.18	24.42 ± 2.81	0.14 ± 0.01	0.66 ± 0.25
S9	-124.3 ± 0.04	28.11 ± 0.08	0.4 ± 0.001	2.16 ± 0.02	-133.82 ± 0.88	35.26 ± 1.91	0.32 ± 0.02	2.17 ± 0.51
SWISH_054411-430611	165.85 ± 0.18	25.17 ± 0.53	0.12 ± 0.07	0.59 ± 0.71	*162.96 ± 1.47*	*17.25 ± 3.9*	*0.08 ± 0.01*	*0.25 ± 0.19*
SWISH_065600-585145	132.97 ± 0.63	19.54 ± 5.67	0.035 ± 0.002	0.13 ± 0.09	0 ± 0	0 ± 0	0 ± 0	0.00 ± 0.00
SWISH_092454-085156	118.64 ± 0.95	14.27 ± 2.13	0.06 ± 0.007	0.15 ± 0.09	0 ± 0	0 ± 0	0 ± 0	0.00 ± 0.00
SWISH_154044-292316	-144.15 ± 1.57	23.42 ± 3.98	0.08 ± 0.01	0.38 ± 0.21	*-137.35 ± 6.92*	*33.48 ± 8.87*	*0.02 ± 0.007*	*0.15 ± 0.16*
SWISH_194846-153607	-243.8 ± 0.62	23.49 ± 1.3	0.1 ± 0.005	0.45 ± 0.10	*-246.68 ± 37.6*	*17.1 ± 4.63*	*0.07 ± 0.05*	*0.23 ± 0.46*
SWISH_211621-112259	-240.14 ± 0.04	21.62 ± 0.1	0.16 ± 0.0006	0.68 ± 0.01	-240.98 ± 0.97	25.85 ± 2.55	0.16 ± 0.01	0.81 ± 0.26

Table 4.1: Parameters of single Gaussian components for 23 sightlines in the SWISH data

Source Name	SWISH Data							GASS Data			
	v_{LSR} (km s^{-1})		FWHM (km s^{-1})		$T_{b,fit}$ (K)		N_{HI} ($\times 10^{19}$ cm^{-2})	v_{LSR} (km s^{-1})	FWHM (km s^{-1})	$T_{b,fit}$ (K)	N_{HI} ($\times 10^{19}$ cm^{-2})
GASS_357.5+5.6+267_10	268.58 ± 0.44	297.78 ± 2.41	38.22 ± 0.93	61.2 ± 2.03	0.15 ± 0.007	0.07 ± 0.005	1.97 ± 0.34	270.67 ± 1.56	43.23 ± 3.12	0.14 ± 0.01	1.14 ± 0.34
GASS_37.2-10.3-256_0	-247.53 ± 0.497	-248.12 ± 0.54	26.02 ± 9.71	7.76 ± 9.86	0.154 ± 0.024	0.182 ± 0.022	1.05 ± 1.59	-250.54 ± 0.67	24.03 ± 1.76	0.197 ± 0.01	0.92 ± 0.23

Table 4.2: Parameters of the double Gaussian components for 2 GASS sightlines.

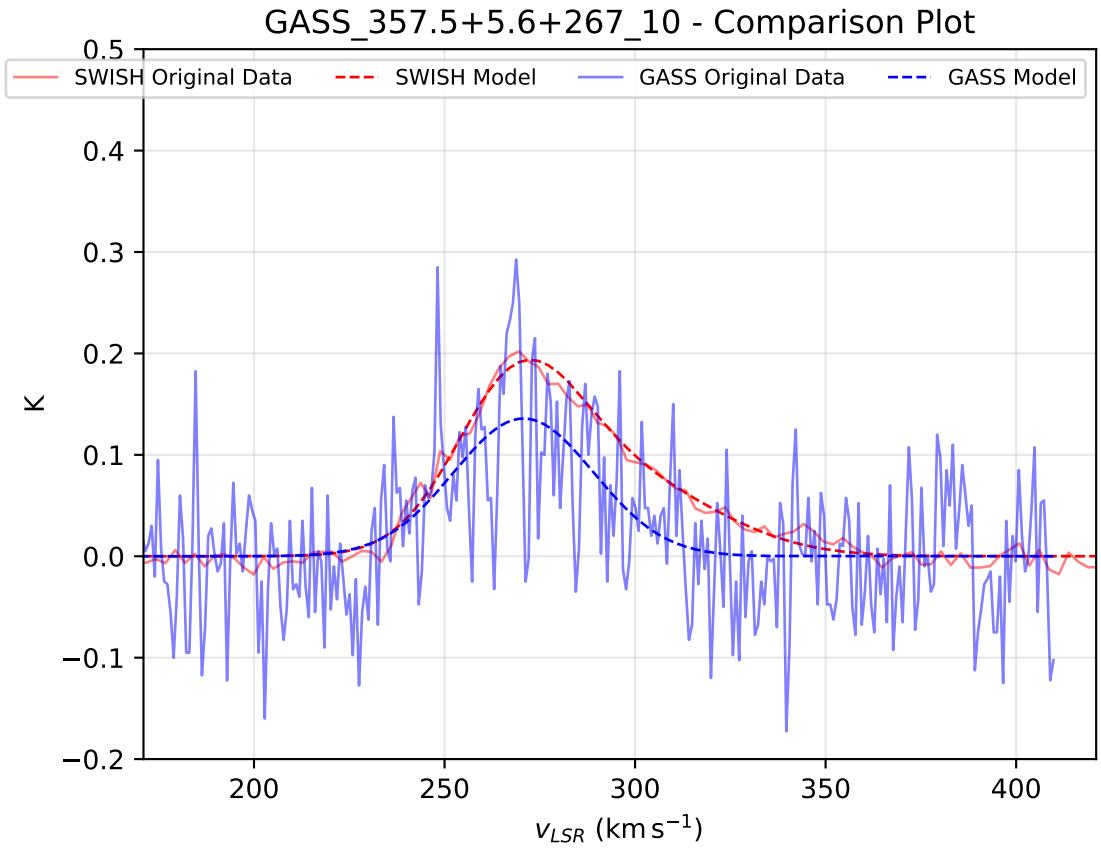


Figure 4.1: Comparison plot showing the previously-catalogued HVC GASS-357.5−5.6+267_10 in both the SWISH and GASS data. The unbroken lines represent the data, and the dashed lines show the best Gaussian fits.

This evidence for a systematic higher brightness temperature in SWISH was further investigated via comparison with values catalogued by Lockman et al. [2002] for the L02 sightlines. (The reduced data was not available for direct fitting.) Of the 20 L02 sightlines reduced in this project, 4 sightlines did not show a detection and preferred the null model. 16 contained detections; however 8 of these sightlines had complex and extended structures, which included capturing bright HVC centres (as well as lower level contamination) in multiple off positions. They were therefore part of the set excluded from this initial analysis, which left 8 sightlines with values that could be compared directly with the L02 catalogue. Figure 4.4 shows the results of this comparison, together with a linear ODR fit of $y = (1.32 \pm 0.19)\,x + (0.046 \pm 0.044)$. (As before, x and y represent the brightness temperature of the L02 and SWISH fits, respectively.) While the number of data points is very small, the two fits are consistent to within the errors. This strongly suggested a real offset in the SWISH data, which needed to be investigated and ideally corrected for before proceeding further with

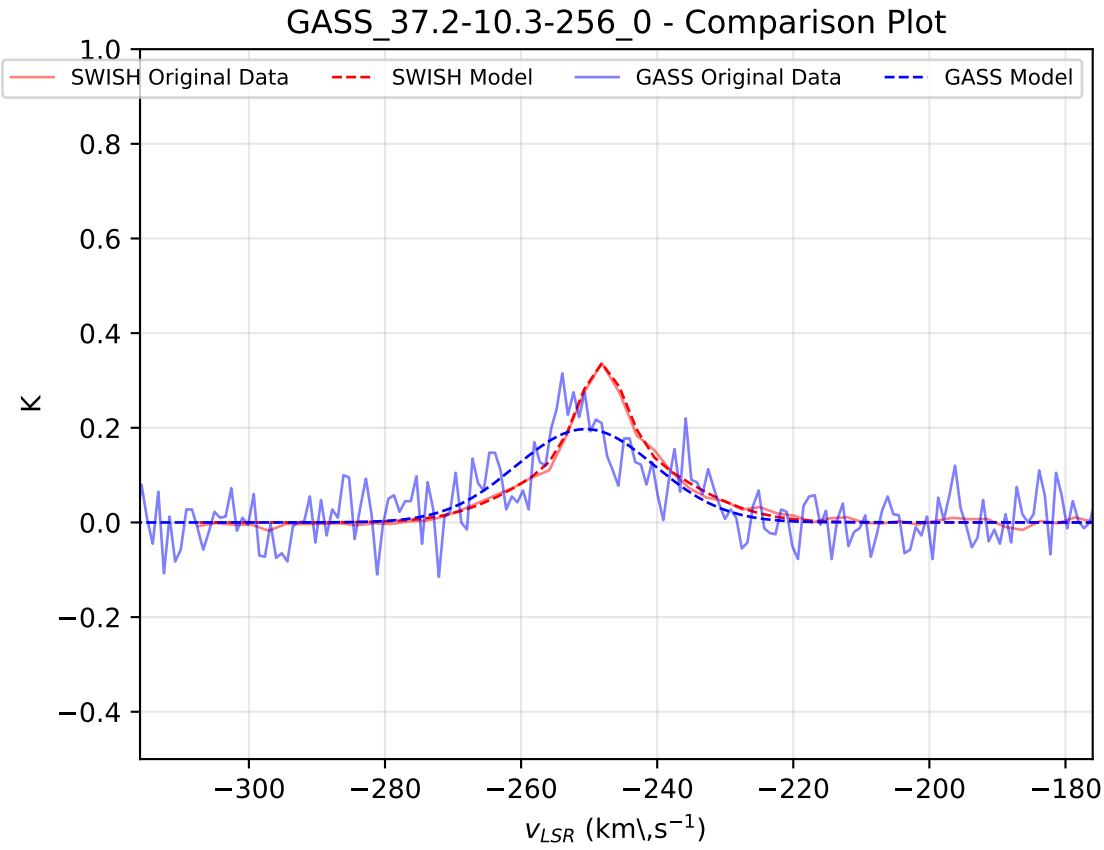


Figure 4.2: Comparison plot showing the previously-catalogued HVC GASS-37.2−10.3+256_0 in both the SWISH and GASS data. The unbroken lines represent the data, and the dashed lines show the best Gaussian fits.

the analysis.

Re-scaling the Brightness Temperature

This systematic scaling difference in the SWISH dataset suggests potential issues in the calibration code inherited prior to the start of this project. This calibration code, which originally used in HIPSR commissioning, was used as-is when processing SWISH data. A preliminary investigation failed to identify any obvious problems: the code correctly identified a suitable calibrator for each sightline, queried model flux densities for each source, performed an appropriate scaling to the data, and computed the system temperature based on the calibrator chosen. Given the compressed timescale of this masters project, it was not prudent to further investigate this problem, but was noted and identified as a high-priority follow up task for future work on the data reduction pipeline. For the purposes of this project,

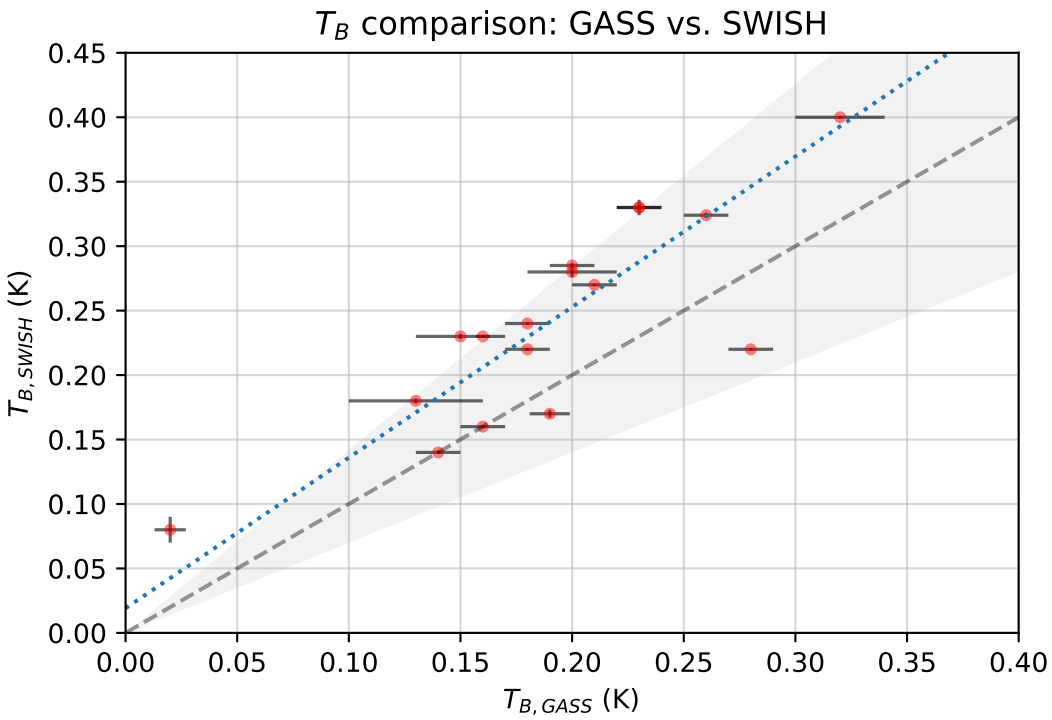


Figure 4.3: Peak brightness temperature values for both the SWISH and GASS data as given by the Gaussian fitting model. The error bars are shown by the horizontal and vertical dark lines on each red data point. The blue dotted line shows the linear fit obtained from orthogonal distance regression, clearly showing there is a systematic difference between the SWISH and GASS data sets. The grey shaded region shows a 30% offset either side of the dashed grey 1:1 line.

the decision was made to manually scale these values according to the line of best fit. From here on-wards, all analysis is performed on these adjusted values, which are tabulated in the appendix in Table B.1.

4.2.2 Velocity Offset

When the Gaussian fits were generated for the sightlines, an obvious offset in velocity became apparent. The results for v_{LSR} from both SWISH and GASS were compared which can be seen in Figure 4.5. The ODR fit generated a best fit line of $y = (0.996 \pm 0.002)\,x + (2.78 \pm 0.35)$. This indicates that there is an offset of 2.78 km s^{-1} of the SWISH data. Since the channel width of the SWISH data is 2.58 km s^{-1}, it was more than likely there was an indexing error when frequencies were converted to velocities in the initial stages of the data reduction. Due

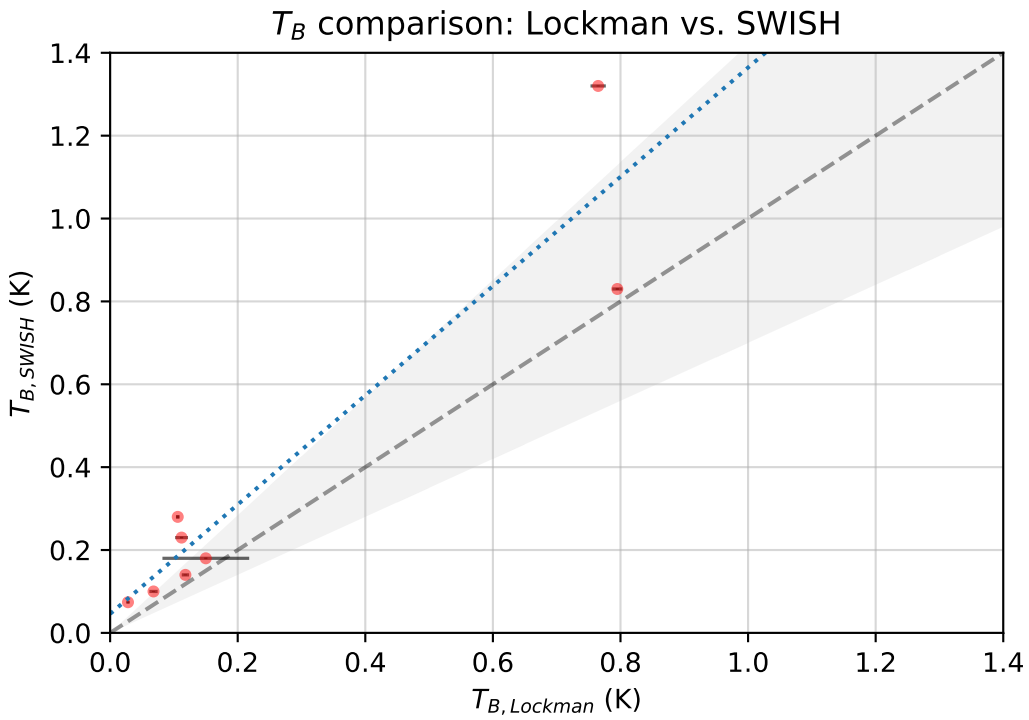


Figure 4.4: Brightness temperature values for the SWISH data as given by the Gaussian fitting model and L02 values as given by Lockman et al. [2002]. The error bars are shown by the horizontal and vertical dark lines on each red data point. The blue dotted line shows the linear fit obtained from orthogonal distance regression, clearly showing there is a systematic difference between the SWISH and Lockman data sets. The grey shaded region shows a 30% offset either side of the dashed grey 1:1 line.

to time constraints, this issue was noted for future work on the pipeline and an offset of one channel was applied to the data going forward.

4.3 Classifying HVCs

The motivation behind the SWISH survey was to investigate the hypothesis that diffuse gas may contribute to up to 75% more mass to the Halo than dense gas. Using the adjusted results from SWISH, our detected clouds were run through the same machine learning algorithm developed in Moss et al. [2017] in order to determine if the clouds observed in SWISH were diffuse in nature. The machine learning algorithm used the SCIKIT-LEARN package [Pedregosa et al., 2012] which classifies in three-dimensions (specifically N_{HI}, T_B and FWHM). Moss

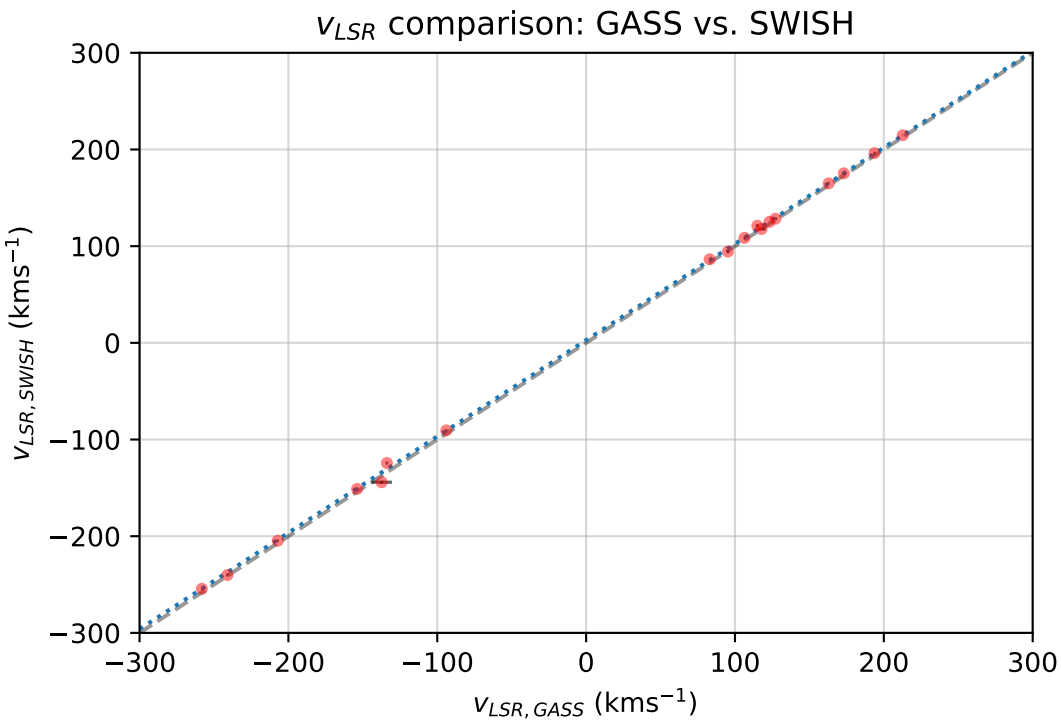


Figure 4.5: v_{LSR} values for the SWISH and GASS data as given by the Gaussian fits. The error bars are shown by the horizontal and vertical dark lines on each red data point. The blue dotted line shows the linear fit obtained from orthogonal distance regression which is almost on top of the 1:1 line.

et al. [2017] separated and classified the cores of GASS HVCs as "dense" and the L02 clouds as "diffuse", and used these two groupings as the initial training set for the algorithm (60% to train and 40% to test). From this, the algorithm then classified a set of HVCs and categorised them as dense or diffuse, which generated parameters for the "average" dense and diffuse cloud. The results of running this classification on the SWISH detections can be found in Table 4.3.

From the results for single-component fits we can see that 63% of the classifications (15/24) returned as dense, 29% were categorised as diffuse (7/24) and 2 clouds were undetermined. A visual comparison of the results of this project and Moss et al. [2017], comparing the three dimensions, can be found in Figures 4.6 and 4.7. This over plot shows that the results of SWISH agree well with the placement in 3D space of the diffuse and dense components from Moss et al. [2017]. Since all SWISH observations were meant to target diffuse regions, this classification result is interesting. The machine learning algorithm was trained

Source Name	Diffuse Probability	Dense Probability	Classification
GASS_10.3-25.0-271_41	0.1478065335	0.8521934665	Dense
GASS_180.6-77.4-116_12	0.1475215108	0.8524784892	Dense
GASS_224.1-34.4+133_44	0.4838609946	0.5161390054	?
GASS_246.2+39.9+312_0	0.7596220128	0.2403779872	Diffuse
GASS_28.2-41.1-221_19	0.2192819752	0.7807180248	Dense
GASS_280.5-56.5+119_19	0.258959589	0.741040411	Dense
GASS_280.5-56.5+119_20	0.2481607192	0.7518392808	Dense
M16.11+	0.07262439235	0.9273756076	Dense
PKS_1009-321-	0.10809563	0.89190437	Dense
PKS_1009-321-	0.05747074435	0.9425292557	Dense
PKS_1011-282+	0.01693192957	0.9830680704	Dense
PKS_1020-103–	0.1236154633	0.8763845367	Dense
PKS_1034-29+	0.3363597171	0.6636402829	Dense
PKS_1034-29+	0.7121444431	0.2878555569	Diffuse
PKS_1101-325+	0.1919282687	0.8080717313	Dense
PKS_1117-248+	0.5927255434	0.4072744566	Diffuse
PKS_1207-399-	0.4580123658	0.5419876342	?
S9	0.2193424792	0.7806575208	Dense
SWISH_054411-430611	0.5612610477	0.4387389523	Diffuse
SWISH_065600-585145	0.8666045895	0.1333954105	Diffuse
SWISH_092454-085156	0.3808779564	0.6191220436	Dense
SWISH_154044-292316	0.6729980418	0.3270019582	Diffuse
SWISH_194846-153607	0.5759649768	0.4240350232	Diffuse
SWISH_211621-112259	0.2856986851	0.7143013149	Dense

Table 4.3: Results from the machine learning algorithm as run on the Gaussian fit parameters for SWISH HVCs. Each cloud is assigned a probability of being either dense or diffuse with the most likely result being taken as the final classification. The question mark indicated the code was uncertain of the classification.

on the assumption that L02 sightlines were "diffuse", yet has classified them in SWISH as dense. Similarly, the previously-known GASS targets selected for this project were chosen to be regions surrounding the cores of known HVCs that had been previously classified as "diffuse" by this algorithm. This presents initially as a potential problem with the SWISH data, as one would expect the classifications not to change depending on the data set used.

One possible factor is that the data reduction pipeline in its current form may not remove contaminated "off-beams" around the wings of HVCs as effectively as thought, which could possibly lead to a reduction in the measured FWHM. During the removal of bad data, off-source spectra were sometimes encountered with slight velocity offsets that would reduce the wings of an HVC detection if not removed. The identification and removal of these spectra proved especially difficult for the L02 clouds due to the low brightness temperature of the emission (on both on-source and contaminated off-source spectra), and often a decision had to be made on a case-by-case basis on whether to keep the beams to reduce the overall noise and produce a detection, or to simply abandon the sightline and classify it as an "off beam" detection. This theory is supported by the L02 catalogued FWHM values, which showed SWISH FWHMs were lower by typically $\sim 20\%$. While catalogued values were not available for the sightlines previously detected in GASS (since Moss et al. 2013 only tabulated Gaussian fit parameters for HVC peaks), it is also possible that these sightlines are similarly affected. However without an independent method of checking the accuracy of the generated parameters, it is unclear of the percentage of other sightlines affected. For the moment, we may note this again as an issue for thorough investigation when the reduction strategy is returned to in future work.

4.3.1 SWISH Blind Sightlines

Of the blind sightlines taken in SWISH we found a detection rate of 26% of which 67% are classified as "diffuse" according to the machine learning algorithm. The locations of the SWISH sightlines analysed in this project can be shown on Figure 4.8, where the type of detections are indicated on the map.

The SWISH "blind" sightlines had previously been classified as non-detections in GASS, and while the Bayesian fitting routine used in this project does favour single component detections in some of the GASS data for these sightlines, the signal-to-noise is poor enough that detections would not have been made using a simple 4-sigma threshold (as performed in

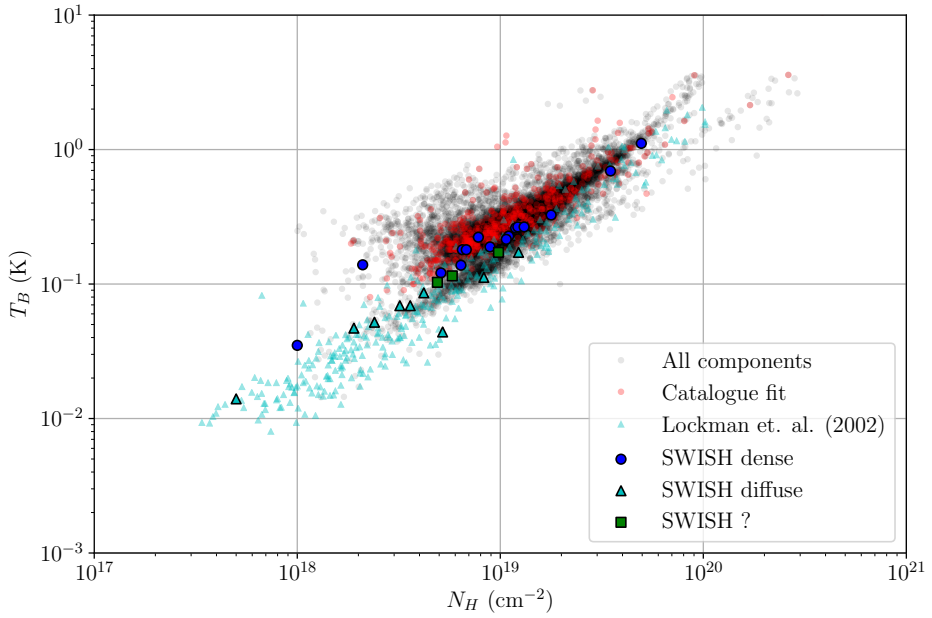


Figure 4.6: Results from the machine learning classification of the SWISH data overplotted on the scatter plots of both the GASS HVC population and the L02 population of -43° $< \delta <$ 0° (the common region observed by both surveys) Moss et al. [2017]. The values of brightness temperature and column density are on the axes (two parameters used in the classification). The all components label refers to all data collected in Moss et al. [2017].

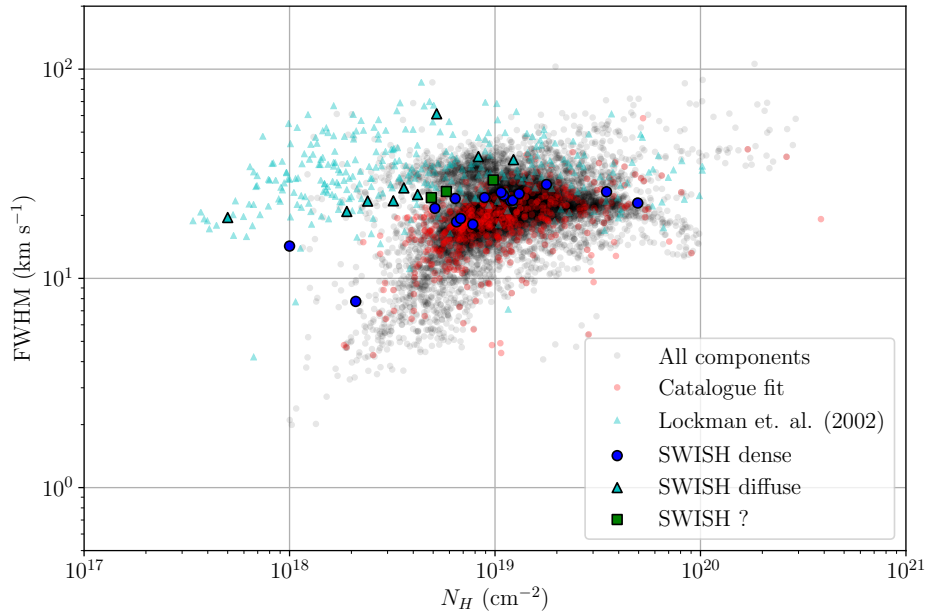


Figure 4.7: Results from the machine learning classification of the SWISH data overplotted on the scatter plots of both the GASS HVC population and the L02 population of -43° $< \delta <$ 0° (the common region observed by both surveys) Moss et al. [2017]. The values of FWHM and column density are on the axes (two parameters used in the classification). The all components label refers to all data collected in Moss et al. [2017].

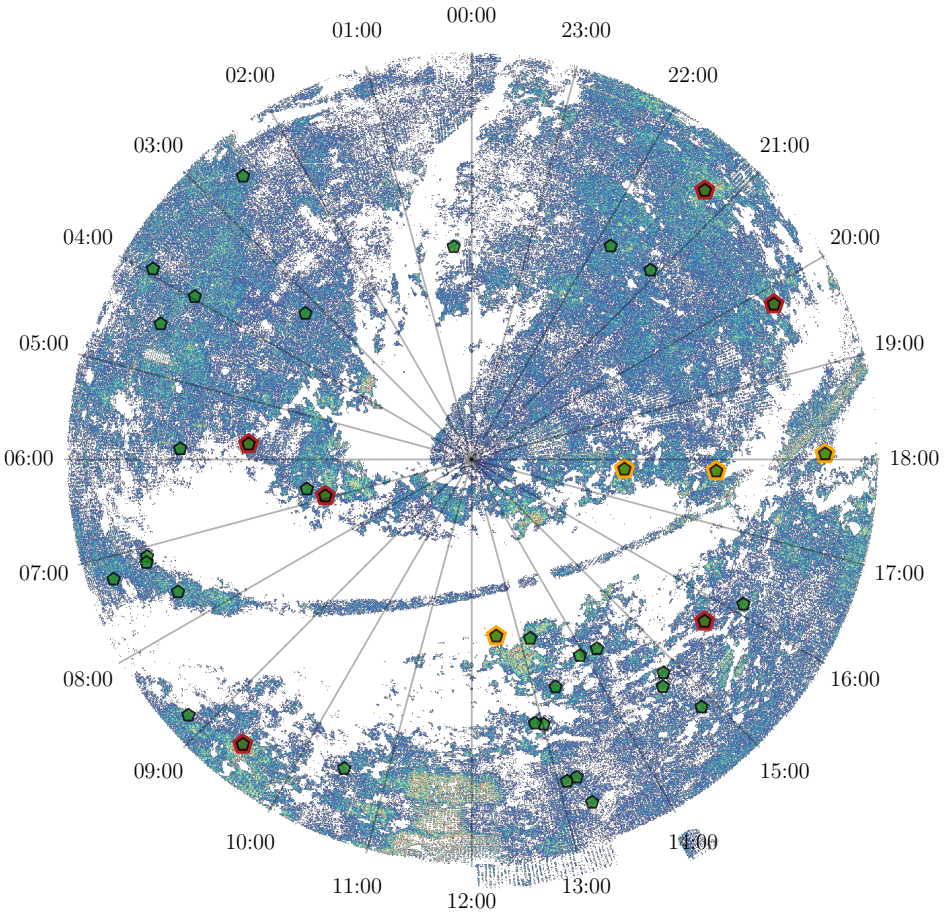


Figure 4.8: Map of the southern hemisphere of the sky with green markers indicating SWISH blank sightlines which were analysed. The red border indicates a confirmed detection with an associated Gaussian model successfully applied to the detection, and orange for a detection found in the off-beams. The white region indicates masked high velocity HI emission and the blue and green regions indicates regions expected to be noise base on GASS sensitivity.

Moss et al. [2013]. Additionally, although our modelling software does indicate a detection, it is provided with the velocity range from a prior SWISH detection. Without this prior knowledge it is unclear whether a robust detection could have been made from GASS data alone.

4.3.2 GASS Double Components

From the two re-observations of GASS-357.5−5.6+267_10 and GASS-37.2−10.3+256_0, double Gaussian components were detected which were not present in the GASS data. These

Source Name	Diffuse Probability	Dense Probability	Classification
GASS_357.5+5.6+267_10	0.8818458816	0.1181541184	Diffuse
GASS_357.5+5.6+267_10	0.9989219393	0.001078060678	Diffuse
GASS_37.2-10.3-256_0	0.4880687234	0.5119312766	?
GASS_37.2-10.3-256_0	0.01288601378	0.9871139862	Dense

Table 4.4: Results from the machine learning algorithm for the Gaussian fit parameters of SWISH double component cloud detections. Each cloud component is assigned a probability of being either dense or diffuse with the most likely result forming the final classification. The question mark indicated the code was uncertain of the classification.

components were run through the classification algorithm and the results can be found in Table 4.4. By comparing GASS-357.5−5.6+267_10 in both SWISH and GASS, there is obviously a broad, diffuse, component detected in SWISH which has been lost in the GASS data due to the high noise levels. This is equally apparent in GASS-37.2−10.3+256_0. Both of these GASS sightlines were targeted toward the diffuse peripheries of known HVCs. This discovery has raised the question: would a similar diffuse gas component be detected when observing the cores of HVCs? To investigate this further, re-observations of known catalogued HVCs should be made at the sensitivity of SWISH. When the components are being decomposed, a statistically robust method of selecting models should be employed to ensure that these underlying components are not missed. Similar to the way that the SWISH blind observations were analysed, we can make assumptions on how much mass these diffuse components make up of an HVC.

4.4 The Missing Diffuse Gas

Whether the new HVCs SWISH has identified are "diffuse" or not is yet to be confirmed, but it is clear there is low column density HI found in SWISH, previously undetected in GASS. We have found two examples of HVCs that have underlying broad components which potentially contribute to up to 40% of the HVCs total column density. Although this has only been found in two clouds, it is indicative that diffuse gas is present but currently undetected in the large all-sky surveys as broad components may lie under the current noise threshold of current surveys. From the blank sightlines, a new detection rate of 26% was discovered of which

the analysed clouds have an average column density of $\sim 4 \times 10^{18}$ cm^{-2}. Depending on the extent of the diffuse gas within the Halo, these preliminary findings from SWISH do confirm that there is an underlying HI population which may contain significant amounts of mass. Knowing that classical HVC populations do not have the mass to support the calculated star formation rate, uncovering HVCs which have low column density and potential sky coverage of 26% supports findings by the UV community of a ubiquitous, diffuse population of HI containing significant mass [Murphy et al., 1995].

Don't let anyone rob you of your imagination, your creativity, or your curiosity.

Dr Mae Jemison

5

Conclusions and Future Directions

Throughout the period of this Masters of Research degree, an investigation was conducted into HI clouds in the Milky Way Halo, specifically focusing on observations conducted by the Survey of Weak Southern HI. The goals of this project were to:

- *Produce a semi-automatic pipeline to reduce SWISH data.* The data reduction pipeline which was created is detailed in Chapter 2, and was used to successfully process a subset of data from the SWISH project. The data reduction strategy will continue to be refined in future work.

- *Perform a comparative analysis of a subset of common sightlines using data collected from SWISH, GASS and LO2.* From the subset of observations made during the first semester of SWISH, several systematic differences were discovered. A velocity offset indicated a channel assignment issue and a brightness temperature scaling difference was discovered during the data analysis. It is very likely these problems arose during the data reduction stage of this project. These issues were discussed and highlighted as issues which should be investigated when work on the pipeline continues.

- *Determine if SWISH produces results which agree with Moss et al. [2017], which compares the results from two HVC surveys.* The results from Moss et al. [2017] suggests that 74% of HVCs are diffuse in nature and could contain 2-3 times the mass of the dense components. The results from this thesis found that GASS indeed did miss lower column density HVC emission along 26% of apparently blank sightlines, as confirmed through the blind sightline observations. Double components were also newly detected in two clouds which contained 40% of the total column density for that cloud. This results strongly supports Moss et al. [2017] regarding the missing mass in GASS.

- *Discuss the results of SWISH and assess the possible implications these results have in the wider context of Galactic evolution.* The improved sensitivity of SWISH has uncovered a "hidden" population of low column density gas which is below the detection limit of GASS. Our observations align with conclusions made through UV surveys that ubiquitous HI is present in the Halo, currently undetected by larger all-sky surveys of neutral hydrogen.

- *Determine the future directions of this research from the results of the analysis.* Chapter 4 highlights the need for more intelligent strategies when dealing with "contamination" and the distribution of detections across both "on" and "off" beams. However this project had identified lower column densities of HI that should be investigated with follow up research into HVCs.

5.1 Future Directions

Given the limitations encountered in the on/off position switching method (even with the techniques applied in this thesis), we will also explore the possibility of applying other methods to reduce the data in future. Since the method used in these observations samples parts of the sky multiple times, we can make use of the information contained in the multiple spectra through techniques such as the least squares method or forward modelling via information field theory. Early experiments using the least squares method in a simulated idealised SWISH like scenario have shown promise with good recovery of contaminated signals. Alternative new techniques for reducing the SWISH data may, if successful, have direct implications for

the ways we approach calibration and visualisation of HI emission spectra in future surveys of faint HI, especially those conducted with multibeam (FAST) or phased-array-feed (Parkes) receivers.

The full SWISH sample expands on the SWISH-100 dataset presented here, to extend the observed sample from verification sightlines to other classes of sightline, namely sightlines chosen for overlap with UV absorption data (RICHTER, COS-UV), a grid over the whole sky (GRID) and stripes across a small set of clouds (STRIPE). Upon analysis of this we will be able to verify whether the preliminary results presented here hold when the sample is larger. The complete map of SWISH sightlines can be found in Figure 5.1.

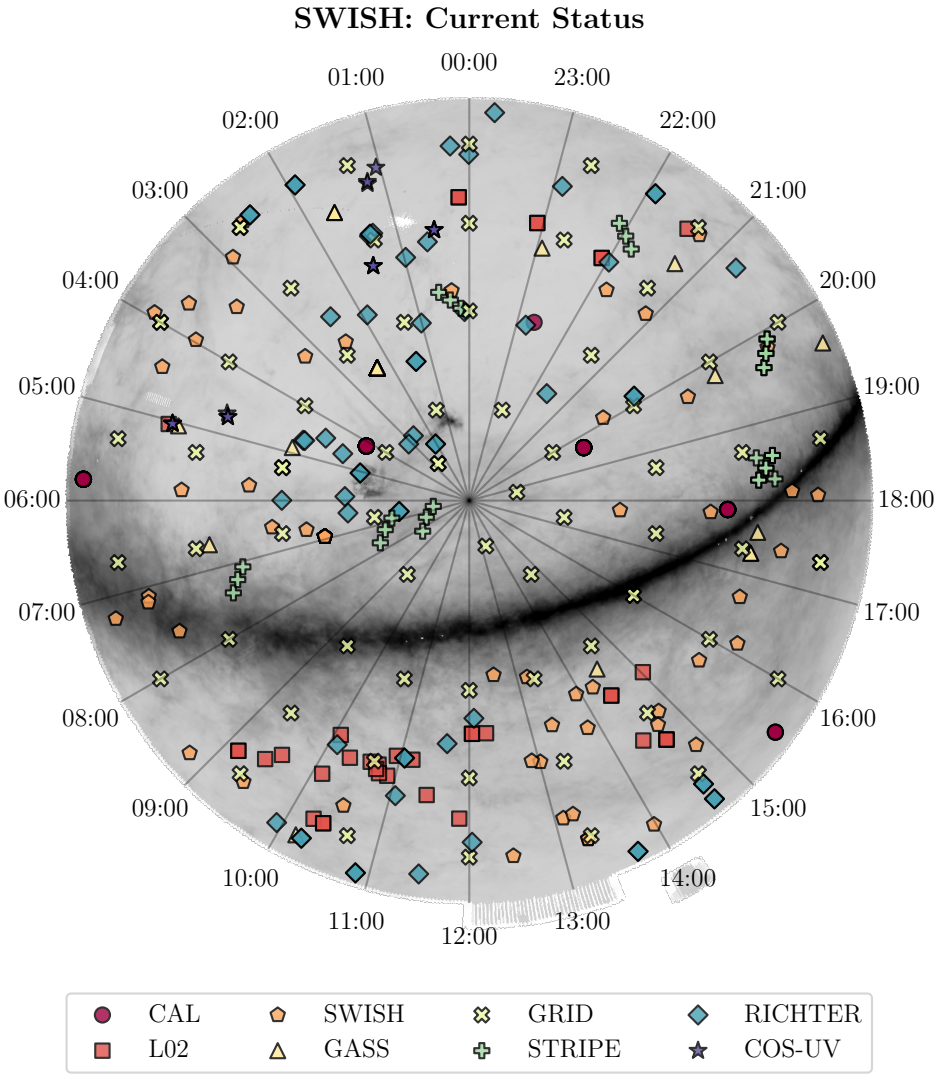
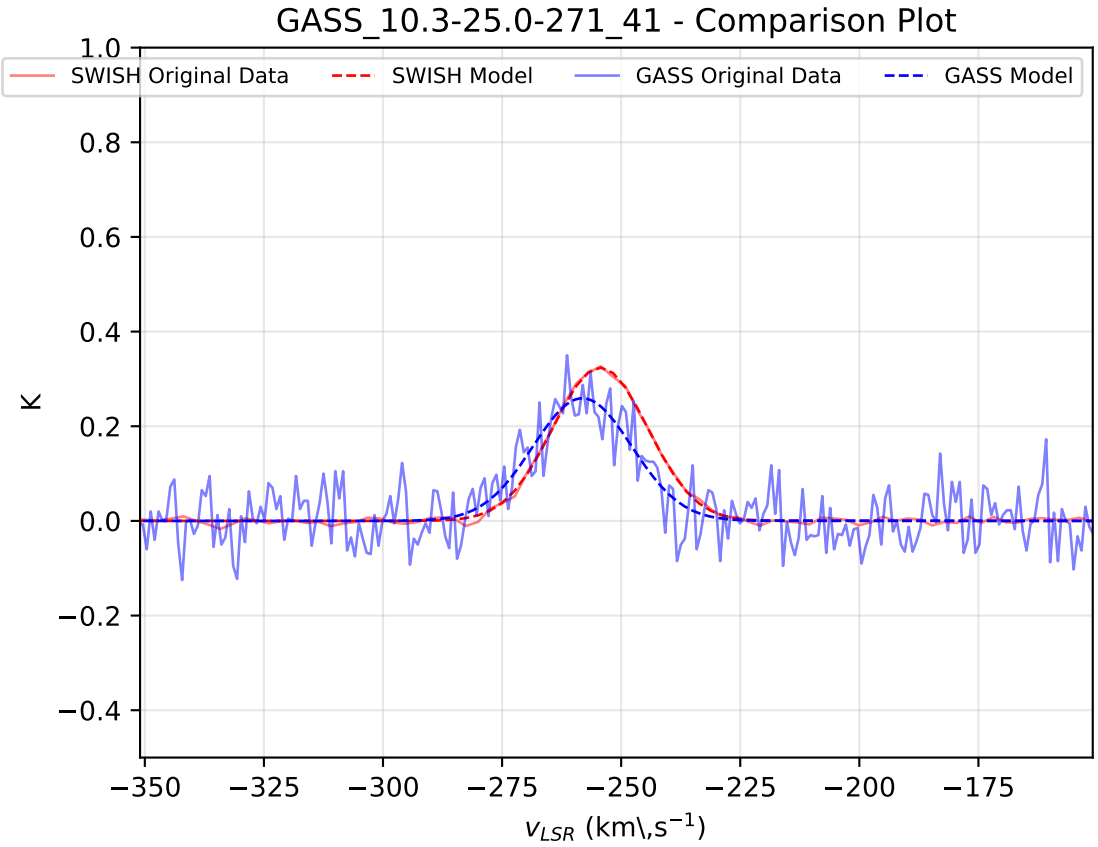
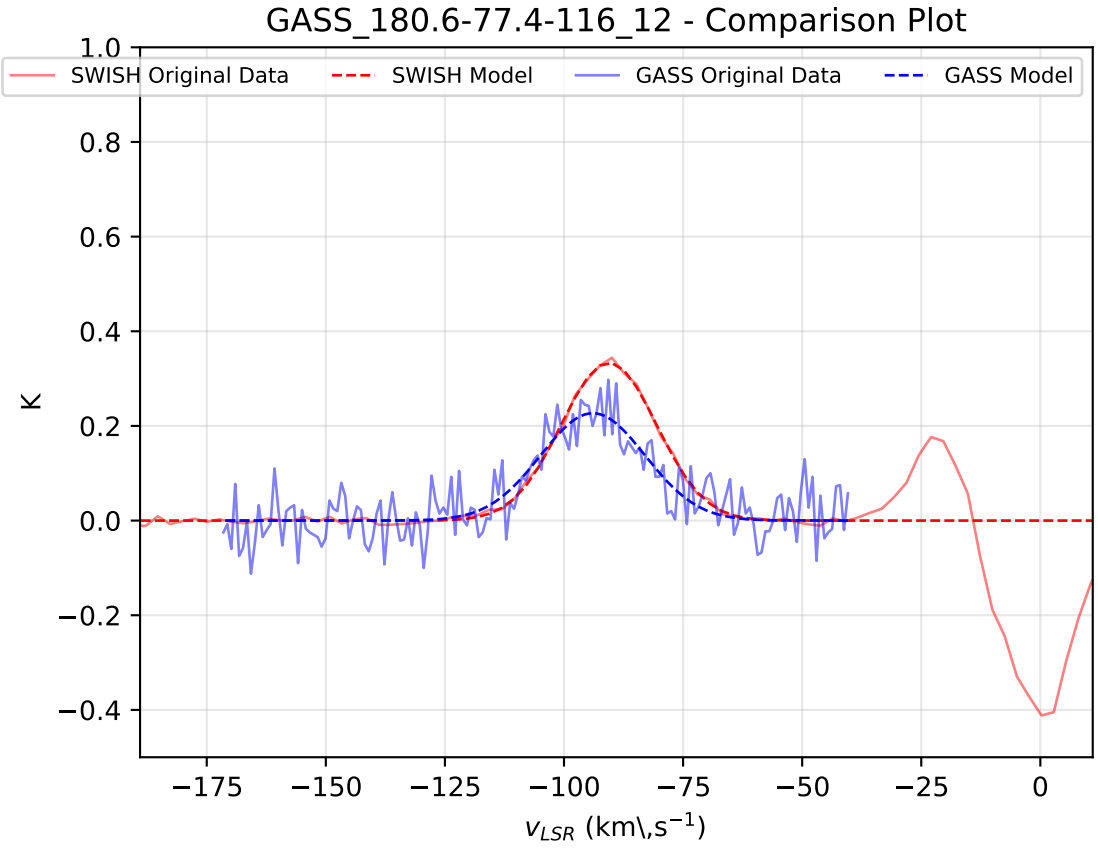


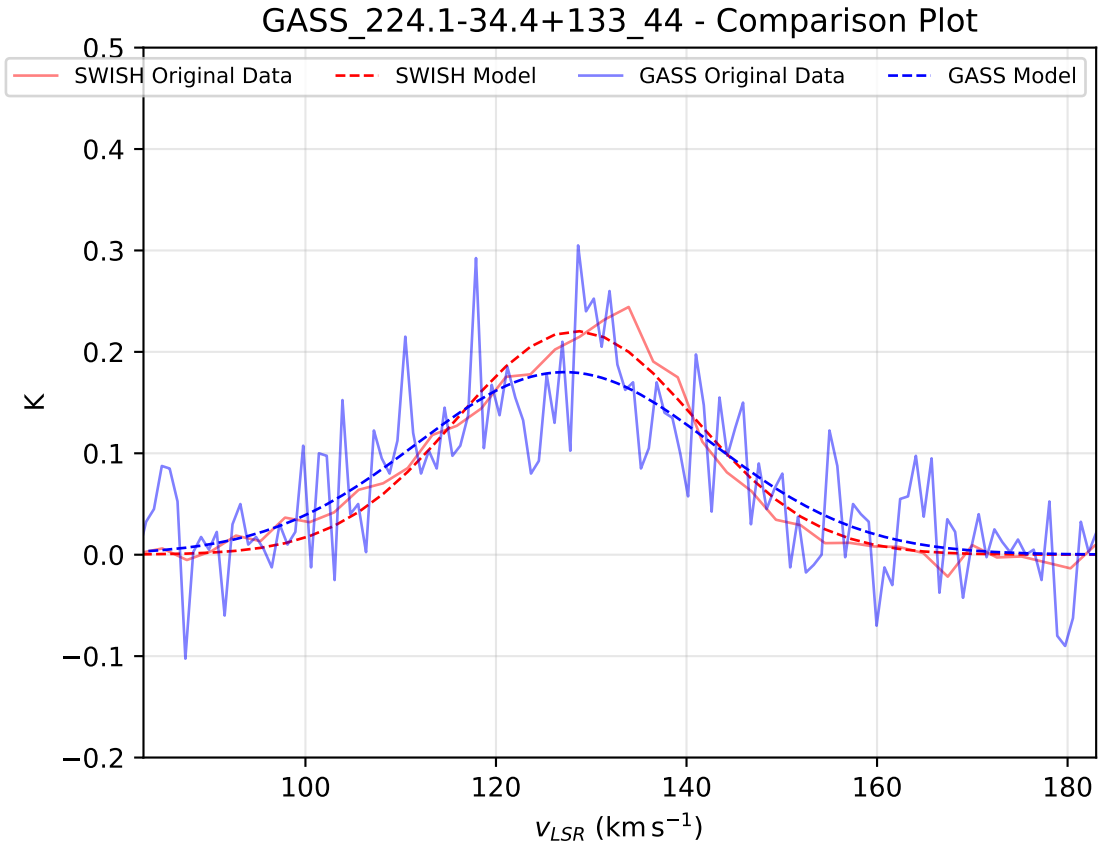
Figure 5.1: A full map of all targets in SWISH over-plotted on an integrated intensity map of HI in the Southern hemisphere. The meanings of the symbols SWISH, GASS and L02 are as used throughout this thesis; the meanings of GRID, STRIPE, RICHTER and COS-UV are as explained in the text. CAL are calibration sources.

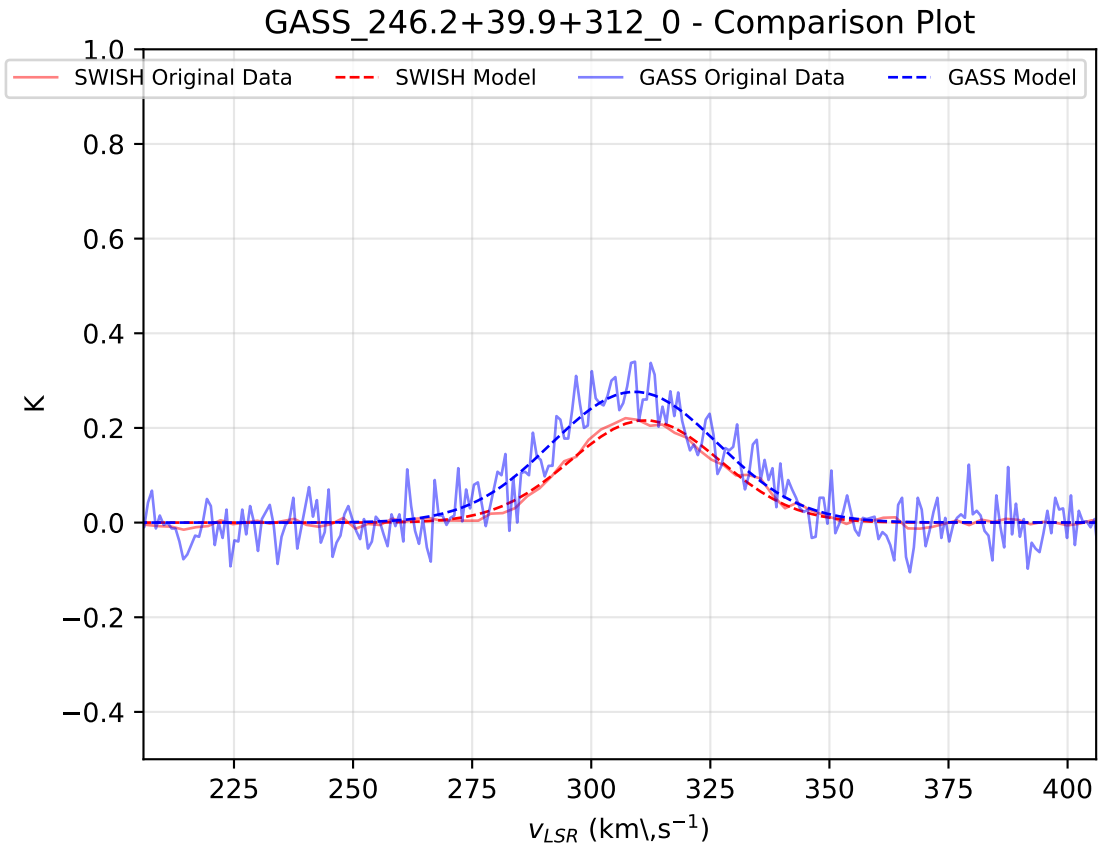
A

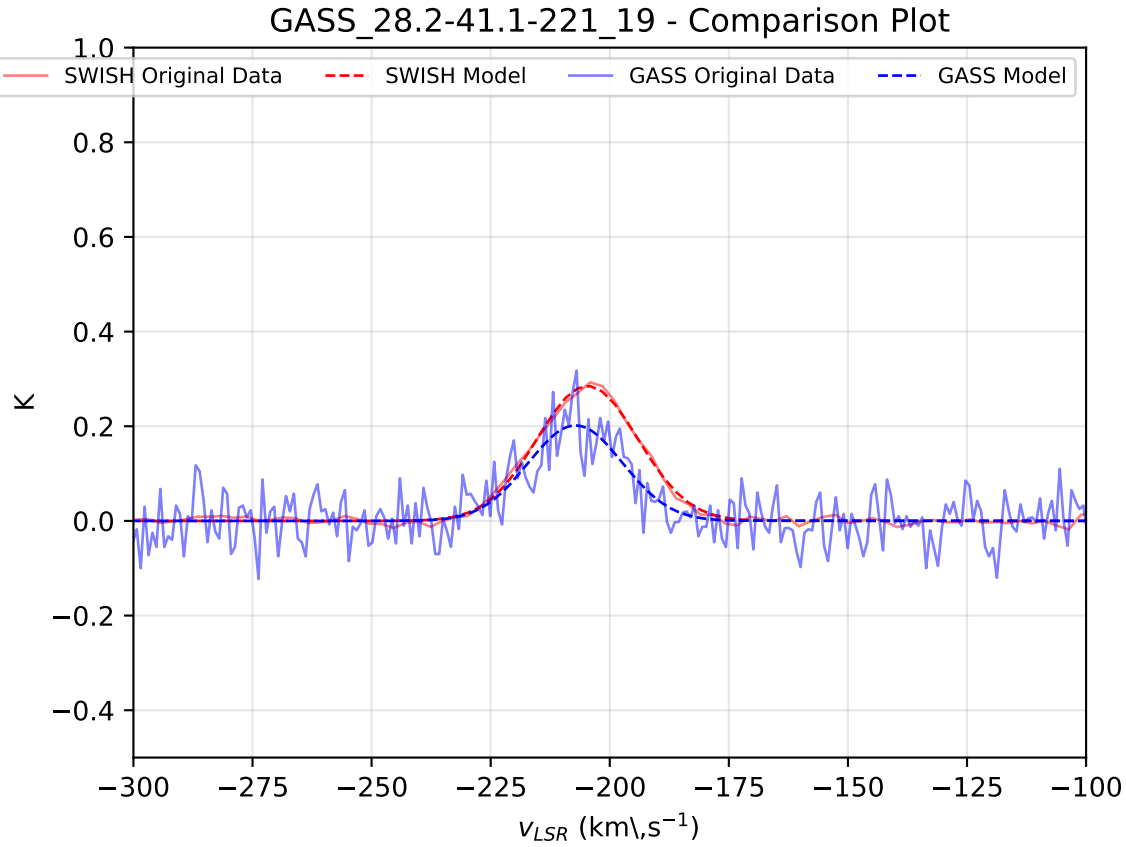
Gaussian Fit Comparisons

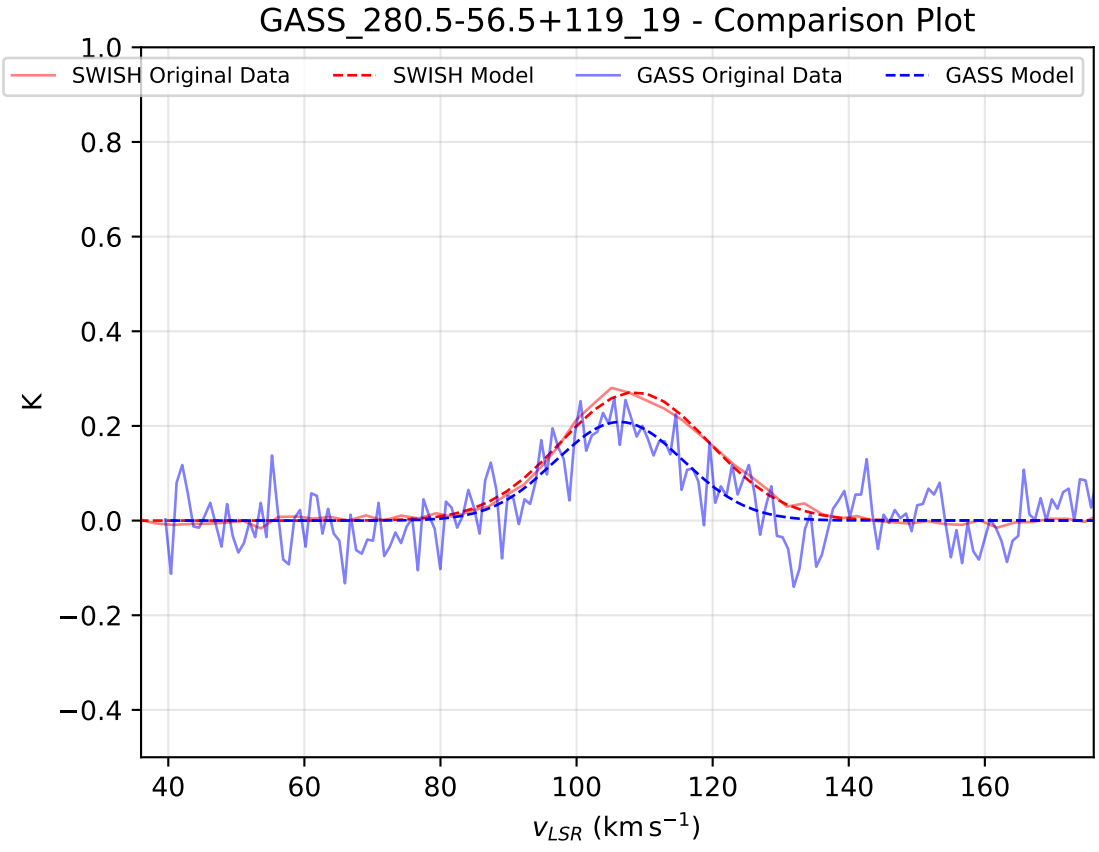

GASS_10.3-25.0-271_41 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})

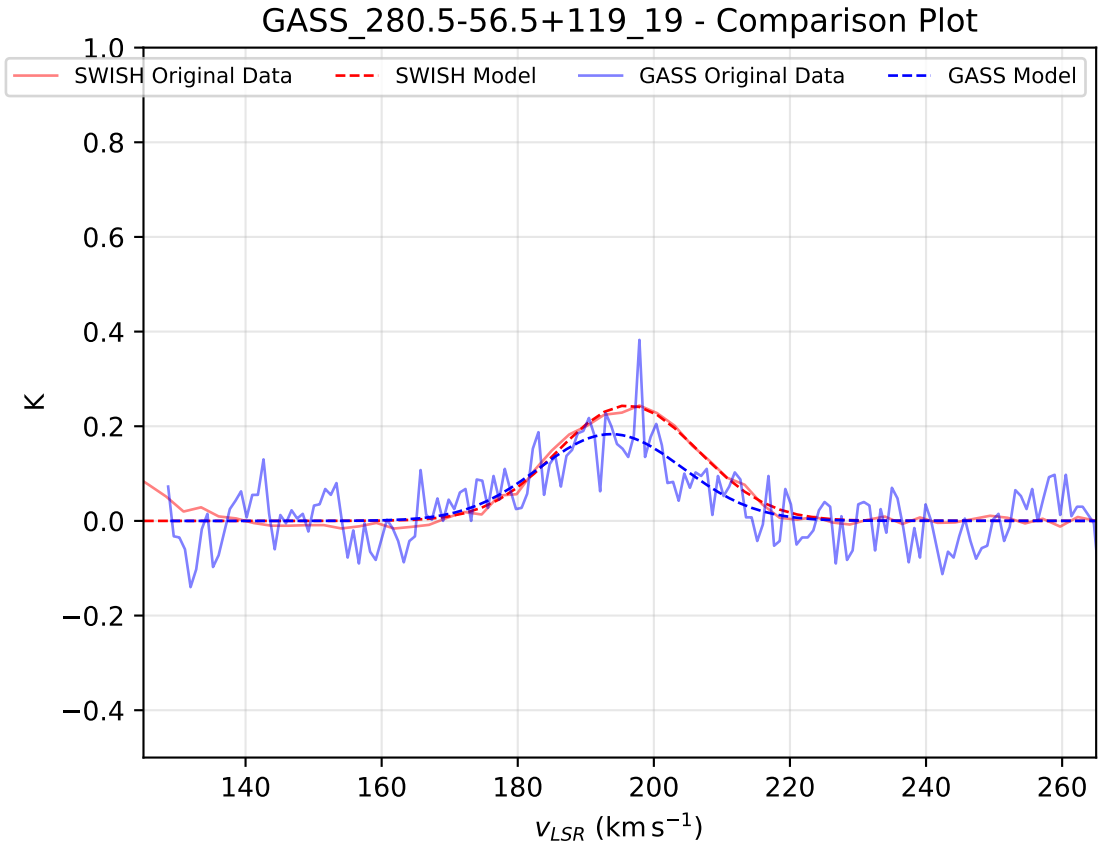

GASS_180.6-77.4-116_12 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})

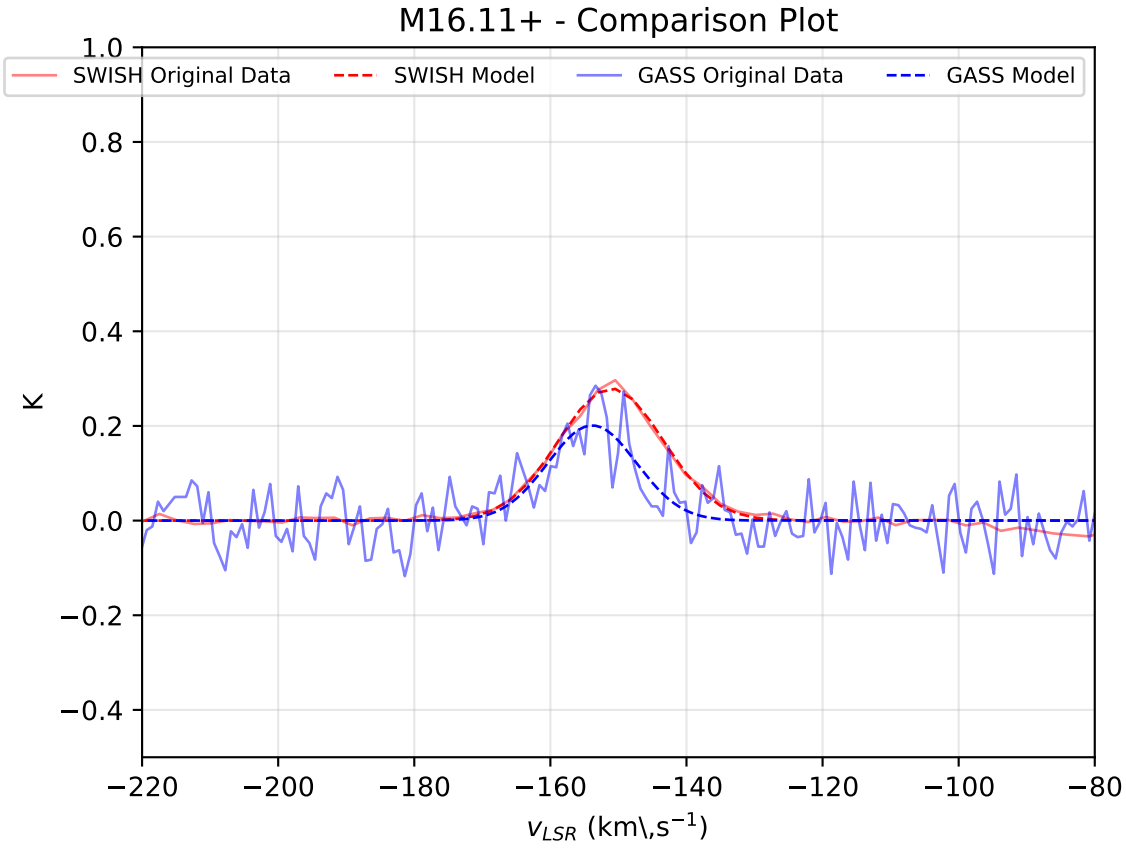

GASS_224.1-34.4+133_44 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km s^{-1})

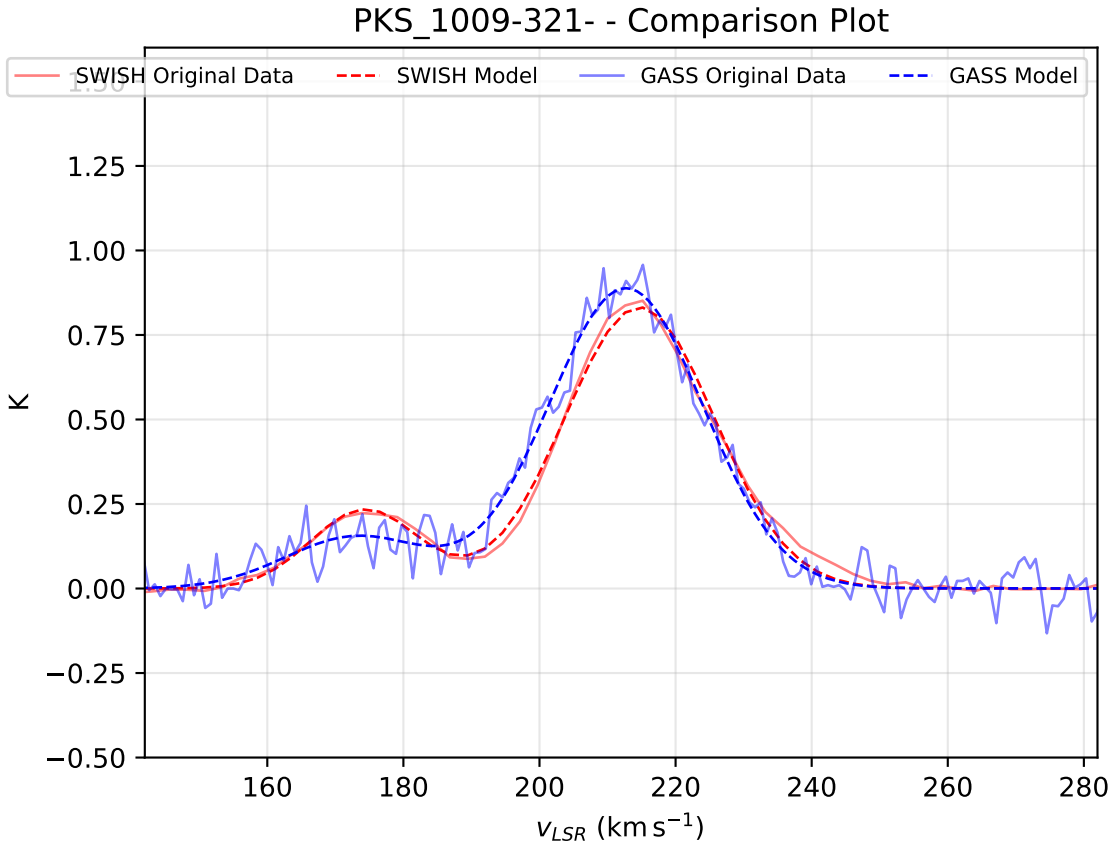

GASS_246.2+39.9+312_0 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})

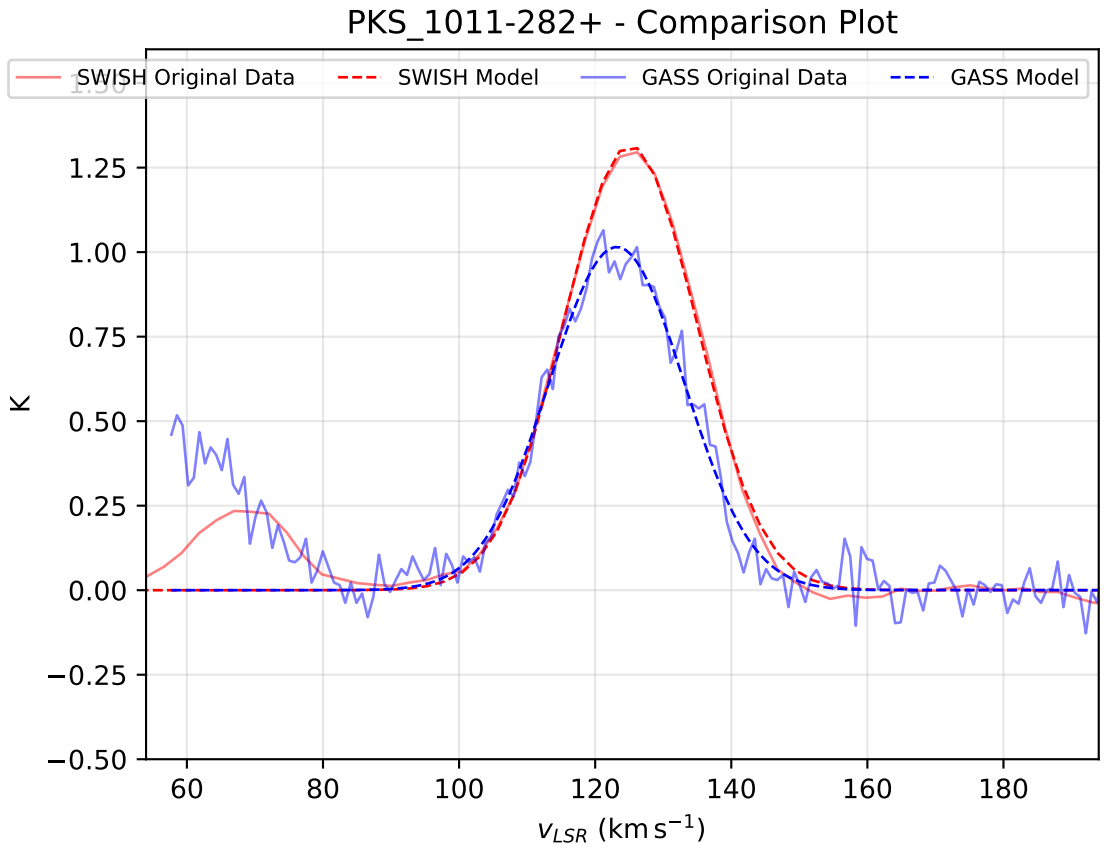

GASS_28.2-41.1-221_19 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})

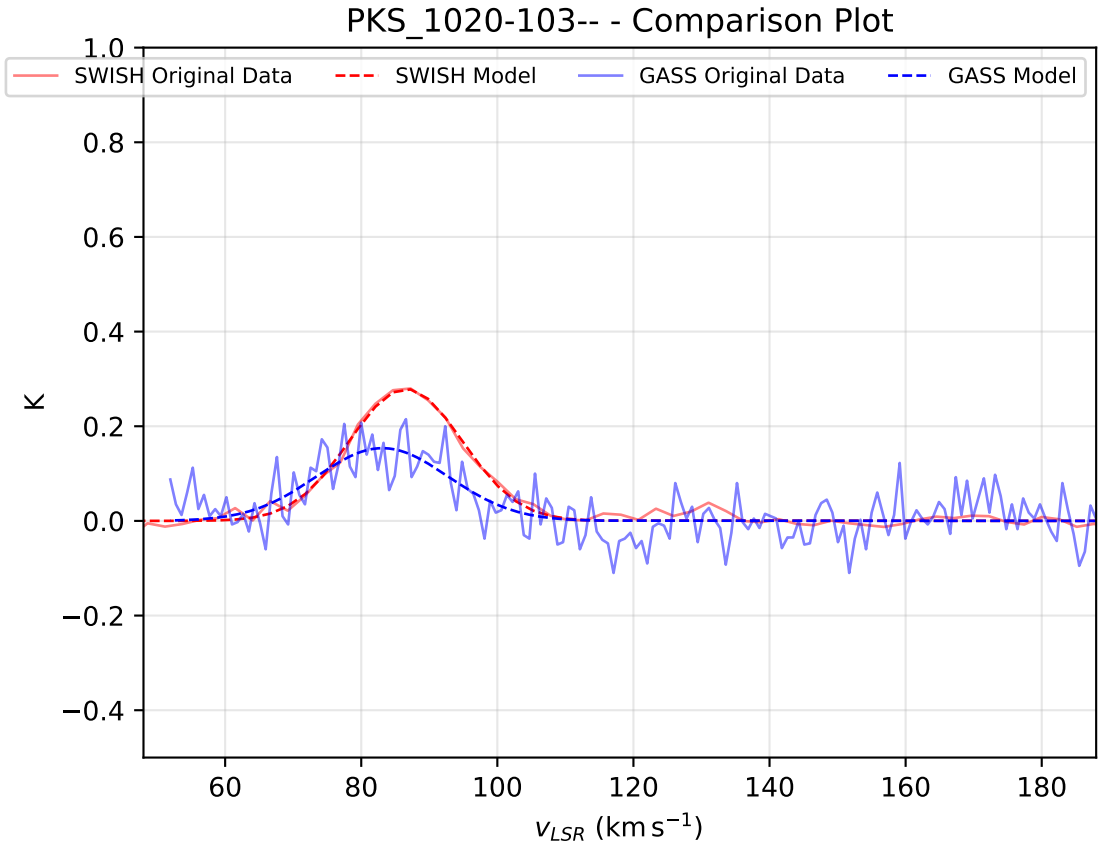

GASS_280.5-56.5+119_19 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km s^{-1})

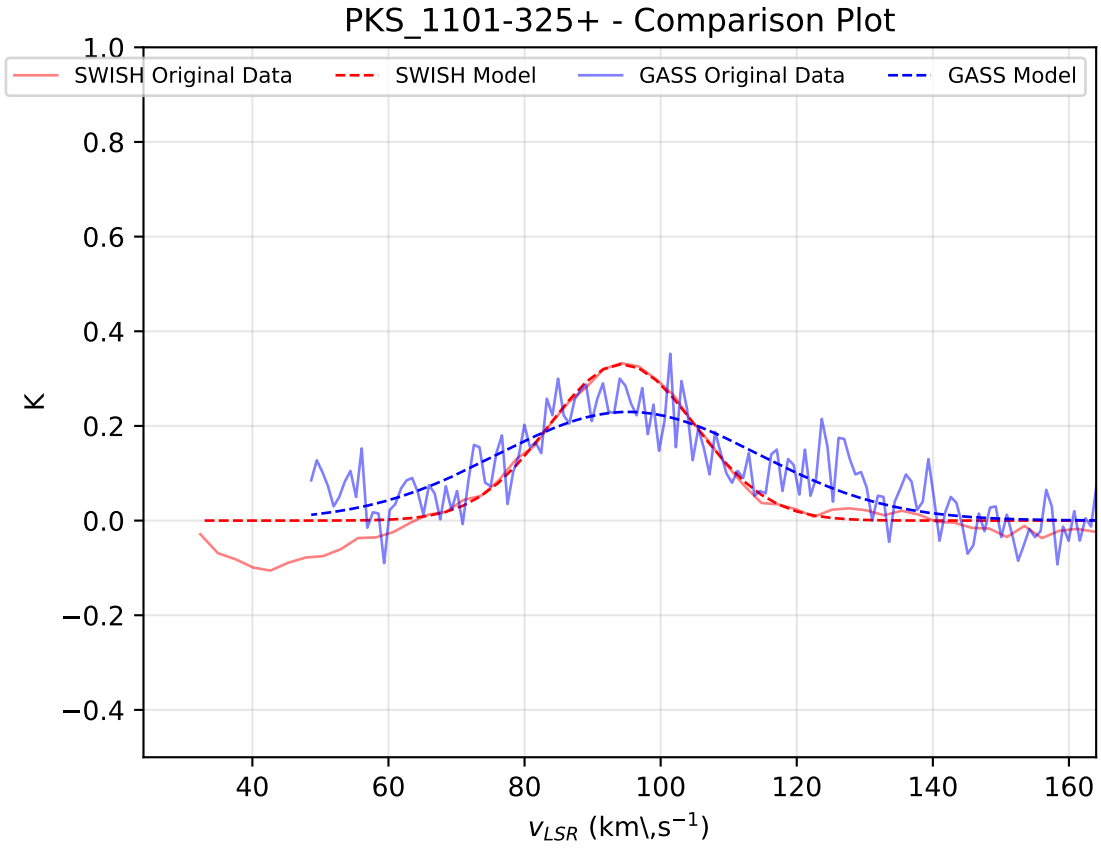

GASS_280.5-56.5+119_19 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
140
160
180
200
220
240
260
v_{LSR} (km s^{-1})

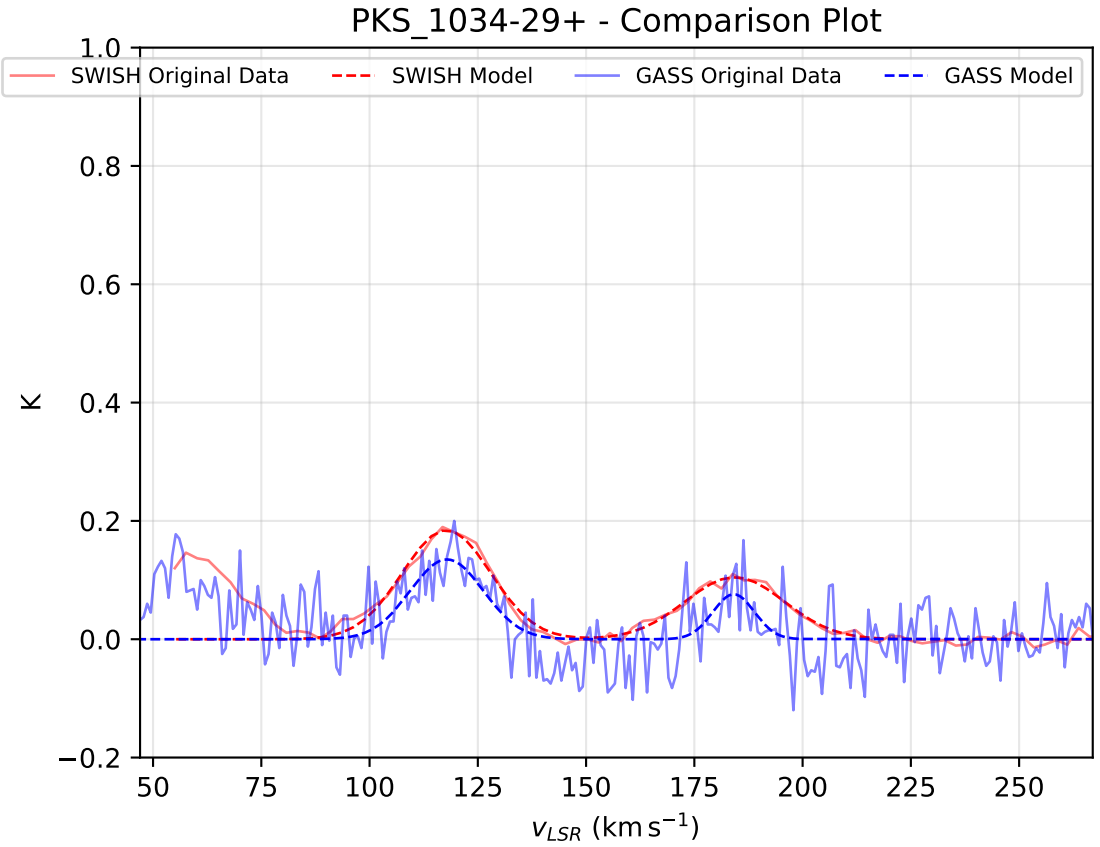

M16.11+ - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−220
−200
−180
−160
−140
−120
−100
−80
v_{LSR} (km\,s^{-1})

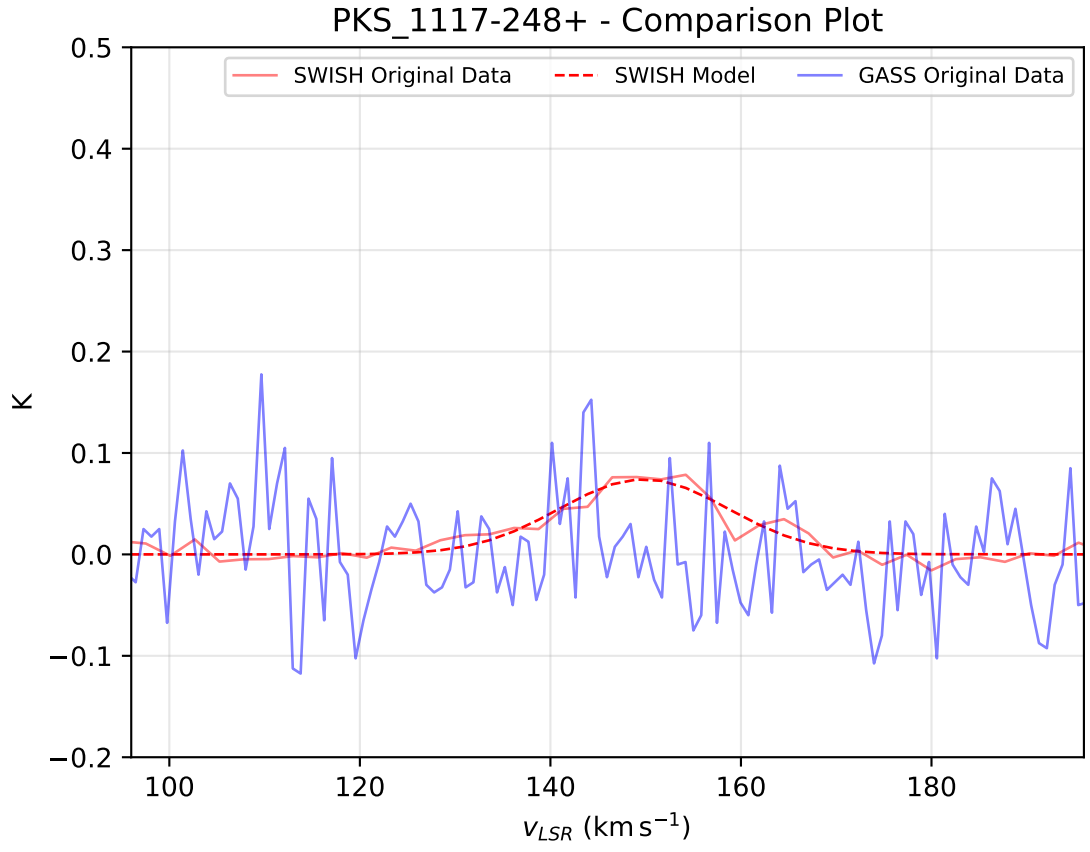

PKS_1009-321- - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
160
180
200
220
240
260
280
v_{LSR} (km s^{-1})

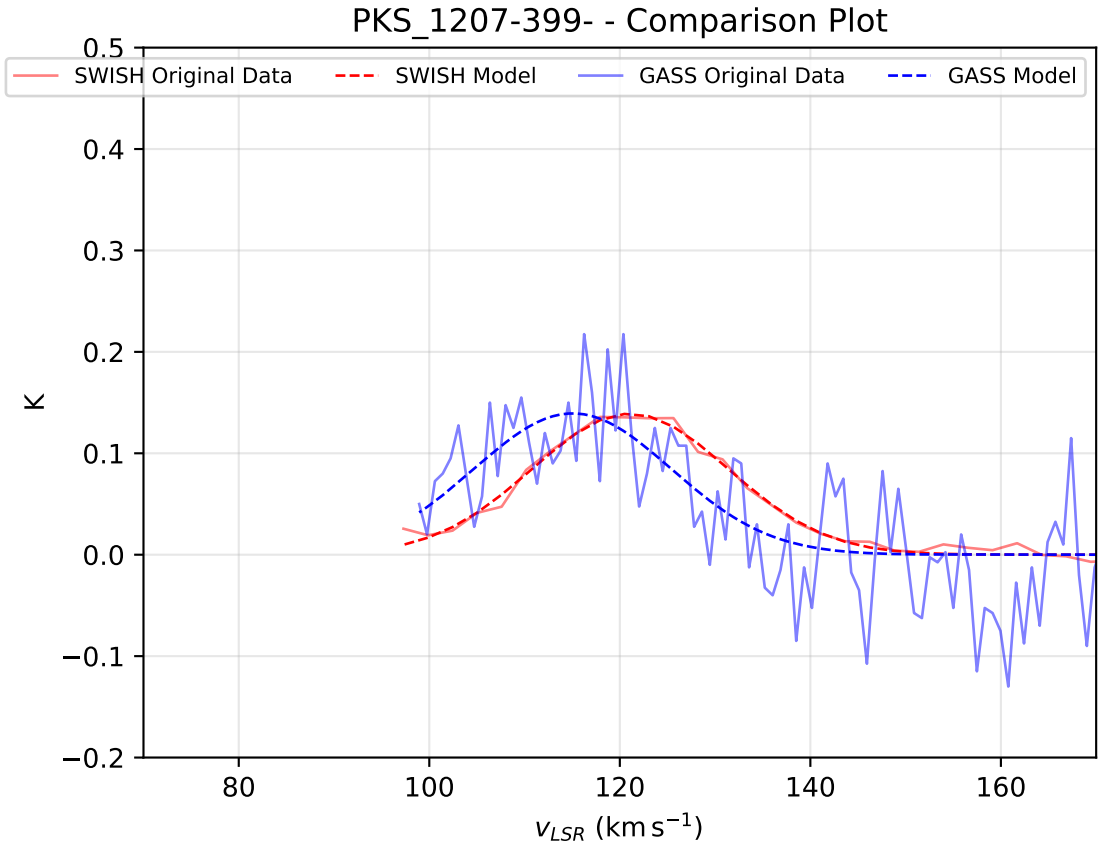

PKS_1011-282+ - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
60
80
100
120
140
160
180
v_{LSR} (km s^{-1})

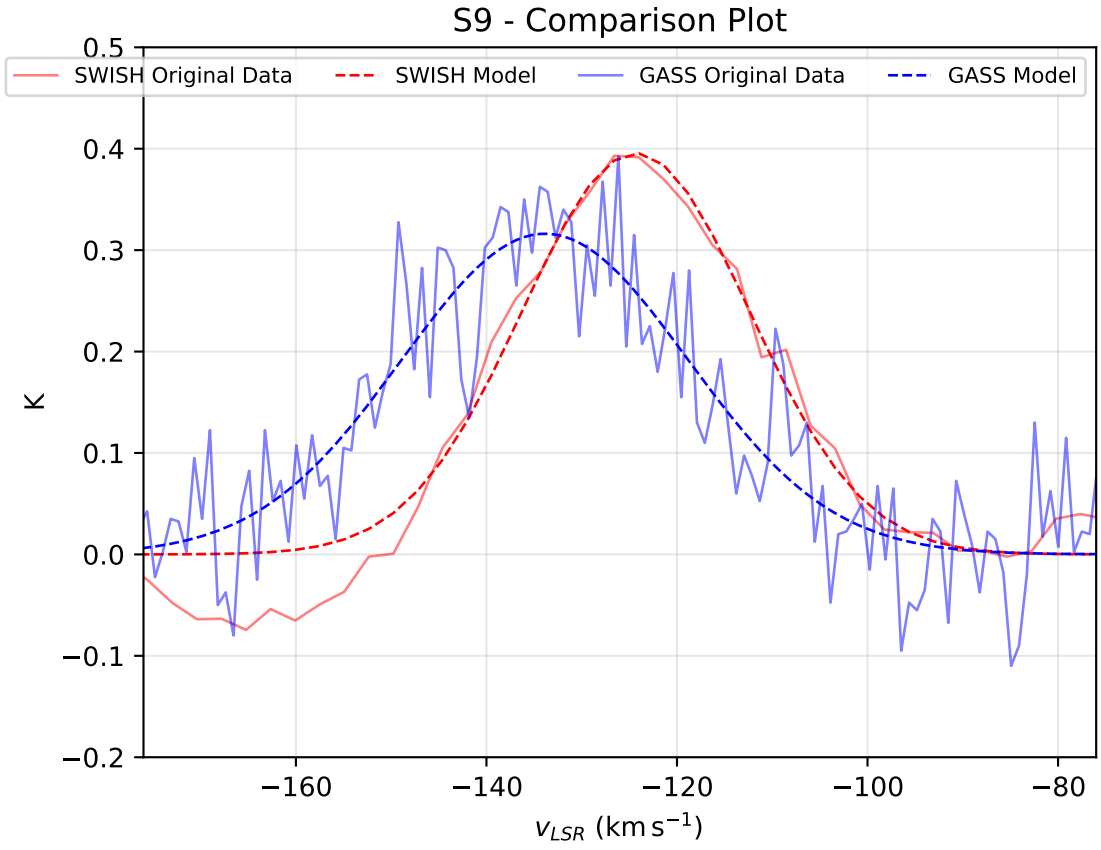

PKS_1020-103-- - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km s^{-1})

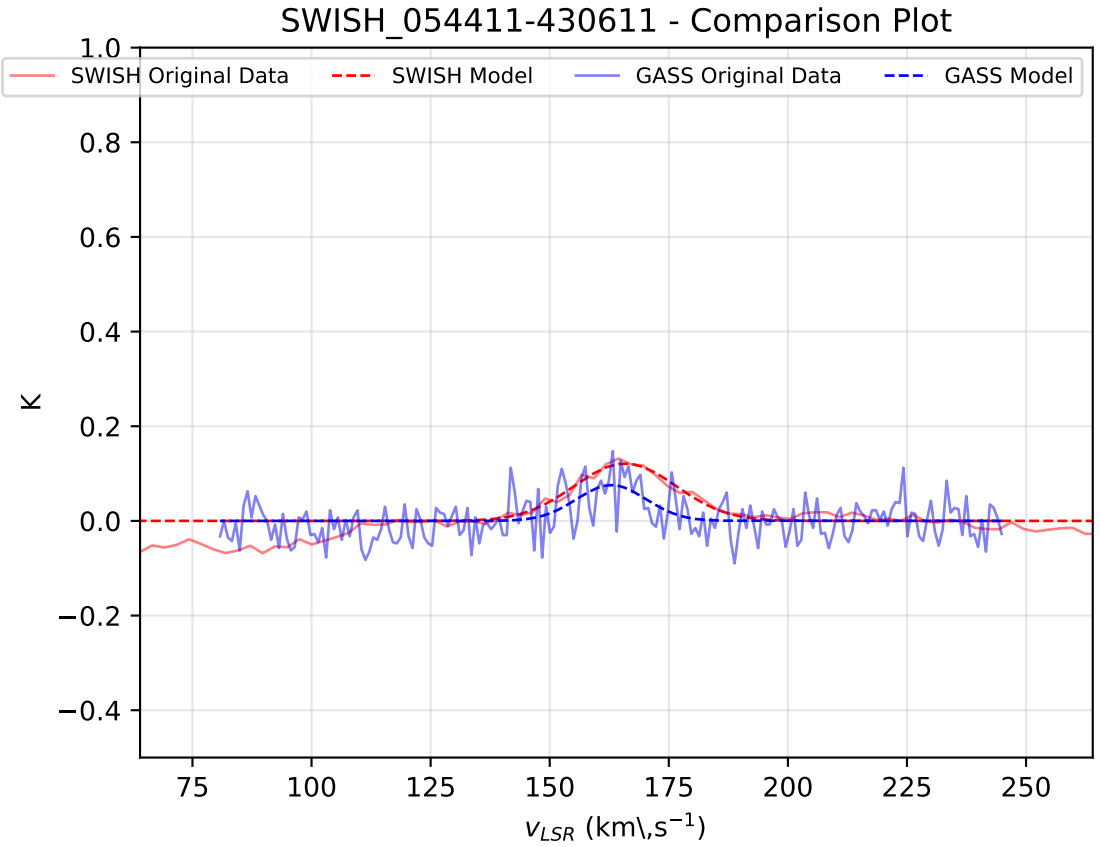

PKS_1101-325+ - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})

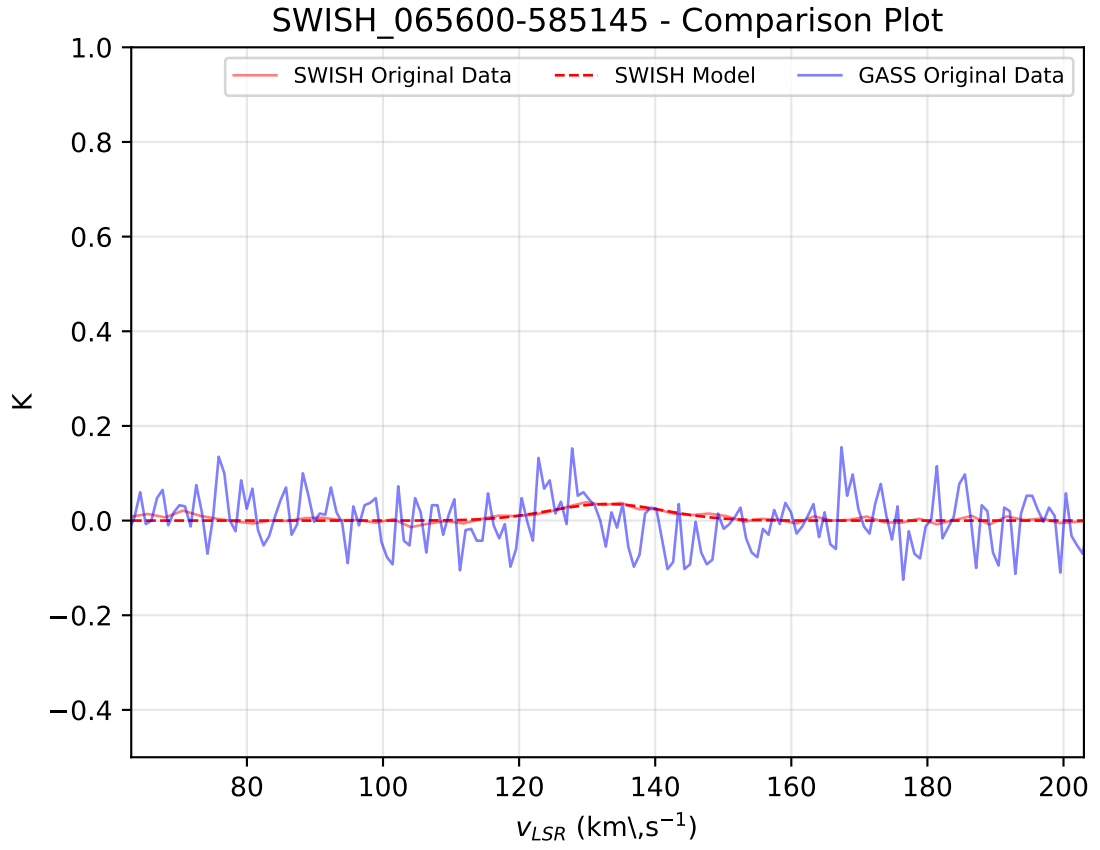

PKS_1034-29+ - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km s^{-1})

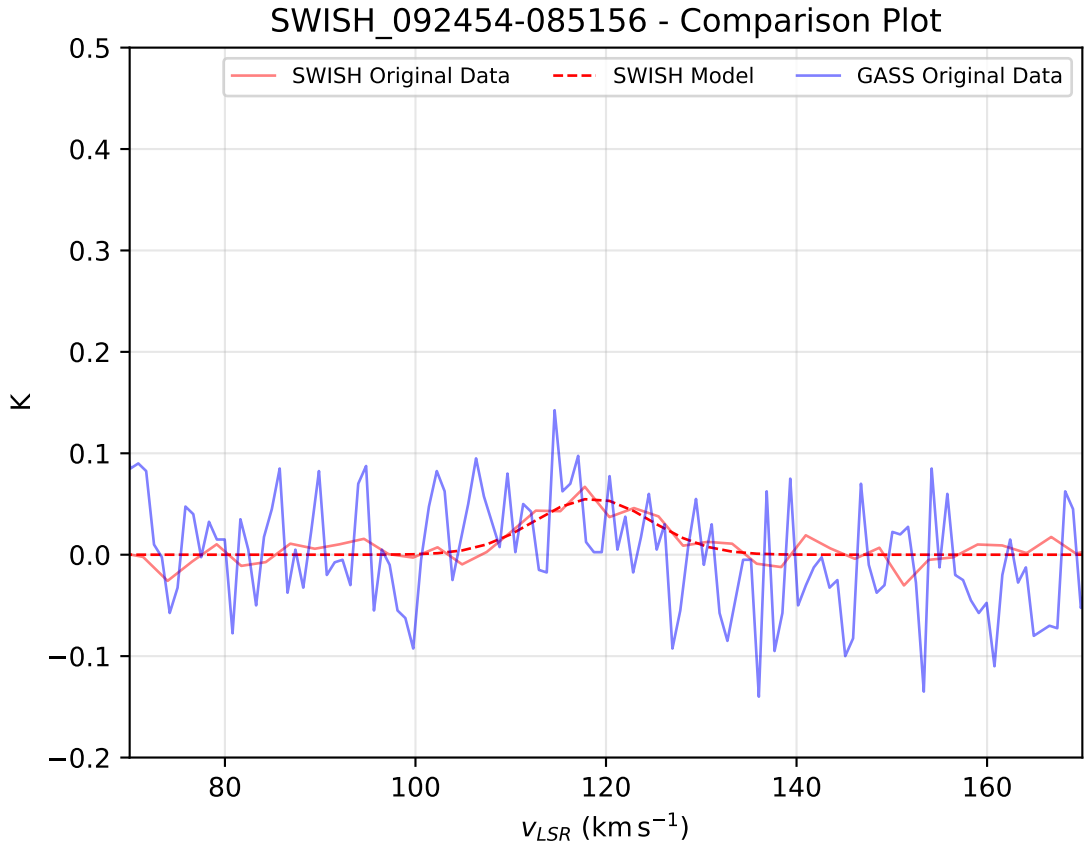

PKS_1117-248+ - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
K
v_{LSR} (km s^{-1})

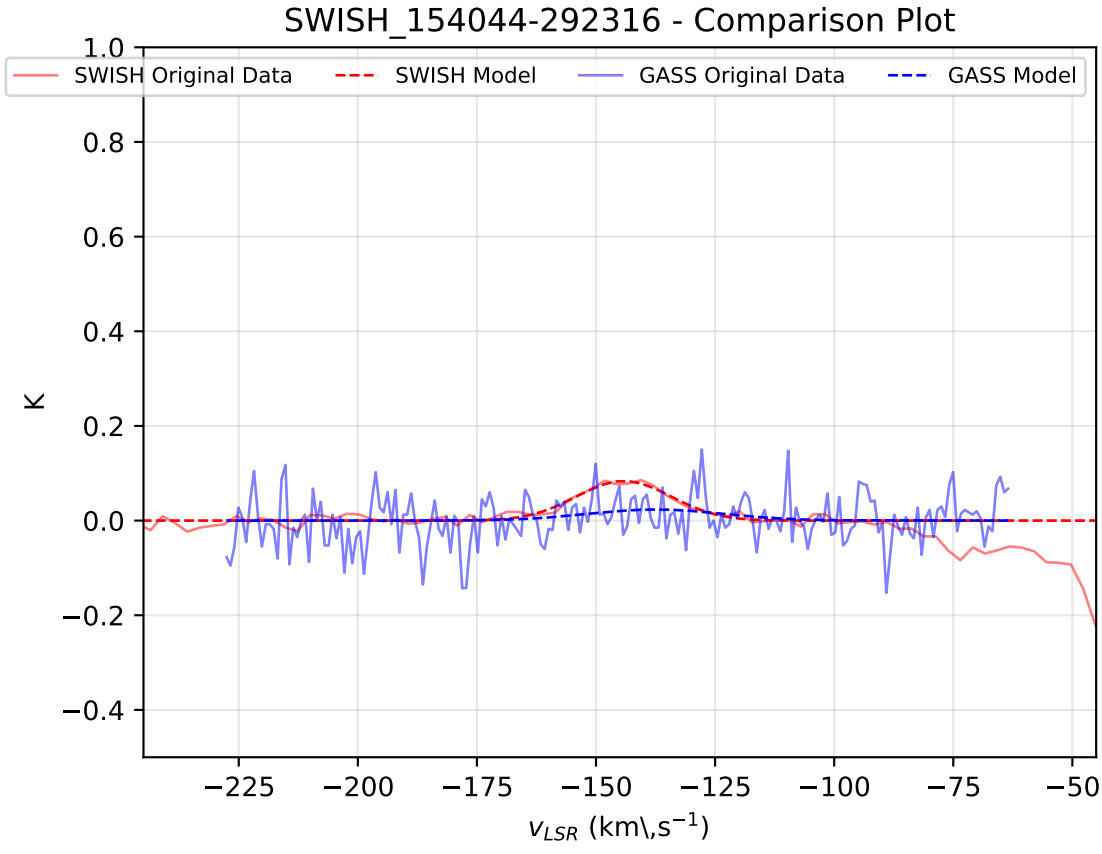

PKS_1207-399- - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km s^{-1})

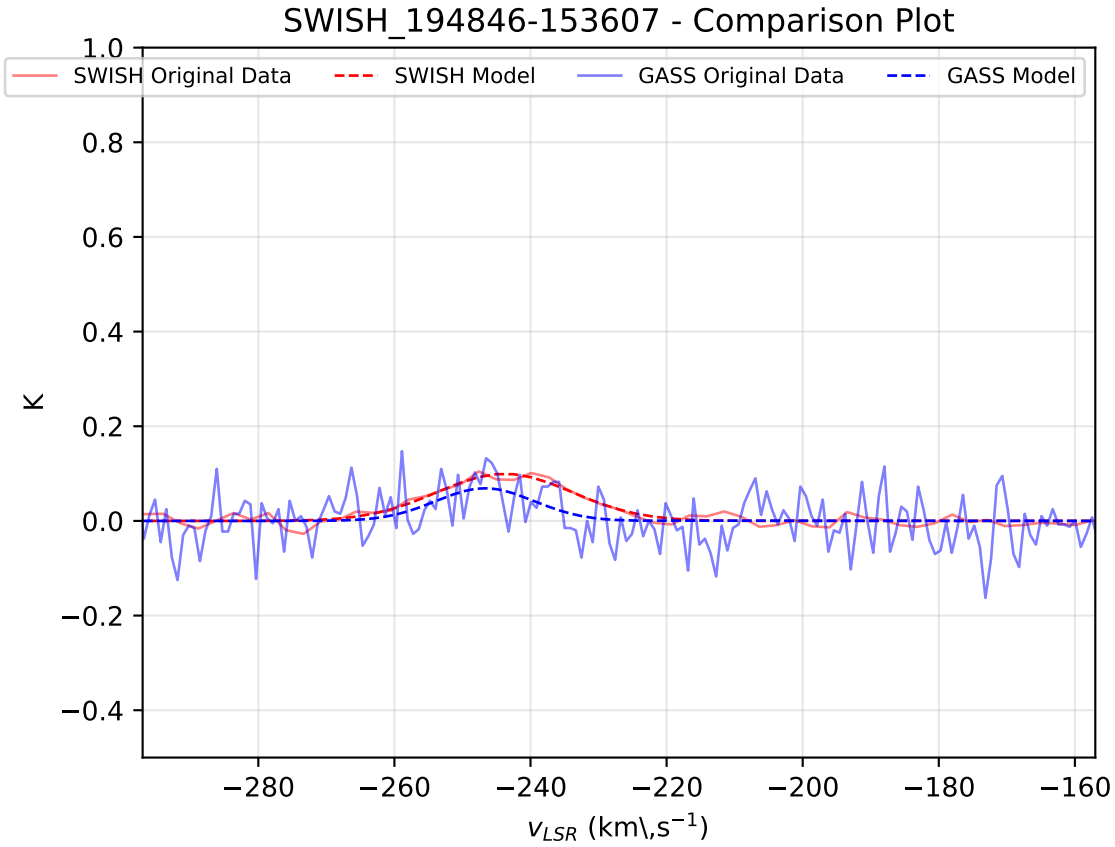

S9 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km s^{-1})

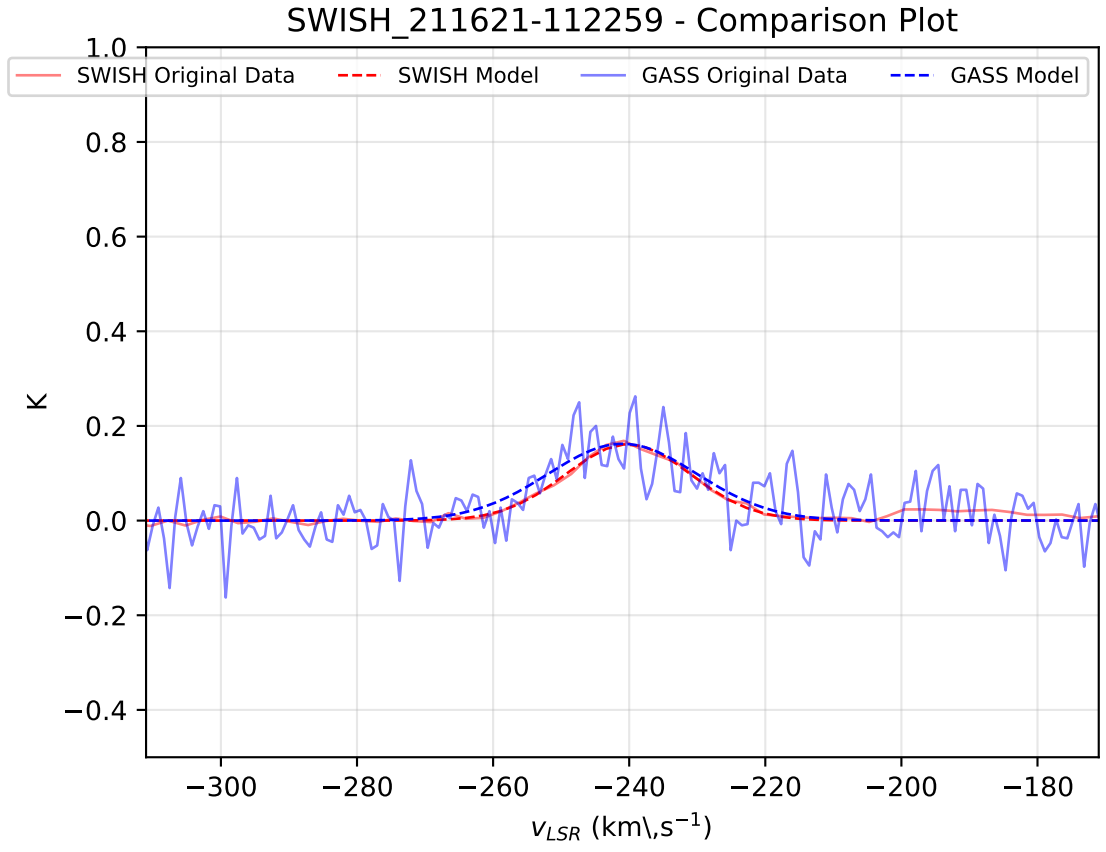

SWISH_054411-430611 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})


SWISH_065600-585145 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
K
v_{LSR} (km\,s^{-1})


SWISH_092454-085156 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
K
80
100
120
140
160
v_{LSR} (km s^{-1})


SWISH_154044-292316 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
K
−225
−200
−175
−150
−125
−100
−75
−50
v_{LSR} (km\,s^{-1})


SWISH_194846-153607 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})


SWISH_211621-112259 - Comparison Plot
SWISH Original Data
SWISH Model
GASS Original Data
GASS Model
K
v_{LSR} (km\,s^{-1})

B

Adjusted Gaussian Parameters

Source Name	SWISH Data				GASS Data			
	v_{LSR} (km s^{-1})	FWHM (km s^{-1})	$T_{b,fit}$ (K)	N_{HI} ($\times 10^{18}$ cm^{-2})	v_{LSR} (km s^{-1})	FWHM (km s^{-1})	$T_{b,fit}$ (K)	N_{HI} ($\times 10^{18}$ cm^{-2})
GASS_10.3-25.0-271_41	-256.97 ± 0.072	23.46 ± 0.17	0.261 ± 0.002	1.19 ± 0.04	-258.13 ± 0.65	24.46 ± 1.39	0.26 ± 0.01	1.23 ± 0.24
GASS_180.6-77.4-116_12	-93.16 ± 0.234	23.61 ± 0.23	0.266 ± 0.006	1.22 ± 0.08	-93.95 ± 0.66	25.87 ± 1.8	0.23 ± 0.01	1.14 ± 0.26
GASS_224.1-34.4+133_44	125.82 ± 0.056	29.52 ± 0.132	0.172 ± 0.001	0.98 ± 0.02	213.9 ± 1.17	36.65 ± 2.69	0.18 ± 0.01	1.28 ± 0.33
GASS_246.2+39.9+312_0	308.73 ± 0.099	36.93 ± 0.25	0.172 ± 0.001	1.23 ± 0.04	309.11 ± 0.58	41.02 ± 1.38	0.28 ± 0.01	2.20 ± 0.31
GASS_28.2-41.1-221_19	-207.26 ± 0.075	24.96 ± 0.18	0.227 ± 0.002	1.10 ± 0.04	-207.19 ± 0.79	23.36 ± 1.92	0.2 ± 0.01	0.91 ± 0.24
GASS_280.5-56.5+119_19	105.98 ± 0.07	25.77 ± 0.16	0.215 ± 0.001	1.07 ± 0.03	106.35 ± 0.55	22.13 ± 1.22	0.21 ± 0.01	0.90 ± 0.19
GASS_280.5-56.5+119_20	193.64 ± 0.06	24.33 ± 0.13	0.189 ± 0.001	0.89 ± 0.02	193.71 ± 0.75	24.98 ± 1.74	0.18 ± 0.01	0.89 ± 0.22
M16.11+	-153.67 ± 0.11	18.11 ± 0.28	0.223 ± 0.004	0.78 ± 0.06	-153.85 ± 0.82	15.33 ± 2.18	0.2 ± 0.02	0.60 ± 0.29
PKS_1009-321-	172.77 ± 0.09	18.63 ± 0.2	0.180 ± 0.002	0.65 ± 0.08	173.14 ± 1.28	24.46 ± 2.65	0.15 ± 0.02	5.35 ± 0.34
PKS_1009-321-	212.13 ± 0.03	25.94 ± 0.08	0.693 ± 0.002	3.49 ± 0.05	212.71 ± 0.2	26.77 ± 0.6	0.89 ± 0.01	4.62 ± 0.31
PKS_1011-282+	122.56 ± 0.01	22.92 ± 0.02	1.112 ± 0.001	4.95 ± 0.02	123.17 ± 0.13	23.3 ± 0.29	1.02 ± 0.01	4.59 ± 0.21
PKS_1020-103–	83.84 ± 0.04	19.33 ± 0.11	0.180 ± 0.001	0.68 ± 0.02	82.97 ± 0.25	23.1 ± 0.6	0.16 ± 0.003	0.70 ± 0.06
PKS_1034-29+	115.18 ± 0.1	24.1 ± 0.24	0.138 ± 0.002	0.64 ± 0.06	117.86 ± 2.68	19.9 ± 36.84	0.13 ± 0.03	0.50 ± 2.09
PKS_1034-29+	181.75 ± 0.18	27.07 ± 0.42	0.069 ± 0.001	0.36 ± 0.03	184.1 ± 18.35	11.31 ± 9.9	0.08 ± 0.1	0.18 ± 0.75
PKS_1101-325+	91.75 ± 0.07	25.38 ± 0.17	0.266 ± 0.002	1.31 ± 0.04	95.24 ± 0.98	45.34 ± 2.73	0.23 ± 0.01	2.02 ± 0.42
PKS_1117-248+	147.22 ± 0.313	20.89 ± 0.81	0.047 ± 0.002	0.19 ± 0.04	0 ± 0	0 ± 0	0 ± 0	0.00 ± 0.00
PKS_1207-399-	118.5 ± 0.12	24.32 ± 0.29	0.103 ± 0.001	0.49 ± 0.03	115.05 ± 1.18	24.42 ± 2.81	0.14 ± 0.01	0.66 ± 0.25
S9	-126.88 ± 0.04	28.11 ± 0.08	0.326 ± 0.001	1.78 ± 0.02	-133.82 ± 0.88	35.26 ± 1.91	0.32 ± 0.02	2.17 ± 0.51
SWISH_054411-430611	163.27 ± 0.18	25.17 ± 0.53	0.086 ± 0.07	0.42 ± 0.71	162.96 ± 1.47	17.25 ± 3.9	0.08 ± 0.01	0.25 ± 0.19
SWISH_065600-585145	130.39 ± 0.63	19.54 ± 5.67	0.014 ± 0.002	0.05 ± 0.09	0 ± 0	0 ± 0	0 ± 0	0.00 ± 0.00
SWISH_092454-085156	116.06 ± 0.95	14.27 ± 2.13	0.035 ± 0.007	0.10 ± 0.09	0 ± 0	0 ± 0	0 ± 0	0.00 ± 0.00
SWISH_154044-292316	-146.73 ± 1.57	23.42 ± 3.98	0.052 ± 0.01	0.24 ± 0.21	-137.35 ± 6.92	33.48 ± 8.87	0.02 ± 0.007	0.15 ± 0.16
SWISH_194846-153607	-246.38 ± 0.62	23.49 ± 1.3	0.069 ± 0.005	0.32 ± 0.10	-246.68 ± 37.6	17.1 ± 4.63	0.07 ± 0.05	0.23 ± 0.46
SWISH_211621-112259	-242.72 ± 0.04	21.62 ± 0.1	0.121 ± 0.0006	0.51 ± 0.01	-240.98 ± 0.97	25.85 ± 2.55	0.16 ± 0.01	0.81 ± 0.26

Table B.1: Parameters of single Gaussian components for 23 sightlines in the SWISH data with systematic difference corrected for.

References

E. Bajaja, C. E. Cappa de Nicolau, M. C. Martin, R. Morras, C. A. Olano, and W. G. L. Poppel. A survey of several southern high-velocity complexes. *aaps*, 78:345–362, Jun 1989.

D. G. Barnes, L. Staveley-Smith, W. J. G. de Blok, T. Oosterloo, I. M. Stewart, A. E. Wright, G. D. Banks, R. Bhathal, P. J. Boyce, M. R. Calabretta, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, P. te Lintel Hekkert, P. A. Henning, H. Jerjen, S. Juraszek, M. J. Kesteven, V. A. Kilborn, P. M. Knezek, B. Koribalski, R. C. Kraan-Korteweg, D. F. Malin, M. Marquarding, R. F. Minchin, J. R. Mould, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, A. Schröder, F. Stootman, R. L. Webster, W. E. Wilson, and T. Ye. The HI Parkes All Sky Survey: southern observations, calibration and robust imaging. *mnras*, 322(3):486–498, Apr 2001. doi: 10.1046/j.1365-8711.2001.04102.x.

N. B. Bekhti, P. Richter, B. Winkel, J. Kerp, P. M. W. Kalberla, U. Klein, and M. T. Murphy. Absorption and Emission Line Studies of Gas in the Milky Way Halo. In R. Kothes, T. L. Landecker, and A. G. Willis, editors, *The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey*, volume 438 of *Astronomical Society of the Pacific Conference Series*, page 170, Dec 2010.

F. Boulanger, P. Cox, and A. P. Jones. Course 7: Dust in the Interstellar Medium. In F. Casoli, J. Lequeux, and F. David, editors, *Infrared Space Astronomy, Today and Tomorrow*, volume 70, page 251, Jan 2000.

D. V. Bowen and J. C. Blades. Interstellar MG II and C IV Absorption toward Markarian 205

by NGC 4319: an “Optically Thick” QSO Absorption System. *apjl*, 403:L55, Feb 1993. doi: 10.1086/186720.

J. A. Collins, J. M. Shull, and M. L. Giroux. Hubble Space Telescope Survey of Interstellar High-velocity Si III. *apj*, 705(1):962–977, Nov 2009. doi: 10.1088/0004-637X/705/1/962.

E. M. de Gouveia Dal Pino, C. Melioli, A. D'Ercole, F. Brighenti, and A. C. Raga. Supernova Explosions and the Triggering of Galactic Fountains and Outflows. In *Revista Mexicana de Astronomia y Astrofisica Conference Series*, volume 36 of *Revista Mexicana de Astronomia y Astrofisica Conference Series*, pages 17–24, Aug 2009.

V. de Heij, R. Braun, and W. B. Burton. High-resolution imaging of compact high-velocity clouds. *aap*, 391:67–81, Aug 2002. doi: 10.1051/0004-6361:20020640.

H. A. Ford, N. M. McClure-Griffiths, F. J. Lockman, J. Bailin, M. R. Calabretta, P. M. W. Kalberla, T. Murphy, and D. J. Pisano. H I Clouds in the Lower Halo. I. The Galactic All-Sky Survey Pilot Region. *apj*, 688(1):290–305, Nov 2008. doi: 10.1086/592188.

H. A. Ford, F. J. Lockman, and N. M. McClure-Griffiths. The Distribution of Disk-Halo H I Clouds in the Inner Milky Way. 438:163, Dec 2010.

D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *pasp*, 125:306, Mar. 2013. doi: 10.1086/670067.

A. J. Fox, B. D. Savage, and B. P. Wakker. A Survey of O VI, C III, and H I in Highly Ionized High-Velocity Clouds. *apjs*, 165(1):229–255, Jul 2006. doi: 10.1086/504800.

K. Freeman and J. Bland-Hawthorn. The New Galaxy: Signatures of Its Formation. *araa*, 40:487–537, Jan 2002. doi: 10.1146/annurev.astro.40.060401.093840.

D. W. Hogg and D. Foreman-Mackey. Data Analysis Recipes: Using Markov Chain Monte Carlo. *apjs*, 236(1):11, May 2018. doi: 10.3847/1538-4365/aab76e.

A. N. M. Hulsbosch and B. P. Wakker. A deep, nearly complete, survey of northern high-velocity clouds. *aaps*, 75:191–236, Oct. 1988.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001. URL http://www.scipy.org/. [Online; accessed <today>].

P. M. W. Kalberla and U. Haud. GASS: The Parkes Galactic All-Sky Survey. Update: improved correction for instrumental effects and new data release. *aap*, 578:A78, Jun 2015. doi: 10.1051/0004-6361/201525859.

P. M. W. Kalberla, N. M. McClure-Griffiths, D. J. Pisano, M. R. Calabretta, H. A. Ford, F. J. Lockman, L. Staveley-Smith, J. Kerp, B. Winkel, T. Murphy, and K. Newton-McGee. GASS: the Parkes Galactic all-sky survey. II. Stray-radiation correction and second data release. *aap*, 521:A17, Oct 2010. doi: 10.1051/0004-6361/200913979.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995. doi: 10.1080/01621459.1995.10476572.

A. Liu and J. R. Shaw. Data Analysis for Precision 21 cm Cosmology. *arXiv e-prints*, art. arXiv:1907.08211, Jul 2019.

F. J. Lockman. The HI halo in the inner galaxy. *apj*, 283:90–97, Aug 1984. doi: 10.1086/162277.

F. J. Lockman, E. M. Murphy, S. Petty-Powell, and V. J. Urick. A very sensitive 21 centimeter survey for galactic high-velocity hi. *The Astrophysical Journal Supplement Series*, 140(2): 331–365, jun 2002. doi: 10.1086/339371. URL https://doi.org/10.1086%2F339371.

N. M. McClure-Griffiths, A. Ford, D. J. Pisano, B. K. Gibson, L. Staveley-Smith, M. R. Calabretta, L. Dedes, and P. M. W. Kalberla. Evidence for Chimney Breakout in the Galactic Supershell GSH 242-03+37. *apj*, 638:196–205, Feb. 2006. doi: 10.1086/498706.

N. M. McClure-Griffiths, D. J. Pisano, M. R. Calabretta, H. A. Ford, F. J. Lockman, L. Staveley-Smith, P. M. W. Kalberla, J. Bailin, L. Dedes, S. Janowiecki, B. K. Gibson, T. Murphy, H. Nakanishi, and K. Newton-McGee. Gass: The Parkes Galactic All-Sky Survey. I. Survey Description, Goals, and Initial Data Release. *apjs*, 181:398–412, Apr. 2009. doi: 10.1088/0067-0049/181/2/398.

N. M. McClure-Griffiths, J. A. Green, A. S. Hill, F. J. Lockman, J. M. Dickey, B. M. Gaensler, and A. J. Green. Atomic Hydrogen in a Galactic Center Outflow. *apjl*, 770(1):L4, Jun 2013. doi: 10.1088/2041-8205/770/1/L4.

V. A. Moss and N. M. McClure-Griffiths. The hidden iceberg structure of the halo with Parkes. 2017.

V. A. Moss, N. M. McClure-Griffiths, T. Murphy, D. J. Pisano, J. K. Kummerfeld, and J. R. Curran. VizieR Online Data Catalog: High-velocity clouds from GASS. I. (Moss+, 2013). *VizieR Online Data Catalog*, 220, Nov. 2013.

V. A. Moss, F. J. Lockman, and N. M. McClure-Griffiths. Tracing Dense and Diffuse Neutral Hydrogen in the Halo of the Milky Way. *apj*, 834(2):155, Jan 2017. doi: 10.3847/1538-4357/834/2/155.

C. A. Muller, J. H. Oort, and E. Raimond. Hydrogène neutre dans la couronne galactique? *Academie des Sciences Paris Comptes Rendus*, 257:1661–1662, 1963.

E. M. Murphy, F. J. Lockman, and B. D. Savage. A Sensitive Search for Galactic High-Velocity H i Clouds. *apj*, 447:642, Jul 1995. doi: 10.1086/175906.

J. H. Oort. Possible Interpretations of the High-Velocity Clouds. *bain*, 18:421–438, 1966.

M. S. Owers, M. J. Hudson, K. A. Oman, J. Bland -Hawthorn, S. Brough, J. J. Bryant, L. Cortese, W. J. Couch, S. M. Croom, J. van de Sande, C. Federrath, B. Groves, A. M. Hopkins, J. S. Lawrence, N. P. F. Lorente, R. M. McDermid, A. M. Medling, S. N. Richards, N. Scott, D. S. Taranu, C. Welker, and S. K. Yi. The SAMI Galaxy Survey: Quenching of Star Formation in Clusters I. Transition Galaxies. *apj*, 873(1):52, Mar 2019. doi: 10.3847/1538-4357/ab0201.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: Machine Learning in Python. *arXiv e-prints*, art. arXiv:1201.0490, Jan 2012.

A. Petzler. Probing the galactic ism in oh absorption, Feb 2018.

D. Price. HIPSR reduction. *GitHub repository*, 2012.

M. E. Putman, V. de Heij, L. Staveley-Smith, R. Braun, K. C. Freeman, B. K. Gibson, W. B. Burton, D. G. Barnes, G. D. Banks, R. Bhathal, W. J. G. de Blok, P. J. Boyce, M. J. Disney, M. J. Drinkwater, R. D. Ekers, P. A. Henning, H. Jerjen, V. A. Kilborn, P. M. Knezek, B. Koribalski, D. F. Malin, M. Marquarding, R. F. Minchin, J. R. Mould, T. Oosterloo, R. M. Price, S. D. Ryder, E. M. Sadler, I. Stewart, F. Stootman, R. L. Webster, and

A. E. Wright. HIPASS High-Velocity Clouds: Properties of the Compact and Extended Populations. *aj*, 123:873–891, Feb. 2002. doi: 10.1086/338088.

M. E. Putman, J. E. G. Peek, and M. R. Joung. Gaseous Galaxy Halos. *araa*, 50:491–529, Sep 2012. doi: 10.1146/annurev-astro-081811-125612.

T. P. Robitaille and B. A. Whitney. The Galactic star formation rate as seen by the Spitzer Space Telescope. *Highlights of Astronomy*, 15:799–799, Nov 2010. doi: 10.1017/S1743921310011774.

B. D. Savage, L. Lu, J. N. Bahcall, J. Bergeron, A. Boksenberg, G. F. Hartig, B. T. Jannuzi, S. Kirhakos, F. J. Lockman, W. L. W. Sargent, D. P. Schneider, D. Turnshek, R. J. Weymann, and A. M. Wolfe. The Hubble Space Telescope Quasar Absorption Line Key Project. III. First Observational Results on Milky Way Gas. *apj*, 413:116, Aug 1993. doi: 10.1086/172982.

K. C. Schlaufman, I. B. Thompson, and A. R. Casey. An Ultra Metal-poor Star Near the Hydrogen-burning Limit. *apj*, 867(2):98, Nov 2018. doi: 10.3847/1538-4357/aadd97.

L. Staveley-Smith, W. E. Wilson, T. S. Bird, M. J. Disney, R. D. Ekers, K. C. Freeman, R. F. Haynes, M. W. Sinclair, R. A. Vaile, R. L. Webster, and A. E. Wright. The Parkes 21 CM multibeam receiver. *pasa*, 13(3):243–248, Nov 1996.

H. van Woerden and B. P. Wakker. Distances and Metallicities of HVCS. In H. van Woerden, B. P. Wakker, U. J. Schwarz, and K. S. de Boer, editors, *High Velocity Clouds*, volume 312 of *Astrophysics and Space Science Library*, page 195, Jan 2004. doi: 10.1007/1-4020-2579-3_10.

M. G. R. Vogelaar, B. P. Wakker, and U. J. Schwarj. Measuring the Fractal Structure of Interstellar Clouds. In E. Falgarone, F. Boulanger, and G. Duvert, editors, *Fragmentation of Molecular Clouds and Star Formation*, volume 147 of *IAU Symposium*, page 508, Jan 1991.

B. P. Wakker and J. S. Mathis. Dependence of Gas-Phase Abundances in the Interstellar Medium on Column Density. *apjl*, 544(2):L107–L110, Dec 2000. doi: 10.1086/317316.

B. P. Wakker and H. van Woerden. Distribution and origin of high-velocity clouds. III - Clouds, complexes and populations. *aap*, 250:509–532, Oct. 1991.

T. Westmeier, B. S. Koribalski, and R. Braun. Gas and dark matter in the Sculptor group: NGC 55. *mnras*, 434(4):3511–3525, Oct 2013. doi: 10.1093/mnras/stt1271.