Automated Processing and Analysis of Galactic Neutral Hydrogen Absorption Observations

James Michael Dempsey

A thesis submitted for the degree of
Bachelor of Science with Honours in Astronomy & Astrophysics of
The Australian National University

26 October 2017

Declaration

This thesis is an account of research undertaken between February 2016 and October 2017 at The Research School of Astronomy and Astrophysics, ANU College of Physical and Mathematical Science, The Australian National University, Canberra, Australia.

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.

James M Dempsey
October, 2017

Acknowledgements

Firstly I wish to thank my advisor, Prof Naomi McClure-Griffiths for her guidance and encouragement throughout the project. I have greatly appreciated the way she has always found time in her busy schedule to provide support and advice, and I admire her deep knowledge of the Milky Way.

Thanks also to Dr Jimi Green for use of his data, and assistance with observing information.

Finally, thank you to my wife Leanne Dempsey for endless patience, support and editing.

Contents

List of Figures

Abstract

In this research project I analysed the composition and structure of the cold neutral hydrogen (H I) gas at the locations of the MAGMO (Mapping the Galactic Magnetic field through OH masers) survey observations. To enable this analysis I developed an automated pipeline and used that to produce H I absorption and emission spectra from the MAGMO observations. This pipeline was designed with the aim of being applicable to future large surveys such as GASKAP, the Galactic Australian SKA Pathfinder survey.

The MAGMO observations targeted the sites of 6.7-GHz methanol masers, which are associated with star-forming regions. These very active regions have complex structures with bubbles formed by stellar winds from young, high mass stars. In these regions it is reasonable to expect there will be significant reserves of cold gas. The MAGMO H I dataset provided an excellent opportunity to examine the cold neutral medium (CNM) towards these regions and to compare it with the overall CNM population.

In Chapter 2 I provide a summary of the current knowledge of the Milky Way's cold H I gas and an introduction to other topics which will assist the reader in understanding the analysis. Chapter 3 provides a detailed description of the data, and the pipeline developed for its analysis. This is more detailed than is typical in order to explain the decisions in developing the automated processing. In Chapter 4 I provide the results from the observations. In Chapter 5 I compare the results to results from similar studies, discuss the overall findings and assess the effectiveness of the processing pipeline. Finally in Chapter 6 I summarise the main findings and discuss future potential for both this data set and the processing pipeline.

Chapter 2

Introduction

2.1 Neutral hydrogen in the Milky Way

Hydrogen, the most abundant element in the universe, fuels stars and is the majority component of the interstellar medium. As a result, understanding its distribution in the galaxy is critical in order to understand the dynamics and history of the galaxy. Observations of neutral hydrogen (HI) have been crucial in mapping the spiral arms of the galaxy (Levine et al. 2006) and, most recently, the mapping of a new outer spiral arm (McClure-Griffiths et al. 2004) in the fourth quadrant of the Milky Way. Such observations have also been used extensively to map the rotation curve of the Galaxy (e.g. Gunn et al. 1979; McClure-Griffiths & Dickey 2016) and to measure kinematic distances of objects within the disc (Jones & Dickey 2012).

Within the Milky Way, H I is located in an exponential disc with a scale height of 3.75 kpc extending out to a Galactic radius R = 35 kpc (Kalberla & Dedes 2008). The disc is warped, with a sine-wave-like vertical shape from R $\geq$ 15 kpc. From $40 \leq R \leq 60$ kpc a more diffuse and turbulent disc with a scale height of 7.5 kpc is apparent (Kalberla & Dedes 2008). This compares with the thin stellar disc with a scale height of $300 - 400$ pc extending out to $\sim$ 15 kpc and a thick stellar disc with a scale height of $\sim$ 1 kpc (Sparke & Gallagher III 2007). Surrounding the centre of the Galaxy there is a 3 kpc radius hole in the H I (Lockman & McClure-Griffiths 2016; Sofue 2017). At a large scale, the H I gas rotates with the galaxy, with gas significantly above and below the disc lagging behind the rotation (Kalberla & Kerp 2009).

Looking closer, there is a wealth of small-scale structure, with abundant shells of H I ejected from supernovae and stellar winds, some broken into fragments that appear as sheets or filaments. Even outside of these expanding bubbles the H I gas is generally accelerated to typical speeds of 10 km/s (Mach 3) (Heiles & Troland 2003b), indicating that the interstellar medium (ISM) is largely not in equilibrium.

H I gas in the Milky Way is split into two stable phases: the Cold Neutral Medium (CNM) and the Warm Neutral Medium (WNM) (Wolfire et al. 2003). These are two long lasting phases of temperature and pressure equilibrium reflecting the heating and cooling balance in the ISM. The CNM is predominantly dense clouds of H I gas at temperatures between 40 and 160K. The WNM is generally deemed to be H I gas between 4100K to 8800K and is much more diffuse (Murray et al. 2015). Significant gas also occurs in the unstable temperature range between these, however the amount is still contentious, with observations ranging from 48% (Heiles & Troland 2003a) to $<$ 30% (Begum et al. 2010) of WNM and simulations by Kim et al. (2014) predicting 20% to 50%. Table 2.1 shows the commonly understood values for the five stable phases of hydrogen in the Milky Way ISM.

Structurally, the CNM consists of small dense clumps and filaments within the diffuse WNM (Kulkarni & Heiles 1988). These clouds are usually detected in H I absorption and have typical diameters of around 3 parsecs (Kalberla & Kerp 2009). They are dynamic features, forming as cores due to pressure fluctuations in the WNM and then combining into filaments or sheets and dissipating over the course of 10^5 years (Kalberla & Kerp 2009). They are shaped both by turbulence and by

magnetic fields. The CNM is found predominantly in the disc of the Galaxy, with a scale height of 150 pc locally, due to the high pressure under which it forms. It makes up 20^{+7}_{-10}% of Galactic H I gas and has column density in the range $(4.4 \pm 0.5) \times 10^{19}$ cm$^{-2} \leq$ $N_H \leq 1 \times 10^{22}$ cm^{-2} (Dickey & Lockman 1990), although Murray et al. (2015) have detected the CNM gas as diffuse as $N_H \geq (3\pm1)\times10^{16}$ cm^{-2}. Very high sensitivity is needed to detect the lower column density population of CNM.

Phase Name	**Phase**	**State of H**	**Temperature (K)**	**Density (cm^{-3})**	**Filling Factor**	**m$_f$ Mass Fraction**
Hot Ionised Medium	HIM	H II	10^6	10^{-3}	25-65%	trace
Warm Ionised Medium	WIM	H II	8,000 - 10,000	0.3	25%	15%
Warm Neutral Medium	WNM	H I	5,000 - 8,000	0.4	35%	35%
Cold Neutral Medium	CNM	H I	20 - 100	1-50	3%	10%
Molecular Medium	MM	H_2	10	10^{2-6}	1%	40%

Table 2.1: Typical Milky Way ISM temperatures and densities for the different phases of hydrogen. The last two columns are the fraction of total ISM hydrogen in each phase by volume and by mass. (Sparke & Gallagher III 2007; Wolfire et al. 2003)

2.2 Radio observations of the CNM

The primary method of observing H I is radio observation of the 21cm H I spectral line. This hyperfine line is caused by the quantum spin-flip transition of the electron in the neutral hydrogen atom and has a rest frequency of 1,420.40575 MHz (Kulkarni & Heiles 1988). With a transition probability $A_{21} = 2.87 \times 10^{-15}$ sec^{-1} it will be around 12×10^6 years before an excited H I atom will emit a photon (Kulkarni & Heiles 1988); it does not saturate observations and thus is an excellent means of determining the column density of H I along a particular line of sight.

There are two broad classes of radio telescopes: single dishes and interferometer arrays. Single dishes generally have a large field of view and a course resolution (e.g. the Parkes 64-metre radio telescope has a 14 arcmin FWHM beam at 1420 MHz) while interferometers have a much finer resolution but often have less total collecting area (e.g. the Australia Telescope Compact Array can achieve an 8 arcsec FWHM synthesised beam at 1420 MHz). According to the Rayleigh criterion, the beam width of a telescope is given by the formula:

$$\theta = 1.22\frac{\lambda}{d} \text{ radians}, \tag{2.1}$$

where λ is the wavelength and d is the baseline length (distance between two antenna) or dish radius. This also gives the inverse relationship between the size of structures that the telescopes can detect. Shorter baselines and smaller dishes detect the large-scale structure, while longer baselines give finer resolution, allowing detection of the small structures but at the expense of missing the larger structures. As a result the two telescope types see different scales of structure in the ISM. Single dishes are well suited to seeing the larger structures and observing the 21 cm emission, which is generally produced from extended sources. Interferometers see only the small structures, and due to their finer resolution they are much better at detecting absorption between a source and the telescope. Their lack of zero spacing (extremely short baselines) means they generally do not detect emission from extended sources. More recent surveys of the Milky Way H I have thus often combined the data from both large dishes and long baseline interferometers.

The dense clouds of the CNM absorb the radiation of background sources at the H I spectral line. We observe this as a dip, or absorption line, in the continuum spectrum from the background source.

The dip is Doppler-shifted according to the relative velocity of the cloud to the observer, or in our case to the average movement of the stars in the Solar vicinity around the Galactic centre, measured as the Local Standard of Rest (LSR). While this velocity is dominated by the Galactic rotation of the target cloud, the line may be broadened by the motion of the gas within the cloud, or the line-of-sight width of the cloud. A 5 km/s velocity range equates to a distance of 1 kpc for most lines of sight (Burton 1988).

Figure 2.1 shows the rotation curve of the Galaxy. The Galaxy is commonly divided into four quadrants numbered anticlockwise from 1 to 4 starting from longitude 0 (i.e. the first quadrant is the bottom right of Fig. 2.1). In the fourth quadrant of the Galaxy, covering the Galactic longitude range $270 \leq l < 360$, gas inside the solar circle will be moving at negative velocities relative to the LSR. For each line of sight (dotted straight lines in Fig. 2.1) there will be a terminal velocity which is the relative velocity of the innermost Galactic radius that it touches. With a uniform rotation curve it is to be expected that each negative velocity greater than this terminal velocity will be crossed twice, preventing a direct conversion of these velocities into a distance measurement. A similar situation exists for positive velocities in the first quadrant ($0 < l < 90$).

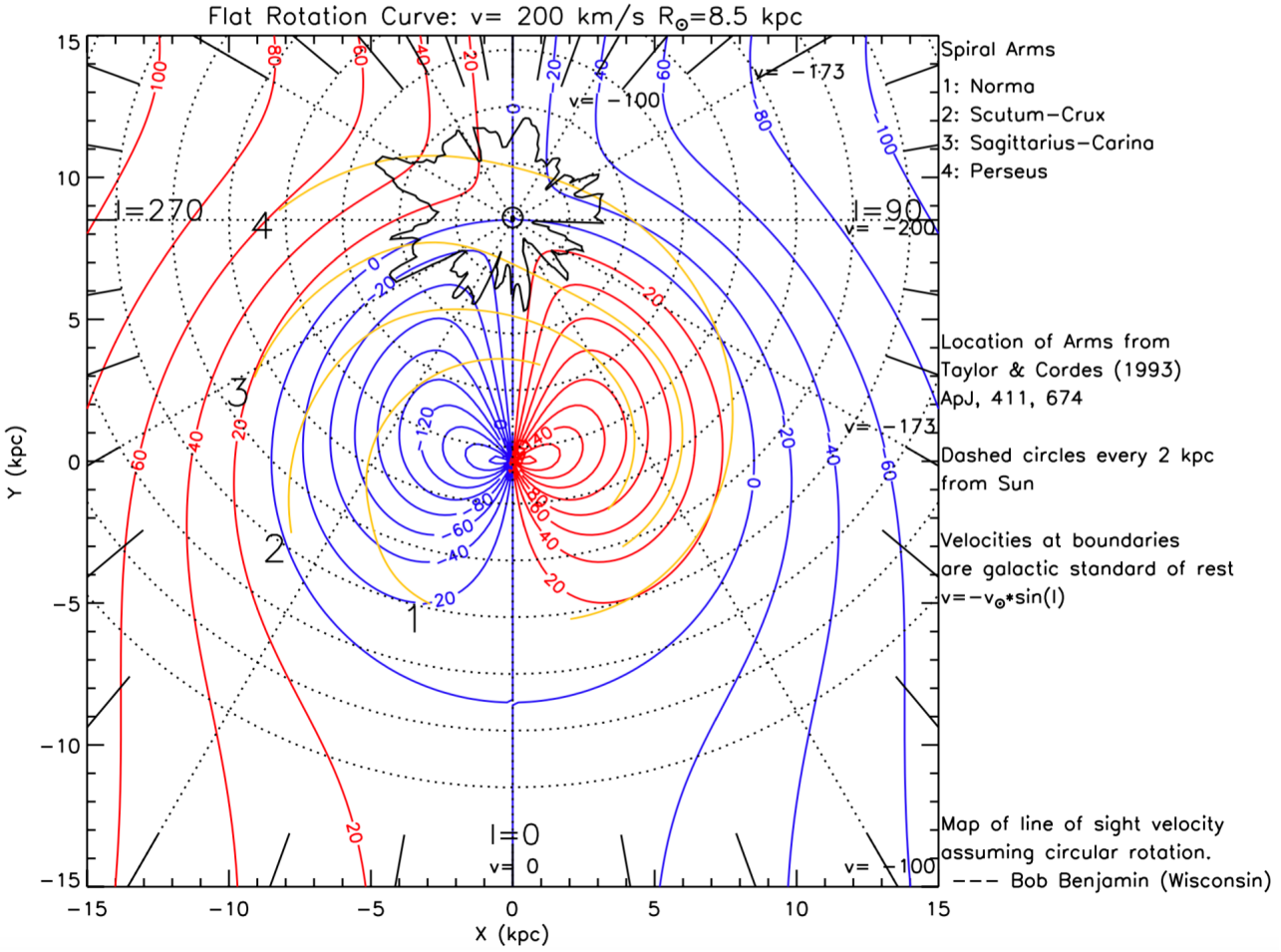

Figure 2.1: Milky Way rotation curve from Benjamin (2017). Positive velocities are shown as red contour lines and negative velocities are shown in blue. The spiral arms are shown in yellow and the dotted straight lines mark Galactic longitude lines.

In some cases this distance ambiguity can be resolved. In most observations the background sources will be a mixture of extragalactic and galactic sources. Where the sources are within the galaxy they will not show any absorption past their own location, exhibiting a flat amplitude at their continuum

emission level. For inner galaxy sources this is complicated due to the double crossing of velocity values, so a background source in the inner galaxy in the fourth quadrant could have absorption at velocities more negative than its own LSR velocity if it were located past the terminal velocity point. Where inner galaxy background sources have had their distances determined by other means, such as parallactic distance measurement using Very Long Baseline Interferometry (VLBI), a unique kinematic distance may be determined as some of the far velocities can be excluded.

Near the centre of the Galaxy, the velocities are confused both by velocity crowding and by the non-circular motion of the bar. Within 15° of the Galactic centre, the inner galaxy motion is largely perpendicular to the line of sight, so the gas is crowded around a velocity of 0 km/s. Similar velocity crowding occurs near the anti-centre. In the Galactic centre ($r < 3$ kpc), the bulge and bar structure at the centre of the galaxy exhibits solid body rotation resulting in a steep terminal velocity line reaching -250 km/s.

When observing absorption of the H I line we measure the brightness temperature against a background continuum source. To convert this temperature into an optical depth τ we must remove as much of the H I emission as possible and then calculate the ratio of the remaining temperature to the background continuum source's temperature. This is generally achieved using the following equation,

$$e^{-\tau(v)} = \frac{T_{on} - T_{off}}{T_{bg}}, \tag{2.2}$$

where T_{on} is the brightness temperature when directly observing the source, T_{off} is the temperature off the source, away from the background source and T_{bg} is the brightness temperature of the continuum background source. The continuum temperature is normally measured at a velocity range where there is no H I gas present.

The emission can also be excluded by filtering out the large-scale structure (where most emission comes from), either by Fourier transforming the image and zeroing out the short spacing data, or by observing using long baselines. In these cases there is no need to subtract the T_{off} as it is no longer present in the measured T_{on} (Dickey et al. 2003).

2.3 Spin Temperature

To calculate the spin, or excitation, temperature T_S we need to have emission measurements as well as the optical depth. The emission is excluded in the longer baselines so it needs to be sourced from shorter baseline or single dish observations. Spin temperature can then be calculated using the following equation.

$$T_S = \frac{T_{off}}{1 - e^{-\tau(v)}} \tag{2.3}$$

This gives a density weighted harmonic mean of the temperatures of the clouds along the line of sight at a given velocity. To separate these out we need to account for each component along the line of sight, including the diffuse continuum emission, the cold absorbing gas and warm gas both in front of and behind the cold gas. This leads to the radiative transfer equation of (Dickey et al. 2003, eq. 10)

$$T_{obs}(v) = T_{w,f} + T_{w,b}e^{-\tau(v)} + T_{cool}(1 - e^{-\tau(v)}) + T_{cont}e^{-\tau(v)} \ , \tag{2.4}$$

where $T_{w,f}$ and $T_{w,b}$ are the apparent temperatures of the foreground and background warm H I, T_{cool} and $\tau(v)$ are the temperature and optical depth of the cold H I, and T_{cont} is the temperature of the diffuse background continuum. Note that the apparent warm gas temperature is proportional to its column density rather than its kinetic temperature as the WNM is optically thin (Kulkarni & Heiles 1988). The background diffuse continuum emission and warm gas emission are both absorbed by the cool gas.

There are many methods in the literature for splitting out these components to produce a spin temperature for the cold gas. In general these seek to identify individual components of the absorption spectrum, match those to components in the emission spectrum and then use the emission component to estimate the brightness temperature of the absorption component and then use Eq. (2.3) to estimate the spin temperature.

Many methods start with fitting a sum of Gaussian components to the absorption spectrum, with the intent of reproducing the physical clouds, each of which may be represented by a component (Roy et al. 2013). This hypothesis is supported by the observation that at high Galactic latitudes, non-blended absorption spectra can be quite accurately reproduced as a sum of Gaussians. However in environments such as the Galactic plane where there are overlapping gas clouds, the Gaussian decomposition is not unique and thus multiple solutions are possible (Dickey et al. 2003). Nevertheless this remains the most physically motivated approach to modelling the ISM probed by an absorption spectrum (Roy et al. 2013). An example of such a decomposition is shown in Fig. 2.2.

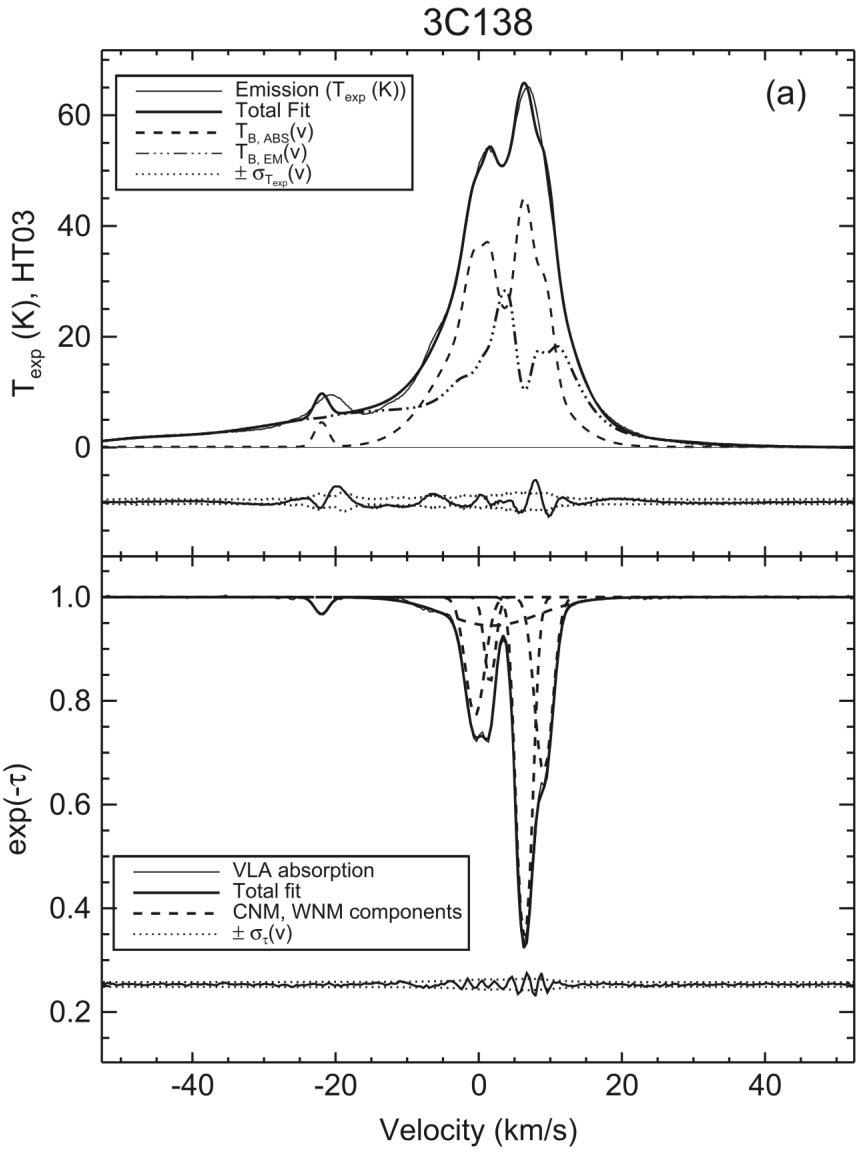

Figure 2.2: Example emission and absorption spectrum from Murray et al. (2015). This shows the decomposition of both spectra into Gaussian components.

In the two component (warm and cold) approach of Heiles & Troland (2003a), the absorption

spectrum is fitted using a sum of Gaussians using a least-squares fitting algorithm. Then the expected emission spectrum is fitted with a combination of the predetermined components for the CNM plus Gaussian components for the WNM. Here the fitting must find both the new Gaussians from the emission but also the spin temperatures from the cold components. Where components overlap, all possible combinations of order are tried to see which one best fits the emission spectrum. The fraction of the warm gas in front of or behind the CNM components are also trialled to find the lowest residuals. Generally the fraction of WNM in front of the CNM does not lead to significantly significant differences in the fit, but is instead used to quantify the errors in the spin temperatures. This approach is also used by Murray et al. (2015) and Stanimirović et al. (2014).

Dickey et al. (2003) describe a two phase linear least-squares fit to the observed emission T_B and absorption $e^{-\tau}$ profiles rather than decomposing the spectra into Gaussians. Here they first subtract the background continuum emission from the observed emission spectrum to get a brightness temperature T_B. They then model the emission as a linear function across each individual absorption component and fit the linear emission and the cool component temperature $T_{cool} - T_{cont}$ to the observed brightness temperature and absorption profile. While different fractions of warm gas in front of the cool gas component are trialled, they generally use the results for a fraction of 0.5 except where it gives an unphysically low spin temperature when they use a fraction of 1, meaning all warm gas is behind the cool cloud. With this approach they avoid the non-uniqueness of Gaussian decomposition for blended lines. A similar approach is used by Strasser et al. (2007) except that they fit a quadratic rather than linear function for the emission.

The final method that Dickey et al. (2003) describe is a two-phase nonlinear least-squares fitting method that aims to provide a continuous solution for the emission either side of the absorption component. In this case they fix the emission at the edges of the absorption component, modelling a linear function between these values, but fitting for the velocity of the component edges for blended absorption components.

Comparisons between iterations of these techniques are made in Heiles & Troland (2003a), Dickey et al. (2003) and Murray et al. (2015). Murray et al. (2015) find that, for their medium and high Galactic latitude sample, the spin temperatures from the Gaussian fitting are closest to the $40 \leq T_S \leq 200K$ as modelled by Wolfire et al. (2003) for the CNM.

In this study we have opted to implement a limited version of the Gaussian decomposition approach. We decompose the absorption spectra into Gaussian components. However rather than doing the complex full decomposition of the emission spectra we instead simply measure the brightness temperature of the cool component as the temperature of the emission spectra at the central velocity of the cool component. This will mean that the calculated spin temperatures will be upper limits rather than actual spin temperatures as some of the emission temperature is likely to have come from warm components rather than just the cool component.

Finally, with a spin temperature and an absorption profile we can calculate the column density $\mathrm{N_H}$ of the cold component. For the CNM, spin temperature is approximately equal to the kinetic temperature (Dickey & Lockman 1990) so we can substitute T_S for T_k in the column density equation.

$$\mathrm{N_H} = 1.823 \times 10^{18}\ \mathrm{cm^{-2}\ K^{-1}} \int T_k \tau(\nu)\ d\nu \tag{2.5}$$

2.4 H I absorption surveys

H I absorption was first reported by Hagen et al. (1955) through the use of a single dish telescope. Early studies with small interferometers (e.g. Clark 1965; Caswell et al. 1975) focussed on measurements of absorption in front of the brightest continuum radio sources. Later work moved to surveys of larger parts of the sky, largely driven by improvements in radio telescope technology, and particularly in the

development of large interferometer arrays. Early surveys of H I absorption in the Galactic plane used the Arecibo and Greenbank single dish facilities (Dickey et al. 1978) and the Very Large Array (VLA) (Dickey et al. 1983).

In 1999, observations started at the Arecibo observatory for the Millennium Arecibo 21 Centimeter Absorption-Line Survey (Heiles & Troland 2003a). This was a set of long integration time observations of 79 mostly high Galactic latitude bright sources. While the primary objective of the survey was to provide H I Zeeman splitting data, the highly sensitive observations provided very high signal to noise H I absorption spectra towards these sources. These were used to analyse the properties of both the CNM and WNM at these specific locations Heiles & Troland (2003b).

In the early 2000s a coordinated effort was made to survey the H I of the Galactic plane. Three surveys were undertaken, the Canadian Galactic Plane Survey (CGPS; Taylor et al. 2003), the VLA Galactic Plane Survey (VGPS; Stil et al. 2006) and the Southern Galactic Plane Survey (SGPS; McClure-Griffiths et al. 2005. These surveys made a huge leap forward in the coverage and detail level of the Galactic H I emission and absorption. They provide coverage of 90% of the galactic disc with a spatial resolution of 2.2 arcmin or better, a velocity resolution of 0.8 km s^{-1} and a 1σ sensitivity of 1.6K. There is a gap in coverage in the third quadrant between $-170 < l < -110^{\circ}$ which is outside the observed areas of all three surveys, and also at the galactic centre where velocity crowding makes observing far more complex. The improved sensitivity and resolution has revealed new features of the Galaxy such as the previously mentioned new spiral arm and proven an excellent tool in tracing the structure particularly of the outer galaxy (Strasser & Taylor 2004). The surveys have obtained a wealth of absorption spectra, with Strasser et al. (2007) cataloguing spectra towards 793 extragalactic sources and Brown, C et al. (2014) cataloguing absorption features recorded in SGPS towards 252 HII regions.

In the Northern hemisphere, the "21 cm Spectral Line Observations of Neutral Gas with the EVLA" (21-SPONGE Murray et al. 2015) project has achieved new sensitivity levels. It is observing selected very bright sources to obtain optical depth spectra with noise levels of $\sigma_\tau = 0.0006$ and has successfully detecting the broad and weak absorption lines from the warm neutral medium.

In the near future, observing for the Galactic ASKAP survey (GASKAP; Dickey et al. 2013) will commence. GASKAP will cover the Southern sky, covering quadrants 3 and 4. It will provide a factor of 5 increase in the number of absorption spectra recorded with $\sigma_\tau \leq 0.05$ over previous surveys such as SGPS and will achieve optical depth noise of $\sigma_\tau \leq 0.02$ for sources of S = 50 mJy. It aims to record absorption at a 10 arcsec spatial resolution, 0.206 km/s velocity resolution and a 1σ sensitivity of 2K at the highest resolutions.

2.5 Interferometry Observations

A short summary of some important radio interferometry concepts is useful for understanding some of the compromises and decisions made in this work.

A radio interferometer consists of multiple antennae observing a region of sky simultaneously. Each pair of antennae forms a baseline, so

$$n_{baselines} = n_{antennae}(n_{antennae} - 1) \tag{2.6}$$

A delay time, based on the length of the baseline, is used to adjust the amplitudes from one antenna in each baseline so that the radiation received from the target sky region interferes constructively. This is termed the cross-correlation of the the two amplitudes.

A synthesized beam is formed from combining the responses on each baseline. The half power beam width of the synthesised beam is based on the length of the longest baseline using Eq. (2.1). So for a 6 km array such as the ATCA, the beam width will be 9 arcsec at 1.4GHz and 6 arcsec at 2

GHz.

With discrete antennae separated by large distances, the instantaneous sampling of the sky is sparse. Due to the cross-correlation between antennae the sampling is actually of the Fourier transform of the sky brightness, the u-v plane. As we are sampling in the Fourier domain, the largest structures are seen only in the shortest baselines. As most emission comes from the larger structures our power is also affected by the baseline length choice.

In an East-West array such as the ATCA, we rely on Earth rotation synthesis to improve the sampling. Over a period of hours the angle of the array changes with respect to the sky region being observed, thus the array will make a more complete sample of the u-v plane. This also affects the beam shape, as with an instantaneous sampling from an East-West array the beam is just a line in the u-v plane. As the array rotates with respect to the sky, a more circular region of the u-v plane is sampled and the beam becomes less elongated. So ideally we would observe for 12 hours to get optimal u-v coverage. However given the need to balance telescope time with coverage, we often interleave observations of multiple fields over time e.g. field a, phase cal, field b, phase cal, field c, phase cal, field a, phase cal etc. This will sample the u-v plane at different angles with the aim being to evenly sample 180 degrees. Figure 2.3 shows the result of such an interleaved observing pattern for a field.

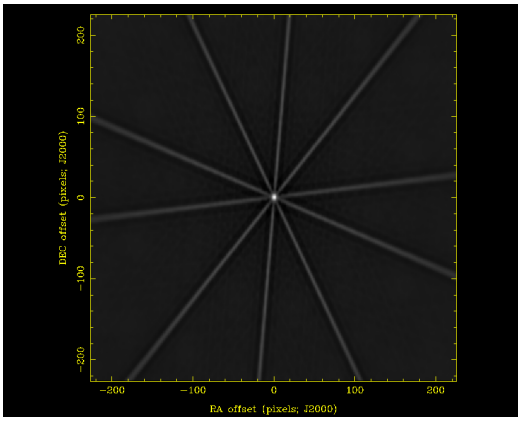

Figure 2.3: Example beam pattern after five interleaved observations over 12 hours.

The noise levels of an observation of a point source with an interferometer are governed by the radiometer equation (Johnston & Gray 2006, eq. 6):

$$\sigma_s \propto \frac{T_{sys}}{A\sqrt{N(N-1)Bt}} \ , \tag{2.7}$$

where σ_S is the rms of the noise level for a point source, T_{sys} is the system noise temperature of the telescope, A is the area of one antenna, N is the number of antenna in the array, B is the bandwidth of each sample in Hertz and t is the integration time in seconds. This means that wider bandwidth and longer integration times result in lower noise. The collecting area of the dish also affects the noise levels of the observation, with a smaller antenna receiving less emission, thus increasing the relative noise level.

The most commonly used data format for radio interferometry data is the visibility format. This format stores the response for each baseline in the u-v plane. This needs to be combined with calibration data and then imaged using Fourier transformation. As u-v plane sampling is sparse we also need to clean the data to reduce effects from bright sources such as sidelobes.

2.6 Processing Pipelines

When reducing small volumes of data, the general approach used has been for each astronomer to have their own set of scripts generally doing a step at a time and leaving some steps as interactive. Of course, processing toolkits such as Miriad, AIPS and CASA have been around for a long time and do much of the heavy lifting for data reduction. However as the volume of data being produced by surveys increases, these techniques are not sufficient to keep up.

The Sloan Digital Sky Survey (SDSS) had one of the first large pipelines for images and spectra (Ivezić et al. 2004). As part of the preparation for SKA, the pathfinder radio observatories are building

automated data reduction pipelines, with LoFAR (van Haarlem et al. 2013) setting up imaging and pulsar pipelines for automated data reduction, ASKAP creating ASKAPsoft (Cornwell et al. 2011) to accomplish imaging and spectra extraction from the raw visibilities. Individual surveys are also examining automated pipelines, such as the current work on the DALiuGE framework (Wu et al. 2017) for the CHILES (Fernández et al. 2016) project.

One area that has emerged as a non-trivial and important problem is source finding (Hopkins et al. 2015). A variety of source finding toolkits have emerged in response to this such as Aegean (Hancock et al. 2012), and Selavy (Whiting & Humphreys 2012).

2.7 The MAGMO Survey

The MAGMO survey was an Australia Telescope Compact Array observing program run from 2010 to 2012 which had the aim of studying "Magnetic fields of the Milky Way through OH masers" (Green et al. 2012b). It was primarily concerned with measuring Zeeman splitting of hydroxyl (OH) ground state spectral lines. Many of the OH masers observed by the MAGMO survey were associated with 6.7 GHz methanol masers. These emissions are linked to high mass star formation (HMSF) regions (Minier et al. 2003). As a secondary objective, neutral hydrogen observations were taken towards most of the same sources (Green et al. 2010). As a result these H I observations will be largely of HMSF regions.

While the OH maser data has been partially analysed in (Green et al. 2012b) and processing of the rest of the OH data is in progress, the H I data has not been processed and analysed previously.

2.8 The MAGMO-H I Project

In the current work we have taken the MAGMO H I absorption data, created an automated processing pipeline, and used it to reduce and analyse the data. The analysis was twofold. First, we evaluated the effectiveness of the pipeline, including comparing both new and previously observed sight-lines with previous observations. Secondly, we calculated major characteristics of the observed H I, compared them against previous studies and examined what this widespread absorption dataset tells us about the observed H I.

This project extends the SGPS analysis of the Southern Galactic plane in three ways.

Firstly, as it will utilise automated Gaussian decomposition, it will allow the automatic decomposition results to be compared to the carefully decomposed absorption spectra of the SGPS. This comparison will inform the analysis strategies of the upcoming GASKAP project.

Secondly, the absorption-only MAGMO data will allow verification of the subtraction of emission spectra in the SGPS to produce absorption spectra.

Thirdly, with a Galactic longitude range more extended than SGPS, this project will provide absorption data for regions not covered in the SGPS nor in other Galactic plane surveys.

As the MAGMO observations were largely aimed towards high mass star forming regions, the HI dataset also provided an excellent opportunity to examine the cold neutral medium towards these complex regions. The high spatial resolution observations are important to be able to map these small scale structures.

Chapter 3

Data Analysis

3.1 The MAGMO Observations

Between January 2010 and June 2012, 644 observations were made of 440 targets with the ATCA, including repeats for observations affected by equipment issues, Radio Frequency Interference (RFI) and weather. The observations are all targeted at sites of 6.7 GHz methanol masers observed during the methanol multibeam survey (Green et al. 2012a). The targets of the observations are shown in Fig. 3.1 with a background image of the Milky Way H I intensity from the HI4PI survey (HI4PI Collaboration et al. 2016).

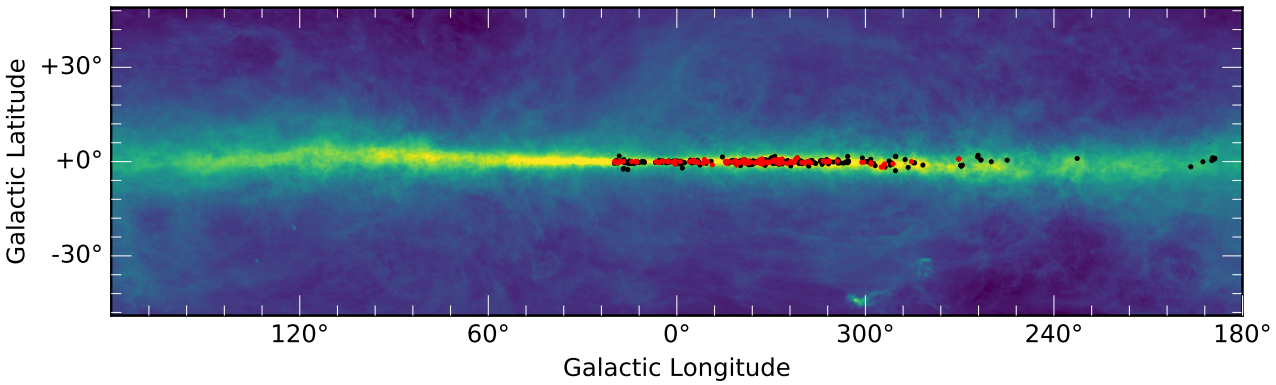

Figure 3.1: Distribution of fields observed with fields used shown in red.

In addition to the OH observations, the ATCA's zoom mode was used to obtain simultaneous high spectral resolution observations (2048 channels each 0.5 KHz wide) of H I with a ∼400 km/s velocity coverage, along with a continuum band (2049 channels each 1 Mhz wide centred on 1757 MHz) (Green et al. 2010). The observations were all at low galactic latitudes ($|b| \leq 2.9$) and covered the longitude range 188° to 19°. As the 6 km baseline configuration of the ATCA was used, the HI observations have a spatial resolution of 7.5". Each source was observed for a total of 30 minutes, in 5 scans of 6 minutes each across a 12-hour observing period.

While this is not highly sensitive, it is sufficient for detecting HI absorption. and is slightly more sensitive than the SGPS, which took 20 minute observations of each source. The observations also have higher angular resolution than the SGPS. Due to the higher angular resolution, the MAGMO noise levels will be slightly higher than SGPS, despite SGPS only taking 20 minute observations of each source. The MAGMO scans were interspersed with calibration observations and observations of other sources.

The response from each pair of antennas, or baselines, was recorded in a raw visibility format as described in Section 2.5.

Listing 3.1: ADQL query to return all MAGMO observation files with H I data

```
SELECT obs_id, configuration, number_scans, frequency,
    access_estsize, access_url
FROM ivoa.obscore
WHERE obs_collection = 'C2291' and frequency in (1420, 1420.5, 1421)
GROUP BY obs_id, configuration, number_scans, frequency,
    access_estsize, access_url
```

3.2 The MAGMO Data

The uncalibrated visibility data are held in the RPFITS[1] format in the Australia Telescope Online Archive (ATOA). These files can be listed using the ATOA Table Access Protocol service at `http://atoavo.atnf.csiro.au/tap` with the query shown in Listing 3.1

Each file contains data for H I , continuum and OH observations for targets and calibration sources. In total for the 43 half days of observing, 1.6TB of data were generated.

The processing of this data, as described in the rest of this chapter, produces the following data products:

- **Continuum Image** – a FITS format image of each field from the 1757 MHz centre band, with a 1 arcsec resolution, generally covering a 19x19 arcmin area
- **H I Image Cube** – A FITS format image cube of selected fields with spatial and velocity axes, with a 5 arcsec spatial resolution and a 0.8 km/s spectral resolution, generally covering a 32x32 arcmin area and a -250 to 250 km/s H I velocity range
- **H I Absorption Spectra** – VOTable format tables of the optical depth vs velocity for selected background continuum sources
- **Catalogues** – VOTable format tables of the details of each field, continuum source, continuum island, H I spectrum, H I Gaussian component and H I gas component
- **Preview Images** – PNG format preview images of each continuum image, H I image cube and H I spectrum

All data are available on request and will be published in conjunction with a paper which is under preparation.

Throughout this thesis we will be using the field '291.270-0.719' which was observed on 10 Sep 2011 (day 21), as an example. From that field, the spectrum 'MAGMOHI G291.277-0.716' against the background source 'OH 291.3-0.7' will be used as an example spectrum. As detailed in Section 3.5, the spectra are named after the Galactic longitude and latitude of their location.

3.3 Data Reduction Strategy

Processing large volumes of data brings a scaling problem which often breaks traditional processing models. Typically an astronomer would have a set of scripts which they modify and run for each data file, or which list the specific data files and commands to be be run. The maintenance of these becomes onerous at large volumes, particularly where a change in approach or parameters may result in the need to change many copies of the same commands. When exploring large data sets it can be

[1] RPFITS is an ATNF specific variant of the FITS astronomical file format

essential to be able to tweak parameters and then rerun the processing to see the effect on the results, as well as to be able to run the pipeline on new data files such as the latests observational results. This has led to the introduction of configurable pipelines for processing large survey datasets as detailed in Section 2.6.

We have implemented such a pipeline with the aim of automating all steps wherever possible and to be simple to rerun in order to incorporate new data and tune the process. This also allowed the processing to be easily repeated both to ensure reproducibility and to allow different settings to be compared. The MAGMO-HI pipeline, as used for producing this thesis, is available at doi:10.5281/zenodo.1036733.

As noted earlier, due to the emphasis on testing the automation of processing, the processing is described in greater detail than typical in this section. This allows for the assessment of the effectiveness of various parts of the automation. It also facilitates the independent reproduction of the results presented in this thesis.

The pipeline was split into phases to allow the tasks to be rerun individually. The main phases were:

- find_data – Identify the RPFITS data files to be downloaded from ATOA
- load_data – Extract the required data from the RPFITS files for a specified day
- process_day – Flag, calibrate, produce 2 GHz continuum images and H I image cubes for a specified day
- analyse_data – Identify sources and extract spectra for a specified day
- analyse_spectra – Aggregate all spectra, assess quality, produce reports, diagrams and catalogues
- decompose – Decompose each spectrum into a sum of Gaussian components
- examine_gas – Calculate physical characteristics of the gas and compare it with results from other studies

While process_day and analyse_data, in particular, operate on data volumes that require use on larger processing servers, the analyse_spectra and later tasks can operate on the lower volume output data products using a normal laptop.

Notable tasks automated for this study included:

- Automatic flagging of the 0 km/s region in the 1420 MHz bandpass calibrator data
- Filtering the fields to be used for spectra extraction based on exclusion of both duplicate observations and poor quality observations
- Source finding using the `Aegean` package
- Selection of sources at which spectra will be extracted
- Extraction of emission data around the sources from the SGPS data
- Fitting Gaussian components to the absorption spectra
- Production of diagnostic plots for manual verification of the results

While many of these were non-trivial to prepare generalised solutions for, in combination they mean that the pipeline can run without manual intervention and process large volumes of data efficiently.

The pipeline was also designed with traceability of results in mind. The observing day, field and source identifications as well as the spatial location were recorded with each spectra. As the source identification may change as the input continuum images were refined, a position-based identifier was also assigned to each spectrum to allow traceability.

To facilitate traceability the VOTable format (Ochsenbein et al. 2013) was used for recording any tabular data such as catalogues and spectra. This is a flexible container for tabular data that allows for full metadata to be held with the data. It has wide support including in `Astropy` (Astropy Collaboration et al. 2013) and `TOPCAT` (Taylor 2005), both of which were extensively used in the production of this work. The use of Virtual Observatory compatible data exchange formats also allows the easier publication and reuse of the data.

Some areas remained which required human intervention. The largest of these was the flagging of known bad data. The list of time, frequencies, sources and fields that should be flagged and ignored was manually curated based on the observing log and manual examination of the diagnostic plots. This meant that there was limited flagging of the target source data and the phase calibrator data.

A potential future improvement for the phase calibrator data would be to automatically identify time periods where the fluxes vary from the expected flat amplitude, or show excessive polarisation. These time periods could then be dynamically flagged. This would automatically filter out short term RFI in the phase calibrator observations. As the target sources may have some variability, this approach would not be usable on those data without the risk of losing valid results.

The processing was mostly single-threaded. Using appropriate infrastructure, the processing problem would be amenable to parallel processing. A simple approach would be to run the process_data and analyse_data steps on a number of days in parallel. This would require minor modification to the control scripts and also testing to ensure that this doesn't slow down the overall result due to I/O contention, or exceeding the memory or CPU capacity of the processing server.

3.4 Miriad Processing

The Miriad (Sault et al. 1995) tool suite was used to extract, calibrate and image the data. A standard Miriad recipe was followed to produce continuum FITS images and 3D H I image cubes from the visibility data. The main steps involved were:

1. Convert the RPFITS format data to Miriad's visibility format and extract the data of interest.

2. Flag the data that was known to be invalid. A combination of set rules (e.g. amplitude > 500 Jy and the 200 channels at the edges), and a curated list of known bad data was used. In addition the local HI absorption was excluded from the bandpass calibration data by flagging the 110 channels around a LSR velocity of 0 km/s. Finally values more than 3 sigma from the mean in the phase calibrator data were also flagged.

3. Produce bandpass, flux calibration and phase calibration solutions and apply them to the continuum and H I data.

4. Fourier invert the continuum visibility data from the u-v plane into the image plane.

5. Iteratively clean (or deconvolve) the continuum images to reduce sidelobes from strong sources in or adjacent to the image, and to account for the incomplete sampling of the u-v plane by the interferometer.

6. Restore the continuum image by convolving the clean image with a Gaussian representation of the beam and adding it to the residual image.
7. Identify which fields would have sufficient signal to noise to be worthwhile producing H I cubes.
8. Average the H I data to 0.8 km/s spectral resolution to improve the signal-to-noise ratio.
9. Repeat steps 4 through 6 for the H I data to Fourier invert, clean and restore the H I image cubes.
10. Output the continuum images and H I image cubes into FITS format.

The configuration used in many of the steps had to be iteratively refined to produce optimal images and data cubes. The quality of the output data was assessed using metrics such as the signal to noise ratio of the final images/cubes, noise levels in the non absorbed velocity ranges of bright background sources and the errors and warnings reported during the processing. Examples of modifications made were the flagging of phase calibrator data where RFI was apparent, and the selection of an appropriate clean function and parameters.

We split the steps for bulk processing into two Python scripts, `load-data.py` and `process_day.py`. The first script implemented step 1 as this has little configuration or requirement to be rerun. The second script covered the steps 2 – 10, which need to be rerun either partially or completely each time the configuration is modified. In addition, numerous diagnostic images and web pages are produced by the `process_day.py` script to allow manual verification of the performance of each step and the end products.

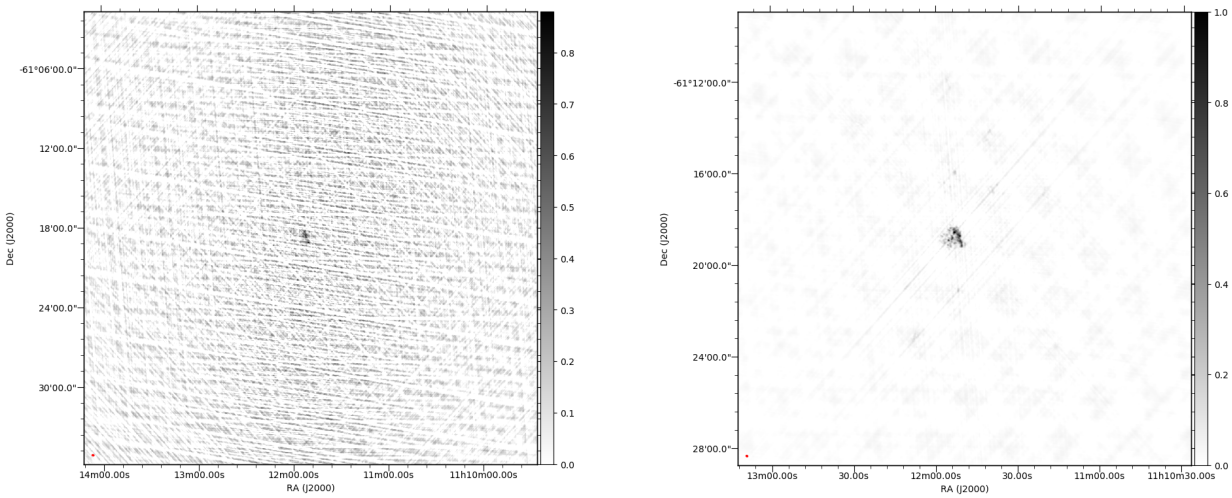

Figure 3.2: Plot of the example field 291.270-0.719 with the total intensity (sum of all channels) of the H I data on the left and continuum (centred at 1757 MHz) on the right. Both plots show flux density in units of Jy/beam on a linear scale with flux increasing from light to dark. The wedge on the right shows the scale.

The end product of the processing was a set of FITS images of each field observed and a set of FITS data cubes of the observations with a signal-to-noise ratio of at least 1.3 in the continuum band image. Figure 3.2 shows H I and continuum images of the example field. The cubes have a spectral resolution of 0.8 km/s and a spatial resolution of 5 arcsec. Out of 522 fields processed, 310 fields had H I image cubes produced for use in the data analysis phase. The breakdown of the fields by day is shown in Fig. 3.3. The processing phase takes approximately 23.5 hours on the RSAA server 'malice' to process 32 days of data.

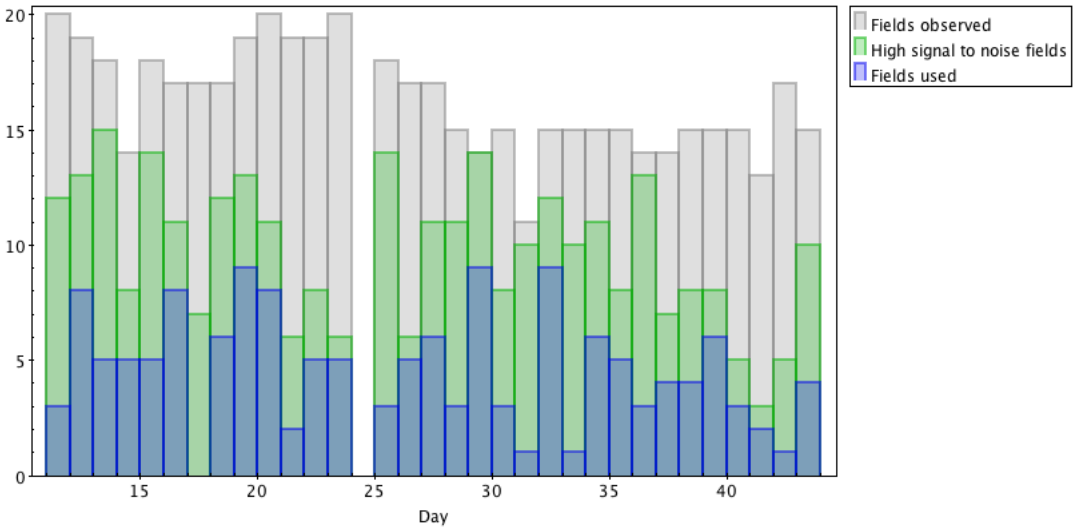

Figure 3.3: Chart of fields observed each day and the number of fields which were used.

3.5 Source Finding and Spectra Extraction

Absorption spectra require the use of a bright background source against which to extract the spectra. To find these background sources we can either use a pre-existing catalogue, or search images of the region for sources. As we had commensal continuum observations available for the H I observations we decided to use these to find suitable sources as they would best reflect the available data. However as described in Section 2.6 source finding is a non-trivial process and as a result we have chosen to use a dedicated source finding package.

Spectra from the MAGMO dataset are produced by identifying sufficiently strong background sources, extracting the flux values at each velocity step at the position of the source, and then dividing the flux by the continuum to produce a spectrum in optical depth units. This process is implemented in the `analyse_day.py` program and run on the data one day at a time.

We have used the `Aegean` package Hancock et al. (2012) to identify the sources from the 1757 MHz continuum images. This is a two-step process, with the `BANE` function first being used to assess the noise levels in each region of the image. The outputs from this are then input into `Aegean`, along with the continuum FITS image to assess sources. We then filter the identified sources to just those with a signal-to-noise ratio (S/N) ≥ 10 and a flux density ≥ 20 mJy. Each accepted source is given an identifier of the form `MAGMOHI Glll.lll+b.bbb` for ease of later reference, where l is the Galactic longitude and b is the Galactic latitude, both in decimal degrees.

At the position of each of the sources in this filtered list we then use `astropy` (Astropy Collaboration et al. 2013) and `numpy` to extract a subcube from the HI image cube around the source. The World Coordinate System (WCS) implementation within Astropy is used to convert the sky position of the source into a pixel position within the image cube. Then a subset of the image cube is taken along the velocity axis of a 5x5 grid around the central source pixel position. The pixels that are not within the ellipse of the source component are zeroed out leaving just the pixels within the source. This subcube is the calibrated flux density across the source at each velocity step.

After this we calculate the brightness of the continuum source in the absence of any H I emission or absorption. Using a standard Galactic longitude-velocity diagram (Hartmann & Burton 1999) we have manually identified velocity regions which are expected to have no Galactic H I gas but that were observed in the MAGMO data. The velocity regions are shown in Table 3.1 and are stored in a lookup table for processing. The average of the flux in this velocity range for each pixel is then used

as the continuum flux for that pixel.

The continuum flux of each pixel is used to calculate the integrated spectrum for the source according to the following equation.

$$1 - e^{-\tau(v)} = \sum_i \left[\left(\frac{c_i^2}{\sum_j c_j^2} \right) s_i(v) \right] \ , \tag{3.1}$$

where c_i is the continuum flux for the ith pixel and $s_i(v)$ is the flux of the pixel at the velocity step. This is based on (Dickey et al. 1992, eq. 2) although it is important to note that the data used here is not continuum subtracted. Finally the flux values at each velocity position are divided by the average continuum level to produce a spectrum of optical depth $e^{-\tau}$ as shown in Eq. (2.2).

Emission data are sourced from the SGPS dataset. A set of 6 points, located 60° apart in a circle of radius 1.08 arcmin (half of the SGPS beam width) is selected. If the point is within any detected island, or has significantly negative emission ($T_B \leq -12$ K), it is moved away from the source location by another half a beam width and this is repeated until it satisfies both criteria. If a point reaches more than ten beam widths from the source then the point is discarded. An emission spectra, in brightness temperature units of K, is extracted from the SGPS data at each specified point. The mean and standard deviation for each velocity step across all of the used points is calculated, with the mean being used as the emission spectra and the standard deviation as the emission noise.

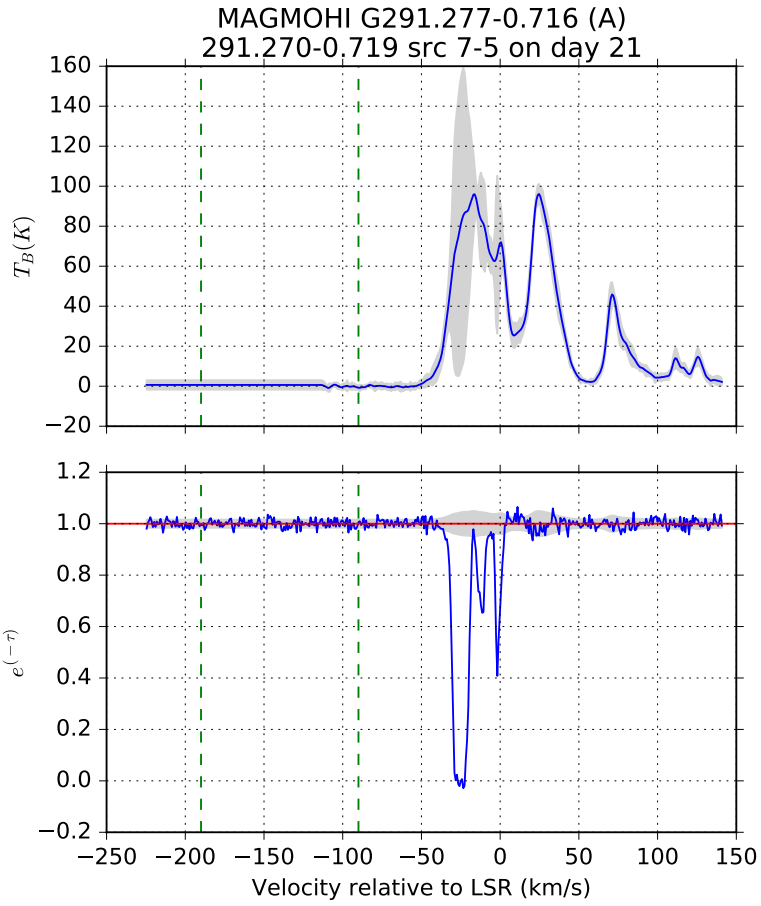

Figure 3.4: Emission (top) and absorption (bottom) spectra for MAGMOHI G291.277-0.716 against the background source 'OH 291.3-0.7'. The continuum region for this Galactic longitude is shown by green dotted lines, the spectrum is in blue and the noise level is in grey.

The pipeline then writes out a VOTable format file for each source with the spectrum in both flux and optical depth units along with general information on the source such as position and identity. It also produces a plot of the emission and optical depth spectra, an example of which is shown in Figure 3.4.

Longitude Range (deg)	Velocity Range (km/s)
0 to 60	-180 to -120
61 to 180	100 to 160
181 to 300	-190 to -90
301 to 310	-200 to -120
311 to 360	120 to 180

Table 3.1: Velocity ranges used to identify the continuum flux level at different galactic longitudes

3.6 Spectra Analysis

The spectra analysis stage, which we implemented in the `analyse_spectra.py` program, was the first of the pipeline stages where the full set of spectra were processed as a whole data set. This stage filtered out duplicate fields and sources, compiled catalogues, assessed the quality of the spectra and produced plots for further analysis.

Each spectrum file was read in and was then automatically assessed for quality using the following quality tests from Brown, C et al. (2014).

- Range – Check that the range $(e^{-\tau}) < 1.5$, reflecting that the optical depth should be between 0 and 1.
- Maximum signal to maximum noise – Check the ratio of maximum signal to maximum noise $\frac{(1 - \min(e^{-\tau}))}{(\max(e^{-\tau}) - 1)} \geq 3$.
- Baseline noise – check the noise level in the continuum region of the absorption spectrum, $3\sigma_{continuum} < 1$.

This results in a rating scale of A – D, with rating A spectra passing all tests and rating D spectra failing all tests.

We then filtered out duplicate spectra. These occur when overlapping fields contain the same sources. Any two sources with a separation of less than four arc seconds were considered duplicates and the poorer quality source was flagged and excluded from further processing. Poorer quality was indicated by a worse quality rating, or for the same rating, a higher baseline noise level.

The pipeline then wrote out the compiled list into a VOTable format file `magmo-spectra.vot` which included additional file metadata to allow the spectra files and their plots to be linked to each catalogue entry. This enabled us to use tools such as TOPCAT (Taylor 2005) to visually explore the spectra.

Similarly the field metadata files written out in the first steps by `process_day.py` were read in by the pipeline, and a combined catalogue of the fields was written out to `magmo-fields.vot`.

3.7 Decomposition of the Spectra

As discussed in Section 2.3 the Cold Neutral Medium can be modelled in absorption as a sum of Gaussian components. Each component represents a discrete cloud of gas. In previous Galactic surveys

each spectrum has been fitted individually by manually determining the number and approximate position of the Gaussians and using a least-squares fitting algorithm to optimise the Gaussian parameters to fit the spectrum. In this study we trialled the use of the Autonomous Gaussian Decomposition (AGD) algorithm as implemented in `gausspy` (Lindner et al. 2015). It is important to note that the decomposition is not unique and that for most spectra there will be multiple possible solutions. AGD automates the standard fitting approach by making the initial estimate of the Gaussians and then applying a least-squares-fit algorithm to produce the final Gaussian components.

AGD was initially trialled on very high signal to noise 21-SPONGE absorption spectra at high Galactic latitudes by Lindner et al. (2015) with good results. Since then it has been used in a comparison with simulated data by Murray et al. (2017) and achieved a recovery rate of 0.53 for low latitude absorption components. The recovery rate is a measure of the algorithm's effectiveness. It is the number of components identified by the AGD compared to the number identified by a manual decomposition by an expert. The lower recovery rate for the low latitude spectra is largely due to the complex and overlapping nature of the components in this environment. Thus with our lower signal to noise spectra of low latitude environments we can expect that there will be limitations in the recovery of components.

As is standard for machine learning techniques, AGD needs to be trained on representative data with known solutions before it is applied to new data.

We generated a sample training dataset of 50 synthetic spectra. Each spectrum was 512 channels wide, and was composed of 4 Gaussians with random parameters in the ranges of 4 – 35 channels wide, 0.25 to 1.25 amplitude, and positioned between 35% and 75% along the spectrum plus a normal distribution random noise with a 1σ noise level of 0.05. This simulates a set of spectra in $1 - e^{-\tau}$ space, each with four potentially overlapping components plus noise.

`Gausspy` includes two approaches to finding the initial components. These are the a) total variation approach, where the first to fourth derivatives of the signal are used to find local minima of the curvature and b) convolution approach, where the data is convolved with a Gaussian filter prior to calculating the derivative. We used the convolution approach, as implemented with the `conv` mode, as this provides the best performance for spectra with significant noise. In this mode the alpha value determines the size of the Gaussian kernel used to filter the data. The training routine was run multiple times with varying seed α_1 and α_2 values to determine to where it would converge. The best matches to the training data were obtained when the alpha values converged on $\alpha_1 = 4.36$, $\alpha_2 = 9.37$ so these values were used for the decomposition.

In the decomposition process, we read in all rating A and B spectra which were not duplicates, convert the spectra to $1-e^{-\tau}$ values and then saved them in an internal format for `gausspy` to process. We then passed in the selected alpha values, and a required signal to noise threshold (5) to `gausspy` which processes all spectra in one run. This generally took under 5 minutes to process all 385 spectra.

To check the results of the decomposition we compared the results of the trained decomposition values with an alternative set chosen by experimentation. The alternative alpha values were $\alpha_1 = 3.5$, $\alpha_2 = 4.36$. The trained values performed better for the noisier rating B spectra, with a mean residual difference ($residual_{experimental} - residual_{trained}$) of 0.02 ± 0.06. However for the rating A spectra, they performed equally with the mean residual difference being 0.00 ± 0.02. Figures 3.5 to 3.7 show examples of the extremes of the fit performance of each set of parameters.

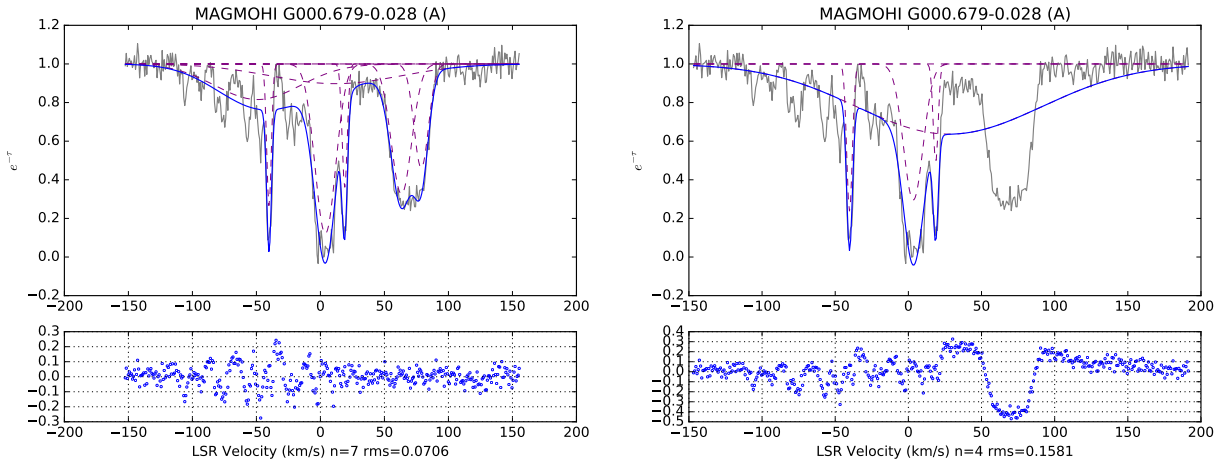

Figure 3.5: Plot of an example spectrum where the parameters produced from training (left) produced a better fit to the spectrum than the experimentally chosen parameters (right). The grey line is the observed spectrum, the red dashed lines are the individual components and the blue line is the total of the components. The bottom plot shows the residual values between the Gaussian component totals and the observed spectrum.

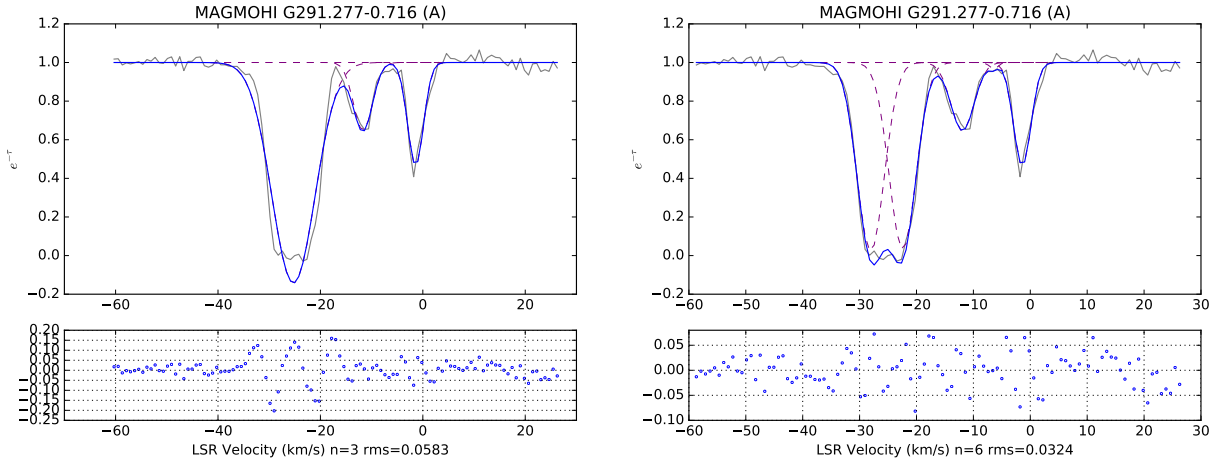

Figure 3.6: Plot of our standard example spectrum, MAGMOHI G291.277-0.716. Here, the trained parameters (left) produced a slightly worse fit for the saturated spectrum than the experimentally chosen parameters (right).

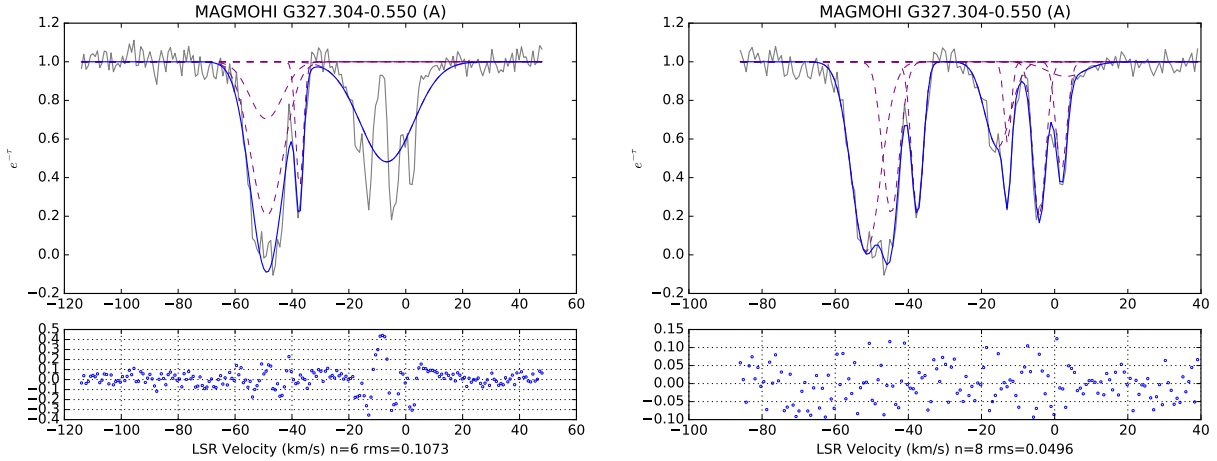

Figure 3.7: Plot of an example spectrum where the parameters produced from training (left) produced a slightly worse fit for the saturated spectrum than the experimentally chosen parameters (right).

Results

4.1 Spectra

In total, we extracted 819 spectra towards unique sources in this study. Of those, 559 were of sufficient quality to be plotted (rating A – C) and 238 were suitable for decomposed (rating A – B). The quality assessment and distribution by rating of the spectra is shown in Fig. 4.1

Absorption was detected in all of these sight lines. The distribution of the absorption is shown in Fig. 4.2. Notably, all of the rating A – C spectra contained velocity ranges with deep absorption where $e^{(-\tau)} < 0.25$, reflecting the ubiquity of cold H I gas at low latitudes. This can be seen in Fig. 4.2 as all longitudes have some bright points.

In Fig. 4.3 we can see that the absorption points in the higher quality spectra are confined to the velocity regions where neutral hydrogen emission was observed in the GASS data (McClure-Griffiths et al. 2009). In the fourth quadrant ($-90 < l < 0°$) the absorption traces the inner Galaxy terminal velocity line well at $v < 0$ km/s.

Comparing Figs. 4.2 and 4.3 we can see that the exclusion of the rating C spectra gives a much lower noise level and still has coverage of much of the survey region. For this reason, further analysis is confined to just the rating A and B spectra.

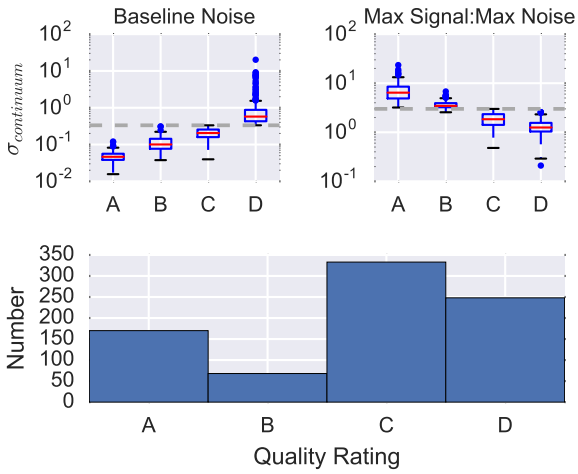

Figure 4.1: Quality metrics for the MAGMO H I spectra (see Section 3.6). The dotted line in the top two plots shows the cut-off value for the test. Both tests are unit-less ratios.

While most of the gas observed is in the inner Galaxy we do observe some outer Galaxy H I gas (outer is where the Galactic radius is greater than that of the solar circle). In the $-70 < l < -45°$ region we see absorption both in the inner and the outer Galaxy. Some absorption in the outer Galaxy can be expected, as while the fields are targeted towards Galactic sources, any other sufficiently bright continuum sources in the field will also be used as background sources for spectra. If these sources are extragalactic (e.g. quasars) then their light may be absorbed by cold H I clouds in the outer Galaxy.

The limited observations of the first Galactic quadrant cover only the inner Galaxy, with almost all absorption in the first quadrant of Fig. 4.3 being in positive velocities.

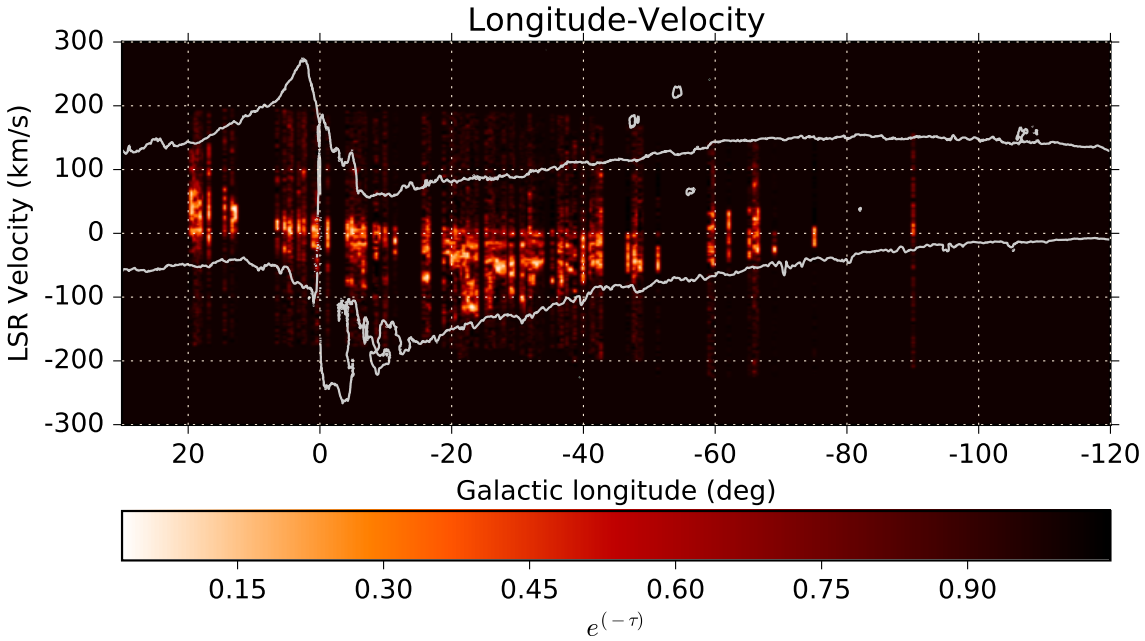

Figure 4.2: Longitude-velocity diagram of the H I absorption measured by MAGMO. The outline is the outer range ($T_B = 1$ K) of H I emission measured by GASS (McClure-Griffiths et al. 2009)

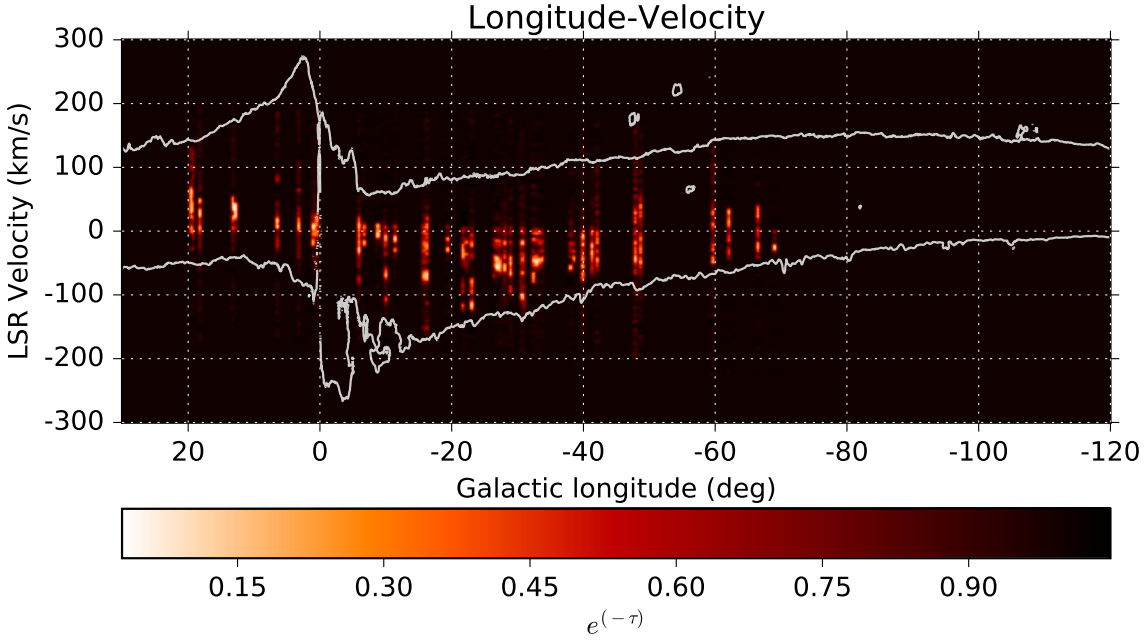

Figure 4.3: Longitude-velocity diagram of the H I absorption measured by MAGMO for just the A and B rating spectra. The outline is the outer range ($T_B = 1$ K) of H I emission measured by GASS.

4.2 Analysis of the gas

The 238 unique rating A and B spectra were decomposed into Gaussian components using `GaussPy`. In total, 922 components were identified, hence an average of four components per spectrum. In four cases, `GaussPy` failed to identify any components for the spectra. All of these were rating B spectra with significant noise. An example of the gas characteristics are shown in Table 4.1.

4.2.1 Gaussian Components

The distribution of optical depth and full width at half maximum (FWHM) of the identified components are shown in Fig. 4.5. The prevalence of saturated velocity ranges in the spectra is reflected in the components, with slightly more than 15% having $\tau > 3$, which is generally regarded as saturated ($e^{-\tau} < 0.05$). The FWHM distribution shows that most components are narrow, with the peak at 5 – 7.5 km/s, reflecting the population of small cold clouds with low internal velocities. This is quite similar to the typical width of 7 km/s given by Dickey & Lockman (1990) for the CNM. There is a significant tail of very wide components which physically imply clouds that either have very high-temperature or are very extended along the line of sight. Both of these are unlikely for cold gas and at the sensitivity of this survey, so a more likely explanation is that multiple narrower components have been fitted as a single component, an example of which can be seen in Fig. 4.4. This is possible when noise obscures the form of components, or where there is strong saturation which causes the line to expand.

Noise in optical depth τ scales according to $\sigma_\tau \times e^\tau$ (Dickey & Lockman 1990). As a result saturated regions are much more affected by noise than optically-thin regions. Given the continuum noise in the spectra and the deep absorption within the Galactic plane, the $exp(-\tau)$ values can be pushed into the negatives where the log function diverges, so the fitting to Gaussians was done in $exp(-\tau)$ space. When the spectra was saturated or negative, the $exp(-\tau)$ value was limited to 1×10^{-16} to calculate the τ value, setting a limit of 36.8 on the τ values.

The equivalent width for each Gaussian component was calculated as $(1 - \exp(-\tau)) \times$ FWHM. The overall distribution of values is shown in the left-hand panel of Fig. 4.6. Much like the FWHM, while 91% of the values were less than 20 km/s, there was a long tail of high values reaching up to a maximum of 280 km/s. These higher values came from the high FWHM values previously discussed, and the highest occurred where there were multiple components fitted with a single broad and deep component. Excluding records with a FHWM $\geq$ 50 km/s, as shown on the right of Fig. 4.6, restricts the highest equivalent width to 45.46 km/s, again highlighting that these are not broad shallow components as expected for absorption in the warm neutral medium.

The distribution of the values by longitude and velocity is shown in Fig. 4.7. The high equivalent width values are most apparent in the regions where the gas is crowded and there are very close or overlapping components, precisely the areas that are hardest to decompose into individual components.

Name	**Velocity (km/s)**	**Optical Depth**	**FWHM (km/s)**	T_S **(K)**	N_H **cm**$^{-2}$	M_t
MAGMOHI G291.277-0.716A	-25.26	-0.14	9.93	81.9 ± 73.3	5.82×10^{22}	10.3
MAGMOHI G291.277-0.716B	-1.41	0.47	3.66	133.0 ± 50.0	7.19×10^{20}	2.2
MAGMOHI G291.277-0.716C	-11.68	0.65	4.36	228.1 ± 64.0	8.38×10^{20}	1.9

Table 4.1: Selected gas component details for the sample spectrum MAGMOHI G291.277-0.716. A full copy of all gas statistics is available in electronic form in the file magmo-gas.vot

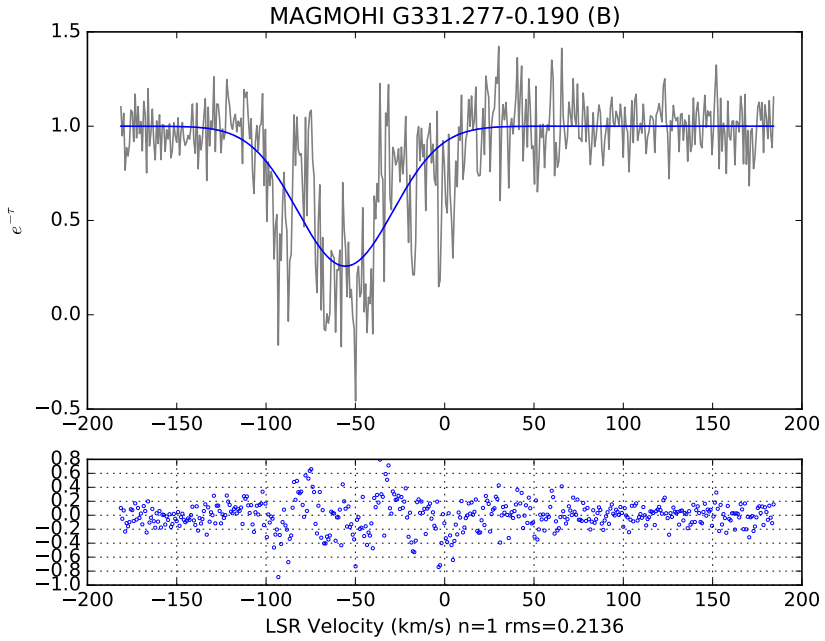

Figure 4.4: Fit for MAGMO absorption spectrum for G331.26-0.19.

4.2.2 Spin Temperature

The spin temperature is calculated using the following equation (Strasser et al. 2007, eq. 2):

$$T_{spin} = \frac{T_B}{1 - e^{-\tau}} \ , \tag{4.1}$$

where T_B is the brightness temperature of emission at the closest velocity step to the centre velocity of the absorption component. This provides a density-weighted average of the temperatures of the gas in the velocity region. The spin temperature provides a lower limit on the kinetic temperature and is generally the same as the kinetic temperature for the CNM (Dickey & Lockman 1990). Alternatives to this calculation are discussed in Section 2.3.

The spin temperature values are shown in the left hand plot of Fig. 4.9 and their distribution by longitude and velocity is shown in Fig. 4.8. There are clusters of higher spin temperatures ($T_S \geq 200$) scattered throughout the longitude range, but predominantly in the velocities closer to Local Standard of Rest (LSR) velocity. These will be discussed further in Section 5.2.

The uncertainty in the spin temperature is calculated using the following equation:

$$\Delta T_S = T_S \sqrt{\left(\frac{\sigma_{T_B}}{T_B}\right)^2 + \left(\frac{\sigma_{continuum}}{1 - exp(-\tau)}\right)^2}, \tag{4.2}$$

where σ_{T_B} is the standard deviation of the emission at the sampled point and $\sigma_{continuum}$ is the measured noise level in the absorption spectrum in optical depth units (see Section 3.5). A comparison of spin temperature to the uncertainty is shown in Fig. 4.10, both as the uncertainty value and the fractional uncertainty. While the median uncertainty is 32%, there are two groups with higher uncertainties.

The first, and largest, group is in the coolest gas, where the uncertainty in the emission (σ_{T_B}) is the main source of uncertainty. In these cases there were large differences between the emission brightness temperature in the locations around the source. The most frequent cause of this is that the gas sampled is still affected by absorption, indicating that our efforts to avoid absorbing gas was not entirely successful. This could be remediated in future work by rejecting points that are significantly

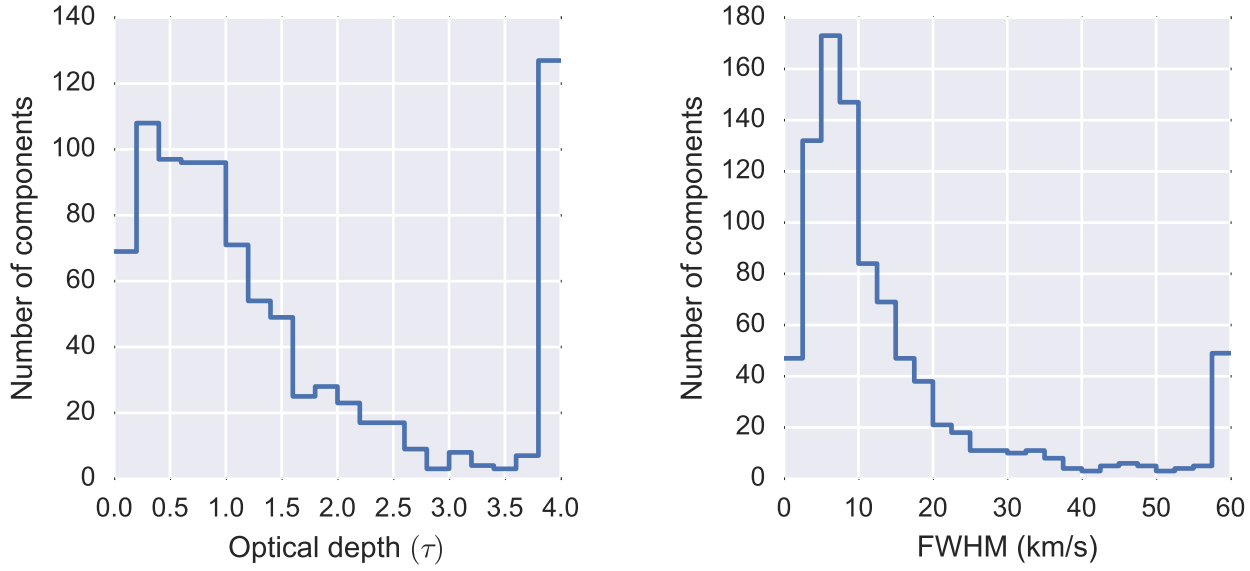

Figure 4.5: Left: Optical depth (τ) of each gas component, right: Full width at half maximum for each component. In both figures, the right bin also contains all components above that value.

below the average. This may require the use a higher number of initial samples to cope with the possibility of discarding some sample points.

The second group are the shallow components which make up all of the gas above 200 K. In some of these cases the identified component has an optical depth only a small multiple of the noise level. While the noisiest spectra have already been rejected, the use of noise values for the spectra in `GaussPy` should be investigated in the future to determine if the matching of these unreasonably shallow components can be avoided.

4.2.3 Column Density

We calculate the column density of the absorbing gas using the following equation (Murray et al. 2015, eq. 13)

$$N_H = 1.823 \times 10^{18} \cdot T_s \cdot 1.064 \cdot \tau \cdot \Delta v \ , \tag{4.3}$$

where T_s, τ and Δv are all characteristics of the Gaussian component and 1.064 converts the product of τ and Δv to be the area under the Gaussian curve. The resulting column densities, shown in Fig. 4.9, are quite high.

H I absorption spectra require a column density of $N_H > 10^{21}$ cm^{-2} to become saturated (Dickey & Lockman 1990). This saturation is thought to indicate the presence of molecular hydrogen (H_2) (Lee et al. 2015). In our spectra, the column densities above this level fall off rapidly and the tail extending beyond $N_H > 5 \times 10^{22}$ cm^{-2} is almost entirely populated by saturated spectra as expected.

4.2.4 Gas Conditions

We calculate the turbulent Mach number of the absorbing gas using the following equation (Heiles & Troland 2003b, eq. 17)

$$M_t^2 = \frac{v_{turb,3D}^2}{C_s^2} = 4.2 \left(\frac{T_{k,max}}{T_s} - 1 \right) \ , \tag{4.4}$$

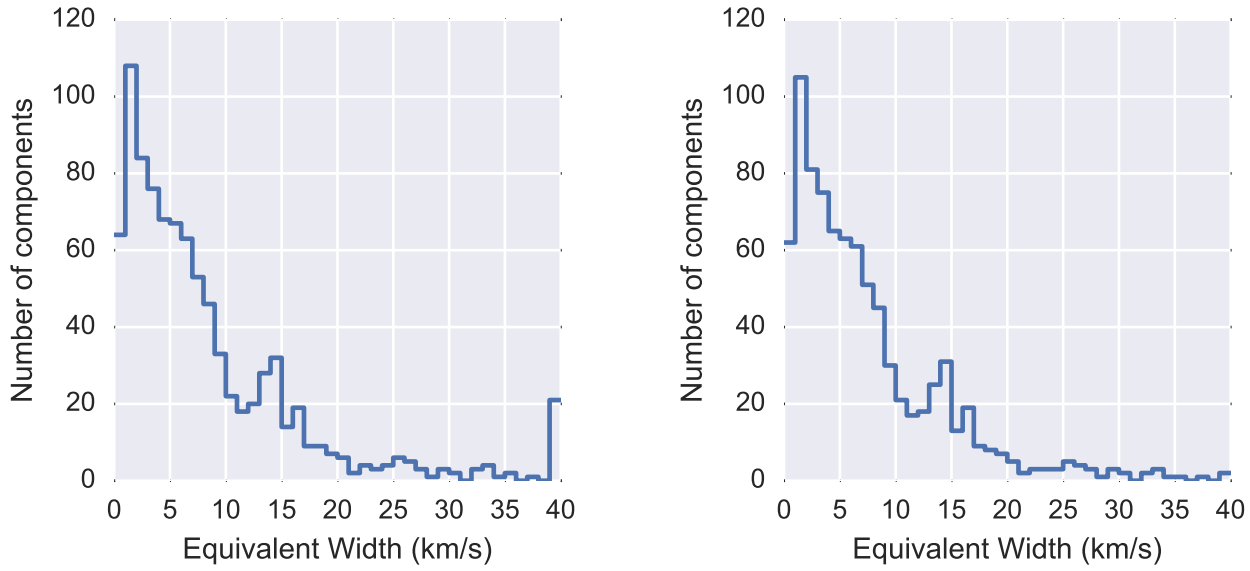

Figure 4.6: Left: Histogram of all equivalent width values. Right: Histogram of equivalent width values for just those components with a FWHM < 50 km/s. In both figures, the right most bin contains all components above that value also.

where $T_{k,max} = 21.866\Delta v^2$ is the maximum kinetic temperature (Heiles & Troland 2003a). As shown in Fig. 4.11 the resulting values show a highly turbulent gas environment. The figure also shows that many of the higher values came from the high FWHM values previously discussed.

Finally we present the fraction of cool gas in this sample using the following equation (Dickey et al. 2000, eq. 7)

$$f_c = \frac{T_c}{T_s} , \tag{4.5}$$

where a cool gas temperature of $T_c = 48$ K (Strasser et al. 2007) was used.

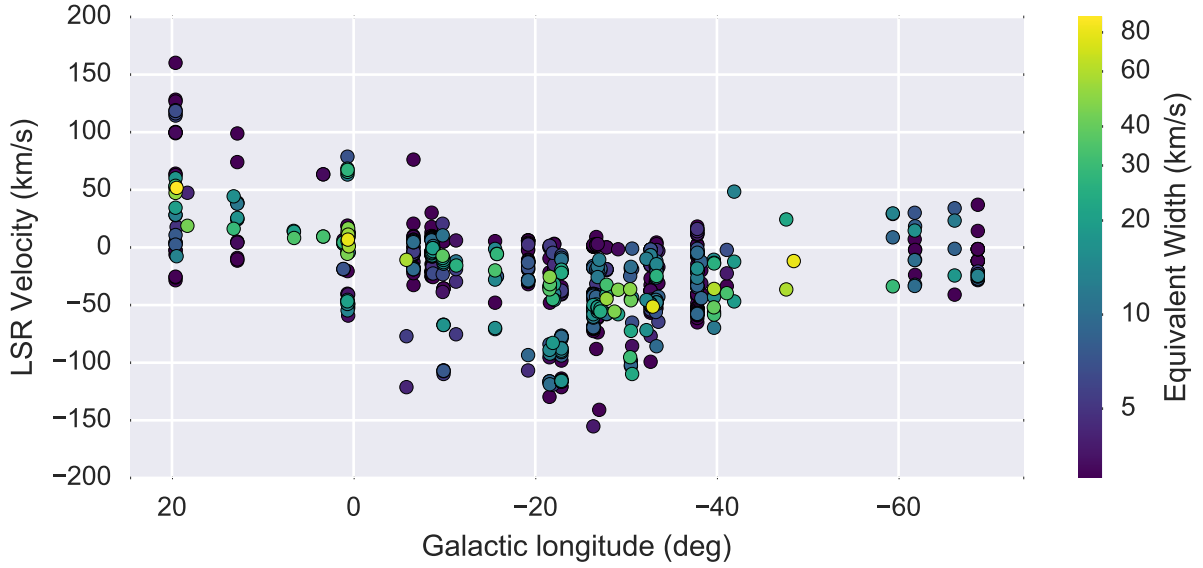

Figure 4.7: Equivalent width plotted by Galactic longitude and LSR velocity. The colour denotes equivalent width on a log scale from 0 to 90 km/s with the scale shown in the wedge on the right.

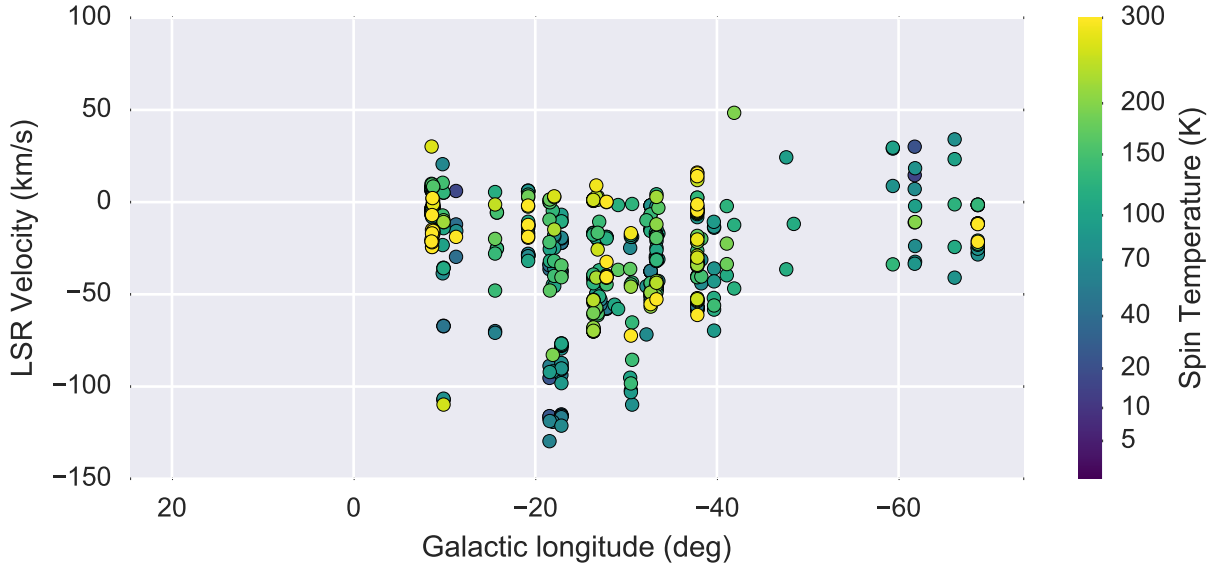

Figure 4.8: Spin temperature plotted by Galactic longitude and LSR velocity. The colour denotes spin temperature on a square root scale from 0 to 300 (or more) K with the scale shown in the wedge on the right.

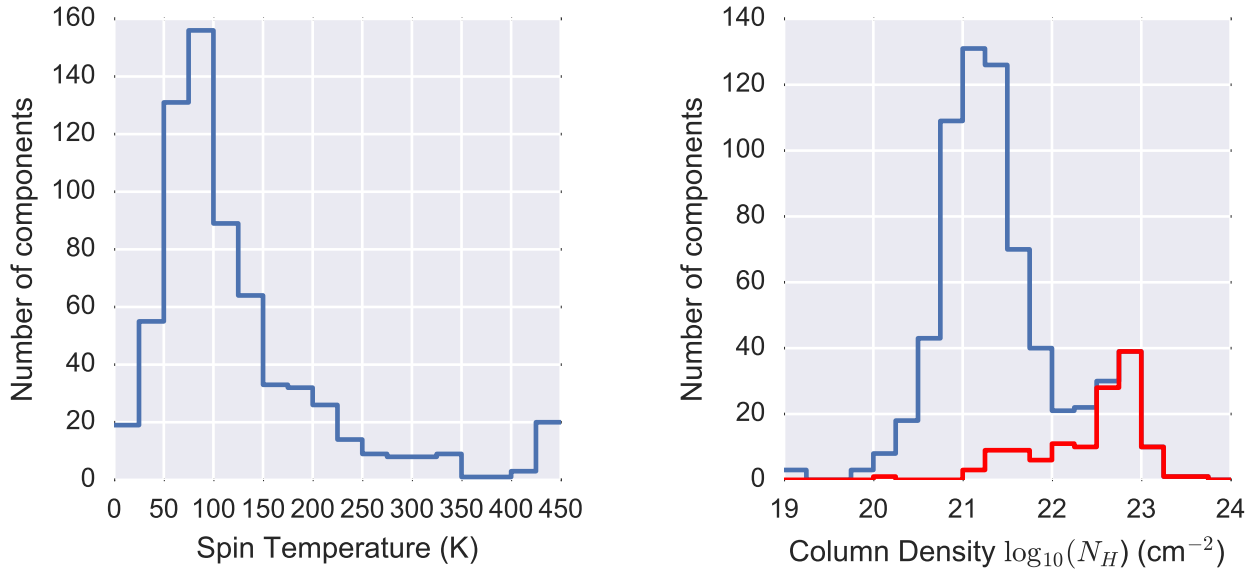

Figure 4.9: Left: Spin temperature of each gas component, Right: Column density of each gas component with the total shown in blue and the density of just the saturated spectra shown in red. In both plots, values are only calculated for gas components where we have emission temperature data from SGPS.

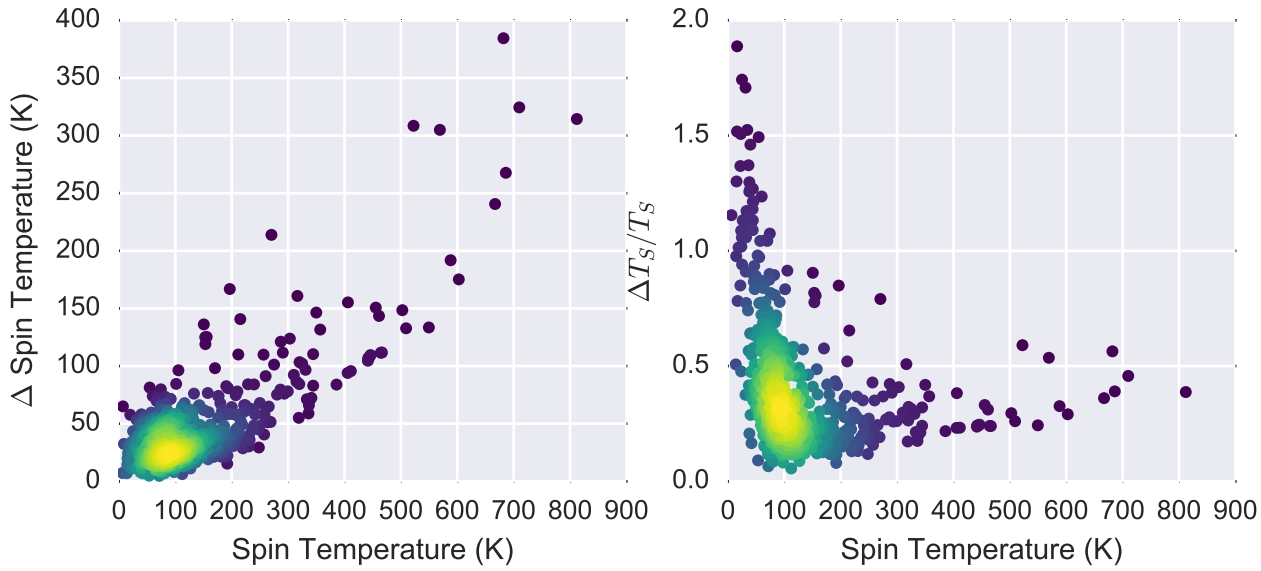

Figure 4.10: Left: Plot of the uncertainty in the spin temperatures against the calculated spin temperatures. Right: Plot of the same uncertainties expressed as a fraction of the measured spin temperature.

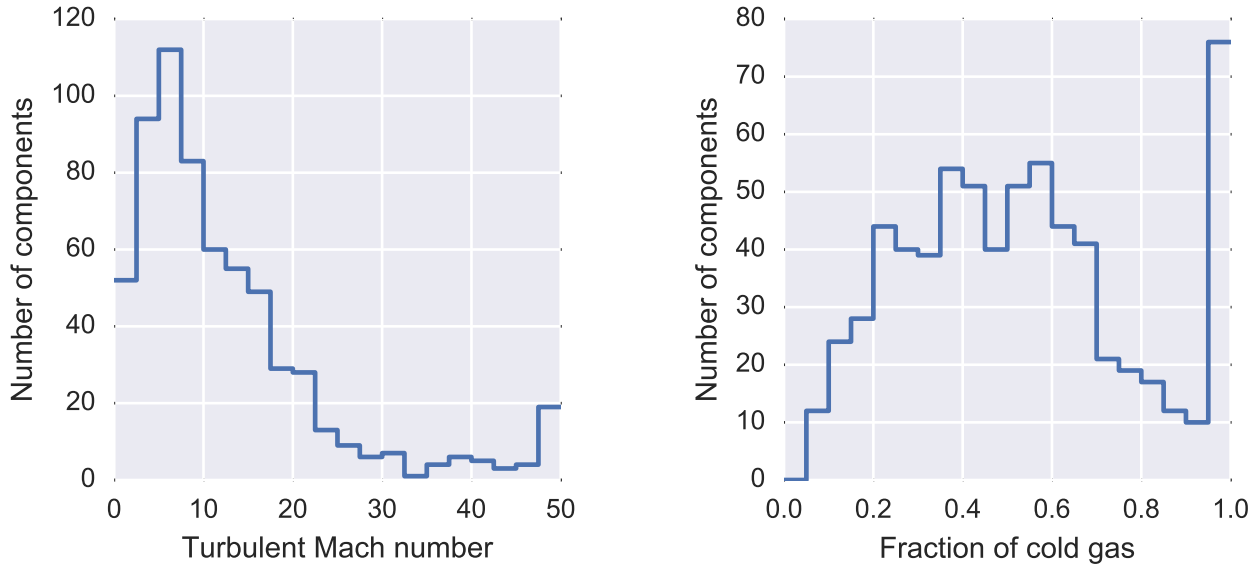

Figure 4.11: Left: Turbulent mach number for each gas component, Right: Fraction of the total gas in the component that is cold. In both cases, values are only calculated for gas components where we have emission temperature data from SGPS.

Discussion

5.1 Comparison with other studies

In this section we compare the results of this study to the literature. We have two objectives here, firstly to see how these results have expanded upon previous studies and secondly to validate the automated processing approach of this study. We start by examining the scientific results in the Section 5.1.1 and then the processing results in Sections 5.1.2 and 5.1.3.

5.1.1 Millennium Arecibo 21 Centimeter Absorption-Line Survey

Heiles & Troland (2003b, hereafter HT03) surveyed absorption against 79 sources, although only 13 where near the Galactic plane, with $|b| < 10$, which are the most comparable to our survey area. They provided a comprehensive description of the physical properties of the CNM. The low latitude sample was further limited to only the eight sources that had good Gaussian fits.

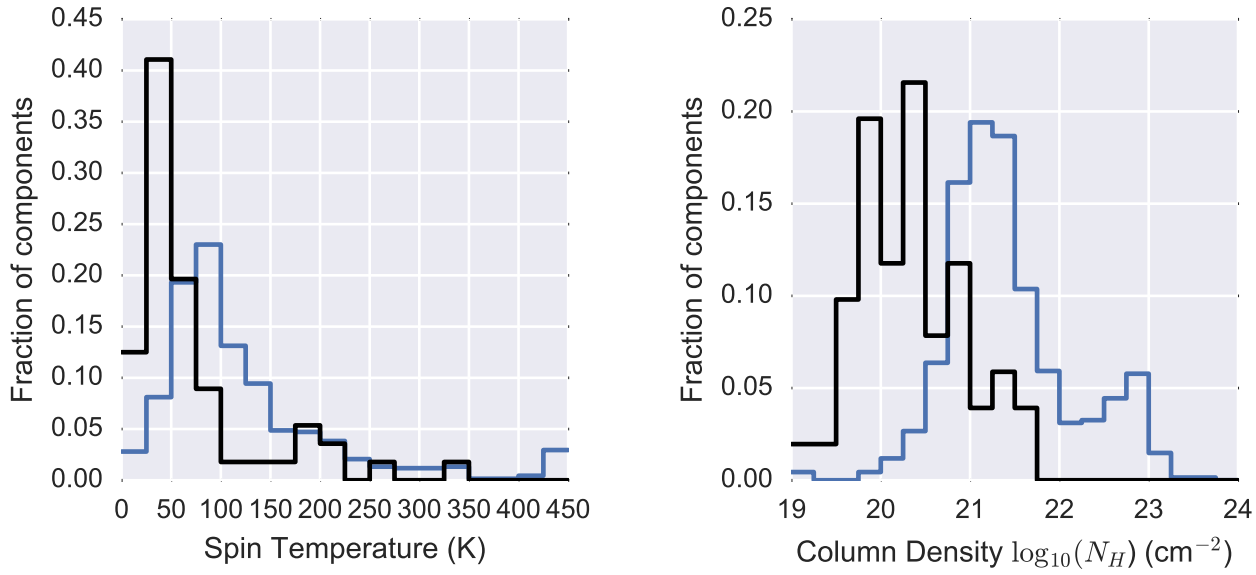

Figure 5.1: Comparison of spin temperatures and column densities between this study (blue line) and HT03 (black line). For HT03 the data is limited to the low latitude ($|l| <= 10$) CNM components

A comparison of the spin temperatures from both studies is shown in Fig. 5.1. Most of the HT03 components were in the 25 – 50 K range with a few individual components with temperatures at 150 K and above. In contrast we see a peak in spin temperature around 75 – 100 K, a steep drop to 150 K and then a long tail of small numbers of higher temperatures. This difference is also reflected in their median spin temperature of 47 K, while our median is 93 K. Using a Kolmogorov–Smirnov (K–S)

		Num Comp		$-- \sigma_\tau --$		
Name	**MAGMO Spectrum**	**SGPS**	**MAGMO**	**SGPS**	**MAGMO**	**Rating**
G326.45+0.90	MAGMOHI G326.446+0.906	6	5	0.011	0.065	A
G326.65+0.59	MAGMOHI G326.657+0.593	7	5	0.018	0.031	A
G331.26-0.19	MAGMOHI G331.277-0.190	5	1	0.029	0.115	B

Table 5.1: Comparison of sources in both the SGPS test region and MAGMO.

test to assess if these two samples of spin temperatures come from the same population of gas gives a $p = 2.7 \times 10^{-12}$ result, showing very high likelihood that they are not from the same population. Similarly, our sample is significantly denser than the HT03 low latitude sample, as shown in Fig. 5.1. Again, a K–S test shows a very high likelihood of the two samples of column densities not coming from the same population of gas ($p = 4.5 \times 10^{-11}$). Our study samples gas predominantly towards the inner Galaxy, with denser H I gas and higher opacity, while HT03 sampled gas predominantly towards the outer Galaxy where less gas is expected. However, as our spin temperature measurements are limited to a maximum spin temperature, it is expected that our values would be higher.

Comparing the turbulent Mach number statistics between the two studies shows that while the bulk of our values are less than 10, there is a long tail in our data which is not seen in HT03. As previously noted the higher Mach numbers are linked to the fitting of multiple components into a single wider component. As a result these higher Mach numbers do not reflect the physical conditions of the CNM.

5.1.2 Southern Galactic Plane Survey Test Region

Dickey et al. (2003) examined the absorption and emission spectra of 13 sources within the SGPS test region of $326 < l < 333°$. They manually fitted Gaussians to the absorption spectra, and from that obtained the characteristics of the cold gas being sampled.

Of those 13 samples, three are also found in the MAGMO data with sufficient quality for analysis. These are shown in Table 5.1 and the MAGMO spectra and their Gaussian decompositions are shown in Figs. 5.2, 5.3 and 4.4. The SGPS spectra are not shown here as the data for them is not published. All three MAGMO spectra are noisier than the SGPS spectra but exhibit the same velocity structure as measured in the SGPS survey. As Gaussian decomposition was done in both studies we can compare the categorisation of the gas. Data are not provided by Dickey *et al.* for the spectra themselves but the details of the Gaussian components are available so a direct comparison of these can be made.

A comparison of the Gaussian decomposition of the three sources in both studies is shown in Table 5.2.

The two rating A spectra show a close agreement between the two studies, with most components recovered in the automated fitting. Two of the four missed components are the shallowest components for their spectra. These are much harder to discern in the higher noise MAGMO spectra. In two cases, the MAGMO decomposition has a single wider component where Dickey et al. (2003) have two narrower components. These are at -42.79 km/s in G326.45+0.90 and at -31.7 km/s in G326.65+0.59, and in both cases the residual would be decreased if two components were fitted. Otherwise, the MAGMO components match or are slightly narrower than the SGPS ones,. The differences in τ between matching components in the two studies is within the noise limits, noting that the noise is amplified for higher τ values. Finally in G326.45+0.90, the automatic decomposition has identified a wide component, however as this has a depth within the noise limit it is not reliable.

The decomposition of the rating B spectrum of G331.26-0.19 (Fig. 4.4) is made more difficult in the higher noise but multiple components are still apparent to a human eye. Having failed to isolate any narrow components, the automatic decomposition has chosen a wide component to fit the overall

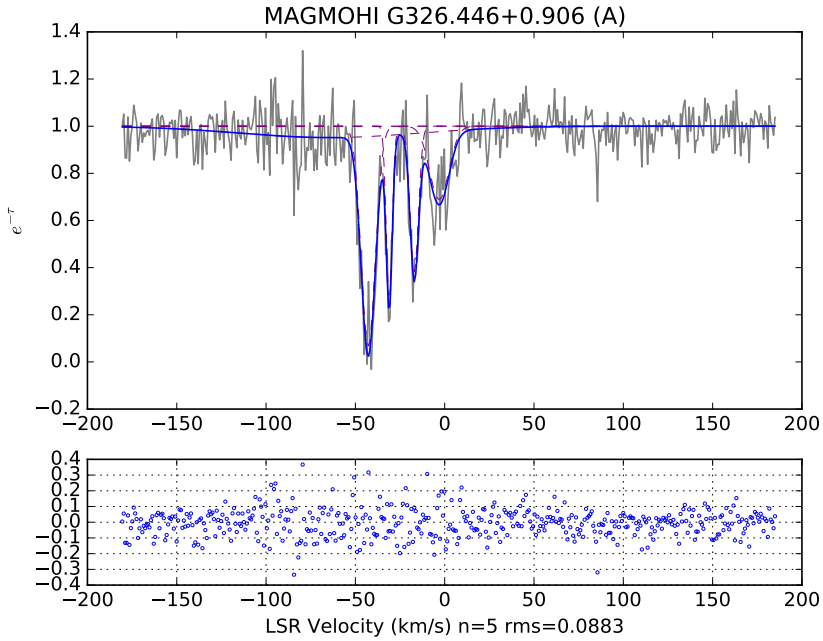

Figure 5.2: Fit for MAGMO absorption spectrum for G326.45+0.90.

absorption. While this is a reasonable match for the noisy data, as discussed earlier, physically it is difficult to explain and the far clearer SGPS data shows significant structure for this spectrum.

Overall from this small sample we can see that the decomposition of the rating A spectra has been successful enough to be usable with the caveat that wide shallow components should be treated with care. However the rating B spectra, with their higher noise, have not been adequately decomposed and again these results should be treated with care.

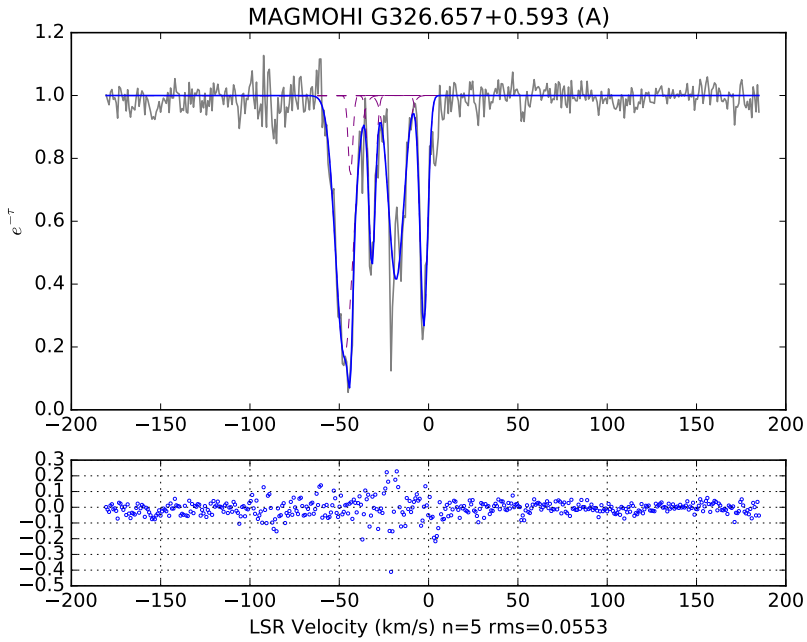

Figure 5.3: Fit for MAGMO absorption spectrum for G326.65+0.59.

Source	**SGPS Comp**	τ	**v**	σ_i	**Magmo Comp**	τ	**v**	σ
G326.45+0.90					E	0.05	-64.5	50.3
MAGMOHI G326.446+0.906	1	1.49	-43.3	3.8	A	2.69	-42.79	4.0
	2	0.87	-40.3	1.5				
	3	1.26	-31.3	2.0	B	1.33	-31.06	1.7
	4	1.12	-17.2	3.0	D	0.96	-17.02	2.3
	5	0.32	-4.5	3.8	C	0.37	-3.06	5.4
	6	0.17	2.2	2.6				
G326.65+0.59	1	0.87	-47.4	7.2	E	1.75	-47.19	4.9
MAGMOHI G326.657+0.593	2	1.06	-43.8	3.4	A	0.3	-43.74	1.3
	3	0.76	-31.3	3.3	C	0.70	-31.7	1.8
	4	1.47	-20.9	1.7				
	5	0.64	-15.7	3.7	D	0.88	-18.1	4.2
	6	1.46	-3.3	1.9	B	1.31	-2.67	2.1
	7	0.26	2.8	3.4				
G331.26-0.19	1	1.75	-89.3	4.7				
MAGMOHI G331.277-0.190	2	1.8	-63.8	7.1	A	1.36	-55.62	26.7
	3	2.23	-47.1	6.8				
	4	0.33	-22.8	7.9				
	5	0.52	-0.6	4.3				

Table 5.2: Comparison of Gaussian components in the SGPS and MAGMO studies. Each row is a single component either identified in both studies, or selected in only one.

5.1.3 Complete Atlas Of H I Absorption Toward H II Regions In The Southern Galactic Plane Survey

Brown, C et al. (2014) produced a catalogue of H I absorption and emission towards 252 Galactic H II regions. We cross-matched this catalogue with the MAGMO H I spectra resulting in spectra from 43 positions which are reproduced in both studies. The SGPS data was recorded at 0.8 km/s spectral resolution, while the MAGMO data has been rebinned from its higher resolution to 0.8 km/s to improve signal to noise ratio.. For this comparison, the MAGMO spectra were also resampled to match the velocity intervals of the SGPS data to allow a direct comparison.

We found the matches using an algorithm that selected the best MAGMO spectrum within 2.6 arcmin of the SGPS target. This range was chosen to match the SGPS beam size. We also tested examining spectra out to two beam widths (5.2 arcmin), however these more distant locations produced spectra where the velocity structure varied from the SGPS spectra, so we retained the single beam width requirement. To resample the chosen MAGMO spectra, we used the python `scipy.interpolate.interp1d` function with a linear interpolation algorithm.

An example comparison for the source OH 291.3 -0.7 is show in Fig. 5.4. The full set of 42 matching spectra is shown in Appendix A. In both sets of plots the left hand panel contains the MAGMO H I absorption spectrum plotted in light blue with the SGPS spectrum in dark blue over it. The right hand panel shows the difference between them with the MAGMO spectrum subtracted from the SGPS spectrum. The $\pm 3\sigma$ MAGMO continuum noise levels are plotted as dashed lines.

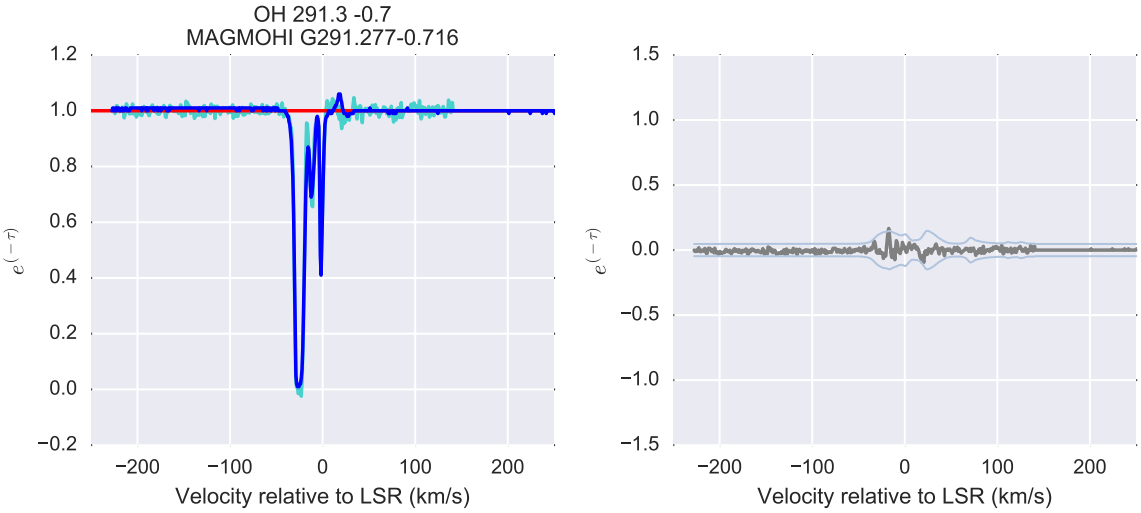

Figure 5.4: Comparison of SGPS and MAGMO absorption spectra for OH 291.3 -0.7. In the left hand panel the MAGMO spectrum, resampled to match the SGPS spectral resolution, is plotted in light blue and the SGPS spectrum in dark blue in front. On the right is the residual taken by subtracting the SGPS spectrum from the MAGMO spectrum.

Examination of the full set of spectra shows that the MAGMO spectra follow the same velocity structure as the SGPS spectra. The MAGMO data are, in general, noisier than the SGPS spectra. Of the 42 spectra, 23 matched within 3σ noise limits (allowing for less than 10 points deviation). In the other cases, the discrepancies between the two sets of spectra can be categorised as follows:

- Noise – HRDS G350.177+0.017 was quite noisy in both studies and no comparison was possible. GAL 332.54-00.11 was too noisy in SGPS to allow a detailed comparison. Four spectra (GAL 298.19-00.78, GAL 321.71+01.16, GAL 329.49+00.21, and [CH87] 324.954-0.584) had greater noise in the SGPS spectra which pushed the residuals outside the MAGMO noise limits. For each of these four sources, the spectra in both studies had the same velocity structure though.

- Wings – In two cases, GAL 298.23-00.33 and GAL 338.45+00.06, the SGPS spectrum has an absorption feature that is wide on one side however this wider feature is not present in the MAGMO spectra. In the case of GAL 338.45+00.06, MAGMO shows emission in this velocity range.
- Absorption depth – GAL 328.81+00.64 has significant noise in one saturated feature in the SGPS spectrum which takes the absorption well below zero, while this noise is not present in the MAGMO spectrum.
- Remnant emission – GAL 328.81+00.64 also shows some emission in the MAGMO spectrum.
- Sloped bandpass – In the SGPS spectrum for GAL 318.91-00.18 there is a slope to the negative velocity region before the absorption. This slope is not present in the MAGMO spectrum. This may be a bandpass problem in the SGPS data.

H II Region Name	**Qual**	**MAGMO Spectrum**	**Sep (")**	**Noise Count**	σ_τ	**Rating**
GAL 285.25-00.05	A	MAGMOHI G285.262-0.053	34	1	0.06	C
OH 291.3 -0.7	A	MAGMOHI G291.277-0.716	24	1	0.02	A
GAL 298.19-00.78	B	MAGMOHI G298.182-0.787	21	233	0.05	A
GAL 298.23-00.33	A	MAGMOHI G298.224-0.340	32	13	0.04	A
GAL 311.63+00.27	B	MAGMOHI G311.626+0.289	72	8	0.09	C
GAL 313.45+00.18	B	MAGMOHI G313.457+0.192	74	8	0.11	C
GAL 318.91-00.18	A	MAGMOHI G318.914-0.165	61	71	0.04	A
GAL 319.16-00.42	A	MAGMOHI G319.162-0.421	23	2	0.11	C
GAL 320.32-00.21	A	MAGMOHI G320.316-0.177	114	0	0.08	B
GAL 321.71+01.16	A	MAGMOHI G321.720+1.173	71	49	0.04	A
GAL 322.15+00.61	A	MAGMOHI G322.162+0.622	49	9	0.02	A
[CH87] 324.954-0.584	E	MAGMOHI G324.923-0.569	124	64	0.12	C
[CH87] 326.441-0.396	A	MAGMOHI G326.446+0.906	34	1	0.07	A
[DBS2003] 95	A	MAGMOHI G326.655+0.592	41	5	0.03	A
[CH87] 327.313-0.536	A	MAGMOHI G327.302-0.553	70	5	0.03	A
[CH87] 327.759-0.351	A	MAGMOHI G327.760-0.352	5	0	0.11	B
GAL 328.31+00.45	D	MAGMOHI G328.300+0.479	118	9	0.38	C
GAL 328.81+00.64	C	MAGMOHI G328.807+0.632	22	14	0.12	C
GAL 329.35+00.14	B	MAGMOHI G329.337+0.146	56	8	0.09	A
GAL 329.49+00.21	D	MAGMOHI G329.471+0.215	70	23	0.12	B
GAL 330.86-00.37	A	MAGMOHI G330.879-0.367	87	6	0.06	A
GAL 331.26-00.19	A	MAGMOHI G331.277-0.190	69	0	0.14	B
GAL 332.15-00.45	A	MAGMOHI G332.155-0.449	30	5	0.05	A
GAL 332.54-00.11	F	MAGMOHI G332.542-0.128	60	146	0.35	C
GAL 332.98+00.79	A	MAGMOHI G332.961+0.773	87	1	0.08	B
GAL 333.11-00.44	A	MAGMOHI G333.130-0.426	80	1	0.04	A
GAL 333.29-00.37	A	MAGMOHI G333.284-0.388	65	2	0.05	A
SNR G333.6-00.2	A	MAGMOHI G333.601-0.218	46	0	0.03	A
GAL 336.38-00.13	B	MAGMOHI G336.359-0.137	58	1	0.16	C
GAL 337.15-00.18	A	MAGMOHI G337.120-0.175	97	3	0.04	A
GAL 337.95-00.48	A	MAGMOHI G337.927-0.458	103	1	0.08	A
GAL 338.41-00.24	A	MAGMOHI G338.405-0.203	128	1	0.07	A

GAL 338.45+00.06	A	MAGMOHI G338.435+0.058	52	30	0.05	B
GAL 338.94+00.60	A	MAGMOHI G338.922+0.624	105	1	0.23	C
GAL 340.05-00.25	C	MAGMOHI G340.051-0.235	69	1	0.21	C
GAL 340.28-00.22	A	MAGMOHI G340.276-0.209	49	0	0.21	C
GAL 340.78-01.01	A	MAGMOHI G340.788-1.019	58	2	0.04	A
GAL 348.72-01.03	A	MAGMOHI G348.725-1.043	57	4	0.05	A
GAL 350.13+00.09	A	MAGMOHI G350.101+0.081	102	3	0.05	A
HRDS G350.177+0.017	D	MAGMOHI G350.180+0.014	15	20	0.14	C
[KC97c] G350.3+00.1	E	MAGMOHI G350.330+0.099	32	6	0.23	C
GAL 351.36+00.67	A	MAGMOHI G351.371+0.655	61	0	0.04	A

Table 5.3: Comparison of SGPS and MAGMO spectra against H II regions.

5.2 Gas

Around 15% of the gas components (99 out of 678 components with spin temperatures) have spin temperatures above 200 K, the expected maximum temperature for the CNM (Wolfire et al. 2003). Of these, 11 are from the set with FWHM > 50 km/s and, as discussed in Section 4.2.1, are likely to be unrepresentative of the actual gas. These 11 also make up the majority of the group with optical depths very close or within the noise level of the spectra, as discussed in Section 4.2.2.

Some of the hottest gas is near $l = -8$, in the direction of NGC 6334, (field G351.242+0.670) a H II region and one of the more complex regions observed. In these regions we would expect significant turbulence and this may have heated the gas. Likewise, there are clusters of hotter gas around the H II regions GAL 322.16+00.62, IRAS 16506-4512 and NGC 3581.

Otherwise, the hotter gas is generally grouped around the low velocities, $-50 < v < 20$, regions where we expect to have significant overlap in the gas from the near and far instances of a particular velocity. This can lead to overlapping components which will boost the optical depth and can be difficult to separate.

Also of note is that 39 gas components have $\mathrm{T}_{kmax} <$T_S which is not expected (Heiles & Troland 2003b). Of these, 10 have T_{kmax} values which fall inside the uncertainties in spin temperature. Another 12 have a FWHM < 1 km/s which is most likely noise that has been fitted. For the remainder, it is most likely that the spin temperature has been overestimated due to the way in which we are sampling the emission to obtain the brightness temperature.

5.2.1 Star Forming Regions

As noted earlier, the MAGMO observations targeted high mass star-forming regions. In these regions it is reasonable to expect there will be significant reserves of cold gas. To test this hypothesis, we divided the the spin temperature population into two sub-populations, one for gas of sources within two arcmin of the maser and in velocity ranges within 10 km/s of the maser range, and the other for the rest of the gas. The comparison of these two sub-populations is shown in Fig. 5.5.

Using a Kolmogorov–Smirnov test to assess if these two subpopulations of spin temperatures come from the same population of gas gives a $p = 0.02$ result, showing that they are likely to be different populations. The near maser gas is slightly colder with a median spin temperature of 83.8 K as compared to the 96.3 K for the rest of the gas. Likewise the optical depths (τ) of the two sub-populations are highly likely to not be from the same population with a $p = 0.00$ result. As shown in Fig. 5.5 there is a substantially higher proportion of high optical depth components in the near maser sub-population, likely associated with high column density gas clouds.

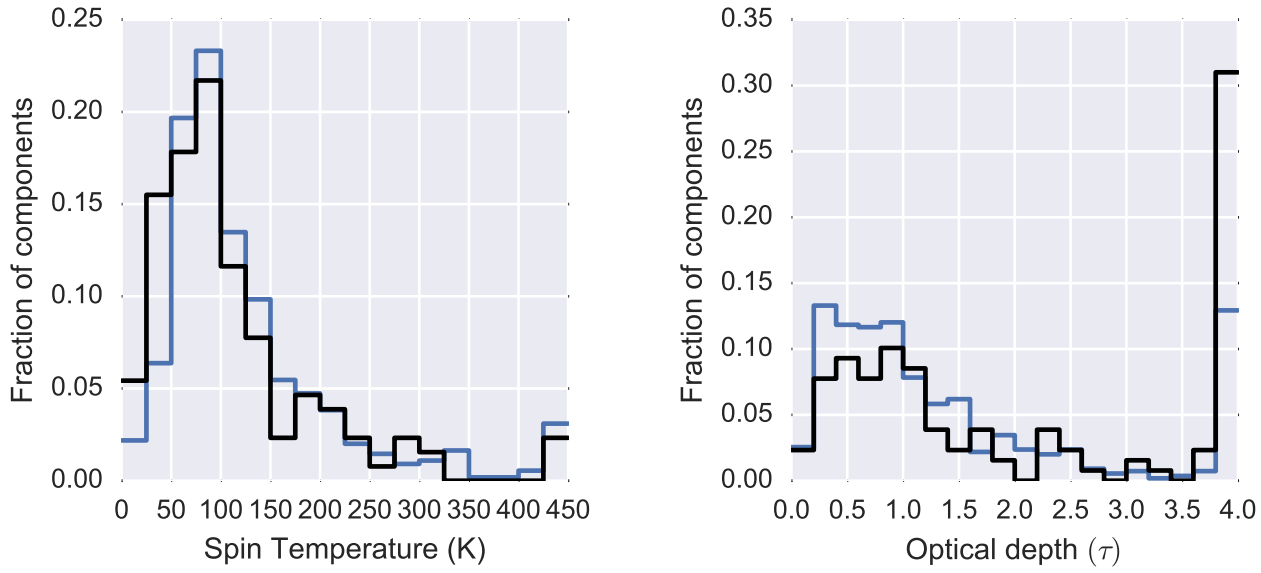

Figure 5.5: Comparison of gas away from (blue line) and near to (black line) the methanol masers in each observed field. Near is defined as within 2 arcmin and ±10 km/s of the maser. Left: Comparison of spin temperatures. Right: Comparison of optical depth (τ).

5.3 Assessment of the Automated Processing

The automated processing in this study has been successful. The source finding and spectra extraction has produced results comparable to other surveys. In particular the weighted combination of spectra from pixels within the identified source component has produced spectra that have compensated for the finer spectral resolution.

The decomposition has successfully identified components for spectra where there were limited or no blended components and where the spectra have a high signal to noise level. However it had difficulty in decomposing blended components and often chose a single component where an expert might have fitted two or more components. Similarly, where there was significant noise the decomposition sometimes failed to correctly extract components. This may be possible to alleviate in the training process, as an experimentally chosen configuration produced better decomposition for many spectra, indicating that the problem is likely in the configuration and that `GaussPy` may be able to manage these spectra.

Despite this, the decomposition of the rating A spectra was accurate enough to obtain a usable model of the cold gas. The pipeline was able to produce good results which reproduce many of the characteristics of previous studies and provide insight into the previously unsampled regions.

The pipeline also proved to be quite capable of being able to be reliably rerun as the processing was refined and improved.

5.4 Applicability to GASKAP

One of the aims for this study was to trial technologies and techniques for use in processing large H I absorption surveys such as GASKAP. The absorption data for GASKAP will be extracted from the non-continuum subtracted spectral cubes. In some ways these are the easiest to automate as they do not require combining with the zero-spacing data such as from the Parkes radio telescope. There is still the need to access the emission data for spin temperature calculation though, which will require zero spacing data.

Much of the processing of the GASKAP survey will be automated by the ASKAP observatory. The data reduction up to an including the production of spectral cubes will be automated and implemented in ASKAPsoft (Cornwell et al. 2011). Source finding and extraction of absorption spectra may be done either within the ASKAPsoft pipeline or outside of it, depending on the project team's requirements. The decomposition of the spectra into Gaussian components and the subsequent analysis will be done by the project team.

Based on this breakdown of the processing responsibilities, the `analyse_data`, `analyse_spectra`, `decompose` and `examine_gas` steps could all be used as starting points for the GASKAP team's processing of the absorption data. As has been discussed, while the spectra extraction has produced good results, the decomposition still needs improvement. Some of that work is already underway within the gas group at RSAA, with investigation of training techniques and simultaneous fitting of both the absorption and emission spectra already occurring.

It is expected that the GASKAP will take 13,000 optical depth measurements within the Galactic plane and 1,000 in the Magellanic system (Dickey et al. 2013). Hence GASKAP H I absorption will be a 17 fold increase in size from the MAGMO-HI sample. Such an increase would mean that more powerful processing resources will be needed for even the combined analysis. However, given the low level of computing resources used for the later stages of processing, it should not be essential to exploit highly parallel compute techniques for successful processing.

Chapter 6

Conclusion

In this study we have taken the H I observations from the MAGMO survey and extracted absorption spectra from all sufficiently bright background sources. We have extracted 819 spectra towards unique sources, and selected the 238 with the highest quality for further analysis. For each of these spectra we automatically decomposed them into Gaussian components and then examined the physical properties of these components, measuring central velocity, velocity width, and optical depth and calculating the spin temperature, column density, turbulent Mach number and cold gas fraction for each component.

We examined the physical properties by comparing the overall population statistics to previous population studies. While noting that we have measured only the maximum spin temperatures, through these comparisons we found that the gas sampled by MAGMO was significantly warmer and denser than those towards the outer Galaxy.

To verify the processing we compared spectra towards specific sources that had also been analysed from the SGPS observations. We found that the spectra were accurately extracted, showing the same velocity structure, with good exclusion of emission but generally a higher noise level. Comparing the decomposition found that while many of the major features were correctly identified, some features were missed and overlapping and saturated spectra had room for improvement.

6.1 Future Work

Given the restricted time-frame of an honours project, there is much work that can still be done that would improve the scientific results.

A major area of future improvement is in the automated decomposition. Firstly we can identify groups of spectra which need separate configuration for `Gausspy` such as the noisier rating B spectra and spectra in particularly crowded regions. We can improve the training by manually decomposing spectra from each identified group and using these to create a series of training datasets for each group. The decomposition should be extended to the emission spectra and provide a full implementation of the Heiles & Troland (2003a) two-phase decomposition and fitting technique to obtain the spin temperatures. Work is currently underway to implement the simultaneous fitting in `Gausspy` and when complete this could be easily adapted.

The range of data which could be fully analysed was limited by the availability of SGPS emission data. Not all of the SGPS data was available for this study, so adding that extra SGPS data would allow expansion of the region analysed. There are also a limited number of observations outside the region observed by SGPS. For these we could use the GASS data, however this would be more complex as, as highlighted in Dénes et al. (2017), the spin temperatures obtained from the two datasets can be significantly different.

Finally, as outlined in Section 5.4 this work can be extended to form the basis of a pipeline for the GASKAP absorption observations.

Bibliography

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, Astronomy and Astrophysics, 558, A33

Begum, A., Stanimirović, S., Goss, W. M., et al. 2010, The Astrophysical Journal, 725, 1779

Benjamin, R. A. 2017, Bob Benjamin's Home Page, `https://wisp.physics.wisc.edu/benjamin/index.html`, accessed: 2017-09-25

Brown, C, Dickey, J M, Dawson, J R, & McClure-Griffiths, N M. 2014, The Astrophysical Journal Supplement Series, 211, 29

Burton, W. B. 1988, in Galactic and Extragalactic Radio Astronomy, ed. K. I. Kellermann & G. L. Verschuur (Berlin and New York), 295–358

Caswell, J. L., Murray, J. D., Roger, R. S., Cole, D. J., & Cooke, D. J. 1975, Astronomy and Astrophysics, 45, 239

Clark, B. G. 1965, Astrophysical Journal, 142, 1398

Cornwell, T., Humphreys, B., Lenc, E., Voronkov, M., & Whiting, M. 2011, ASKAP Science Processing, Tech. Rep. ASKAP-SW-0020, CSIRO

Dénes, H., McClure Griffiths, N. M., Dickey, J. M., Dawson, J. R., & Murray, C. E. 2017, Manuscript submitted for publication.

Dickey, J. M., Brinks, E., & Puche, D. 1992, Astrophysical Journal, 385, 501

Dickey, J. M., Kulkarni, S. R., Heiles, C. E., & van Gorkom, J. H. 1983, The Astrophysical Journal Supplement Series, 53, 591

Dickey, J. M. & Lockman, F. J. 1990, Annual Review of Astronomy and Astrophysics, 28, 215

Dickey, J. M., McClure-Griffiths, N., Gibson, S. J., et al. 2013, Publ. Astron. Soc. Aust., 30, e003

Dickey, J. M., McClureGriffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, The Astrophysical Journal, 585, 801

Dickey, J. M., Mebold, U., Stanimirović, S., & Staveley-Smith, L. 2000, The Astrophysical Journal, 536, 756

Dickey, J. M., Terzian, Y., & Salpeter, E. E. 1978, Astrophys. J. Suppl. Ser., 36, 77

Fernández, X., Gim, H. B., van Gorkom, J. H., et al. 2016, The Astrophysical Journal Letters, 824, L1

Green, J., Caswell, J., McClure-Griffiths, N., Harvey-Smith, L., & Robishaw, T. 2010, MAGMO: Mapping the Galactic Magnetic field through OH masers, ATNF Proposal

Green, J. A., Caswell, J. L., Fuller, G. A., et al. 2012a, Monthly Notices of the Royal Astronomical Society, 420, 3108

Green, J. A., McClure-Griffiths, N. M., Caswell, J. L., Robishaw, T., & Harvey-Smith, L. 2012b, Monthly Notices of the Royal Astronomical Society, 425, 2530

Gunn, J. E., Knapp, G. R., & Tremaine, S. D. 1979, Astronomical Journal, 84, 1181

Hagen, J. P., Lilley, A. E., & McClain, E. F. 1955, Astrophys. J., 122, 361

Hancock, P. J., Murphy, T., Gaensler, B. M., Hopkins, A., & Curran, J. R. 2012, Monthly Notices of the Royal Astronomical Society, 422, 1812

Hartmann, D. & Burton, W. B. 1999, VizieR On-line Data Catalog, 8054

Heiles, C. & Troland, T. H. 2003a, The Astrophysical Journal Supplement Series, 145, 329

Heiles, C. & Troland, T. H. 2003b, The Astrophysical Journal, 586, 1067

HI4PI Collaboration, Ben Bekhti, N., Flöer, L., et al. 2016, Astronomy & Astrophysics, 594, A116

Hopkins, A. M., Whiting, M. T., Seymour, N., et al. 2015, Publ. Astron. Soc. Pacific, astro-ph.IM, e037

Ivezić, Z., Lupton, R. H., Schlegel, D., et al. 2004, Astronomische Nachrichten, 325, 583

Johnston, S. & Gray, A. D. 2006, SKA Memo 72: Surveys with the xNTD and CLAR, Tech. rep.

Jones, C. & Dickey, J. M. 2012, The Astrophysical Journal, 753, 62

Kalberla, P. M. & Kerp, J. 2009, Annual Review of Astronomy and Astrophysics, 47, 27

Kalberla, P. M. W. & Dedes, L. 2008, Astronomy and Astrophysics, 487, 951

Kim, C.-G., Ostriker, E. C., & Kim, W.-T. 2014, The Astrophysical Journal, 786, 64

Kulkarni, S. R. & Heiles, C. 1988, Neutral hydrogen and the diffuse interstellar medium, 2nd edn. (Berlin and New York: Springer-Verlag), 95–153

Lee, M.-Y., Stanimirović, S., Murray, C. E., Heiles, C., & Miller, J. 2015, The Astrophysical Journal, 809, 56

Levine, E. S., Blitz, L., & Heiles, C. 2006, Science, 312, 1773

Lindner, R. R., Vera-Ciro, C., Murray, C. E., et al. 2015, The Astronomical Journal, 149, 138

Lockman, F. J. & McClure-Griffiths, N. M. 2016, The Astrophysical Journal, 826, 215

McClure-Griffiths, N. M. & Dickey, J. M. 2016, The Astrophysical Journal, 831, 124

McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., & Green, A. J. 2004, The Astrophysical Journal, 607, 11

McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., et al. 2005, The Astrophysical Journal Supplement Series, 158, 178

McClure-Griffiths, N. M., Pisano, D. J., Calabretta, M. R., et al. 2009, The Astrophysical Journal Supplement Series, 181, 398

Minier, V., Ellingsen, S. P., Norris, R. P., & Booth, R. S. 2003, Astronomy and Astrophysics, 403, 1095

Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2015, The Astrophysical Journal, 804, 89

Murray, C. E., Stanimirović, S., Kim, C.-G., et al. 2017, ApJ, 837, 55

Ochsenbein, F., Taylor, M., Williams, R., et al. 2013, IVOA Recommendation 20 September 2013

Roy, N., Kanekar, N., & Chengalur, J. N. 2013, arXiv.org, 2366

Sault, R. J., Teuben, P. J., & Wright, M. C. H. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, 433

Sofue, Y. 2017, arXiv.org, arXiv:1706.08771

Sparke, L. S. & Gallagher III, J. S. 2007, Galaxies in the Universe: An Introduction, 2nd edn., Galaxies Universe An Introd. Ed. Linda S.˜Sparke John S.˜Gallagher, III.˜ISBN-13 978-0-521-85593-8 (HB); ISBN-13 978-0-521-67186-6 (PB).˜Published by Cambridge Univ. Press. Cambridge, UK, 2007. (Cambridge, UK: Cambridge University Press)

Stanimirović, S., Murray, C. E., Lee, M.-Y., Heiles, C., & Miller, J. 2014, The Astrophysical Journal, 793, 132

Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, arXiv.org, 1158

Strasser, S. & Taylor, A. R. 2004, The Astrophysical Journal, 603, 560

Strasser, S. T., Dickey, J. M., Taylor, A. R., et al. 2007, The Astronomical Journal, 134, 2252

Taylor, A. R., Gibson, S. J., Peracaula, M., et al. 2003, The Astronomical Journal, 125, 3145

Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, Astronomy and Astrophysics, 556, A2

Whiting, M. & Humphreys, B. 2012, Publ. Astron. Soc. Aust., 29, 371

Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, The Astrophysical Journal, 587, 278

Wu, C., Tobar, R., Vinsen, K., et al. 2017, Astronomy and Computing, 20, 1

Comparison Plots for H II Regions in the Southern Galactic Plane Survey

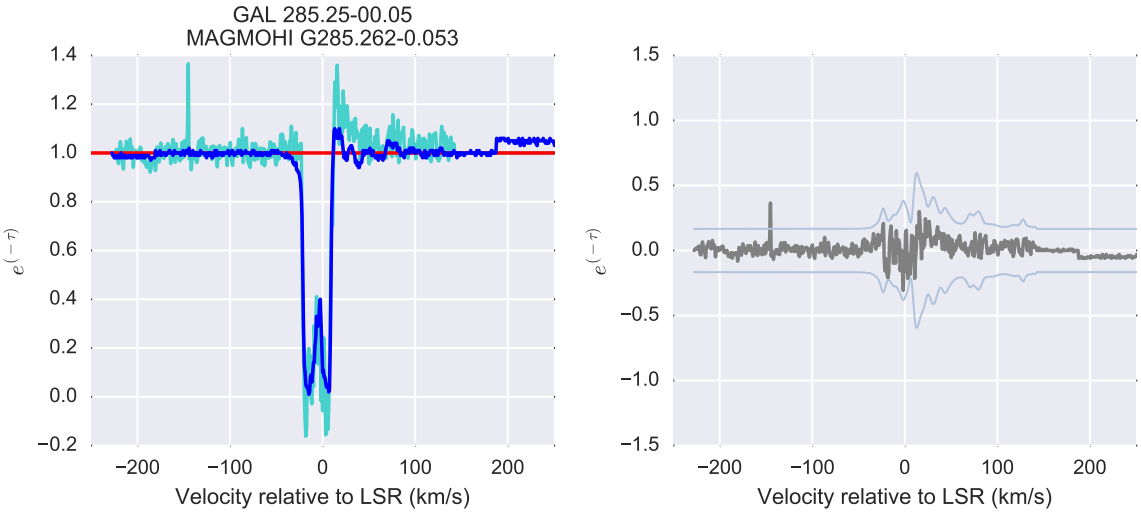

Figure A.1: Comparison of MAGMO and SGPS absorption spectra for GAL 285.25-00.05

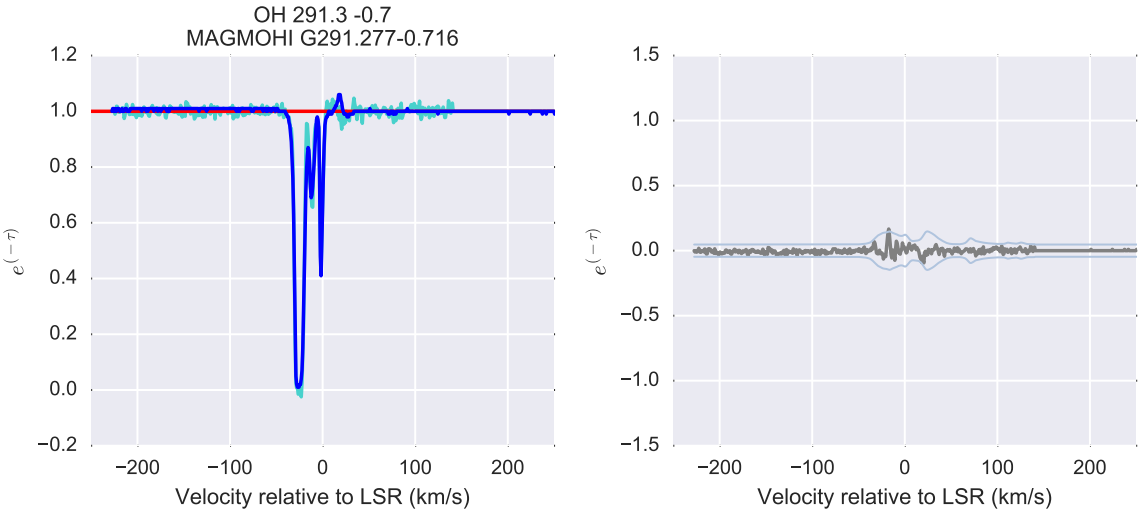

Figure A.2: Comparison of MAGMO and SGPS absorption spectra for OH 291.3 -0.7

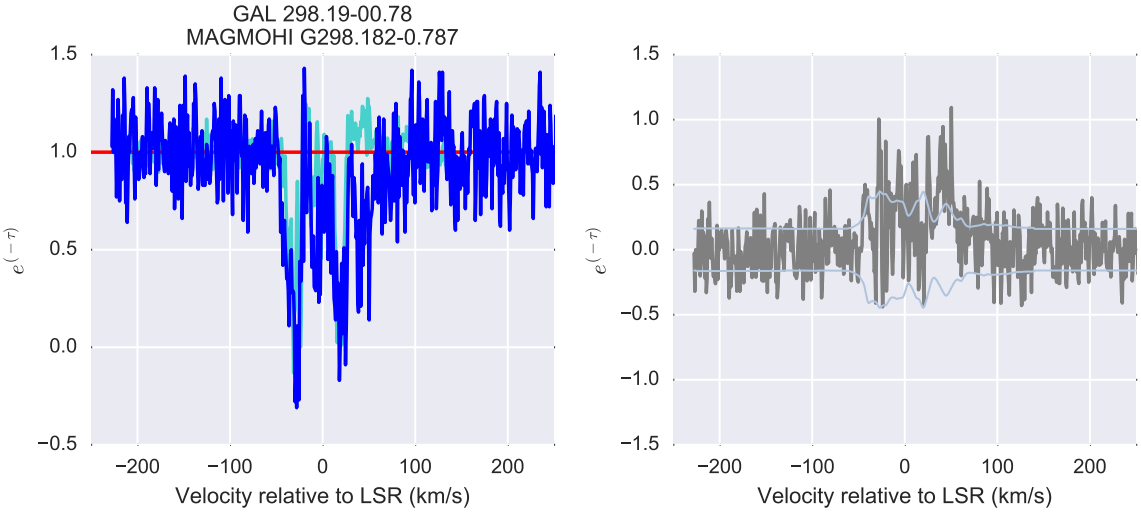

Figure A.3: Comparison of MAGMO and SGPS absorption spectra for GAL 298.19-00.78

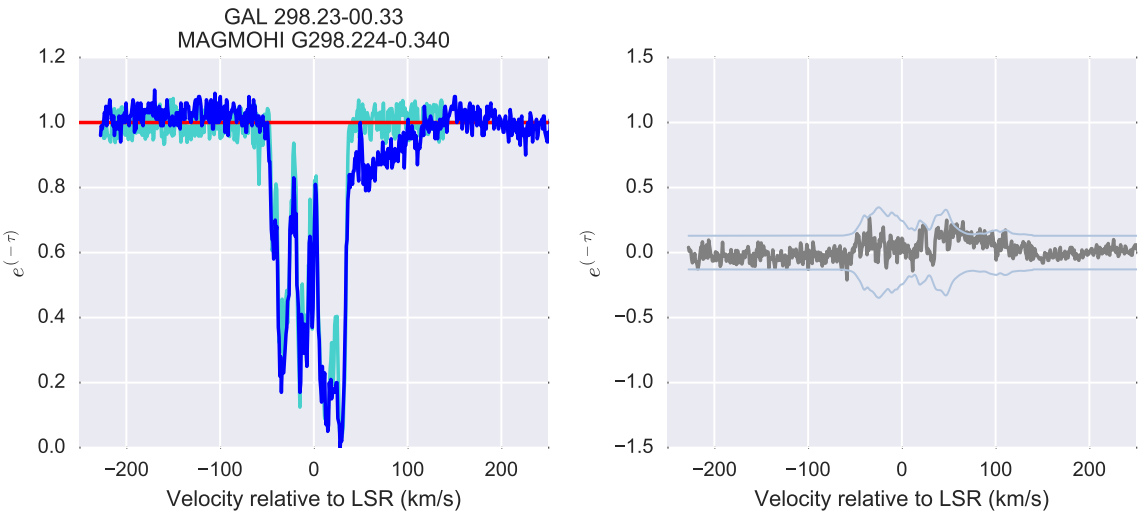

Figure A.4: Comparison of MAGMO and SGPS absorption spectra for GAL 298.23-00.33

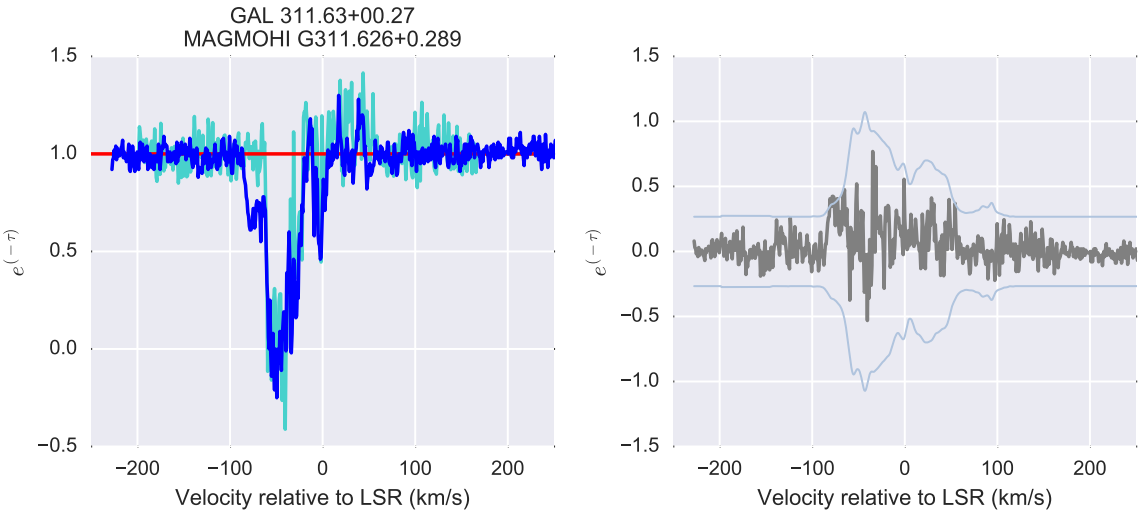

Figure A.5: Comparison of MAGMO and SGPS absorption spectra for GAL 311.63+00.27

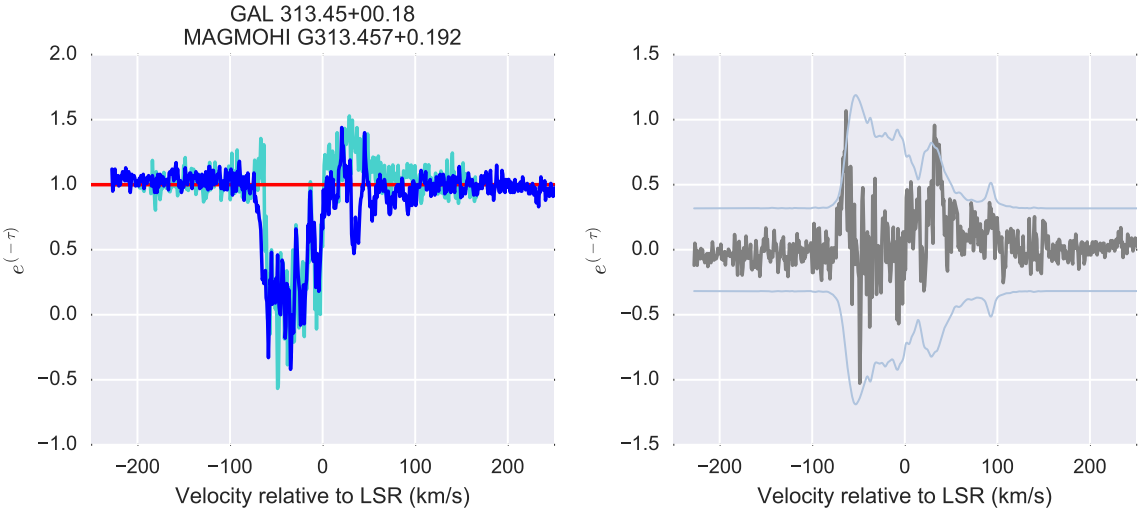

Figure A.6: Comparison of MAGMO and SGPS absorption spectra for GAL 313.45+00.18

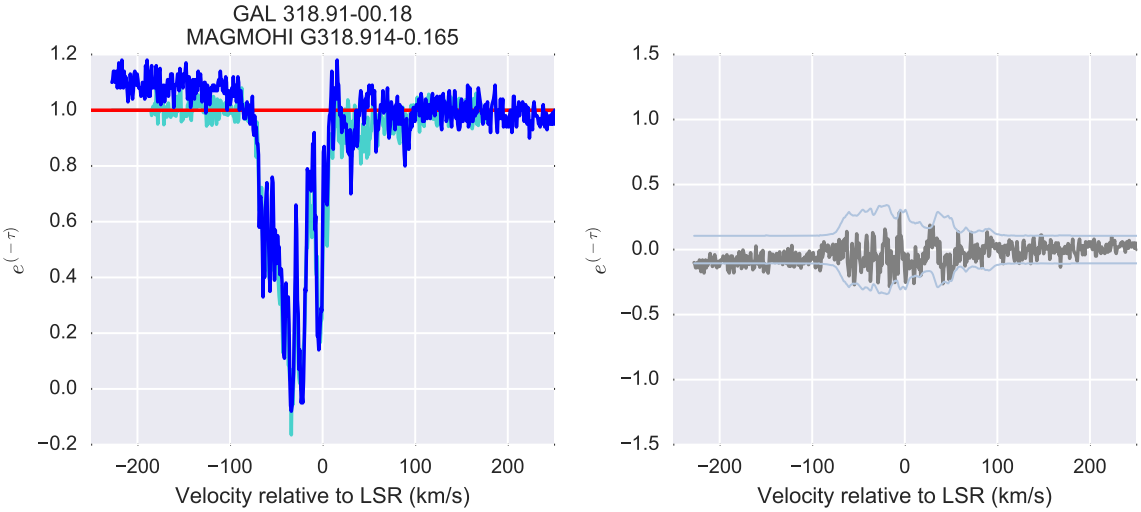

Figure A.7: Comparison of MAGMO and SGPS absorption spectra for GAL 318.91-00.18

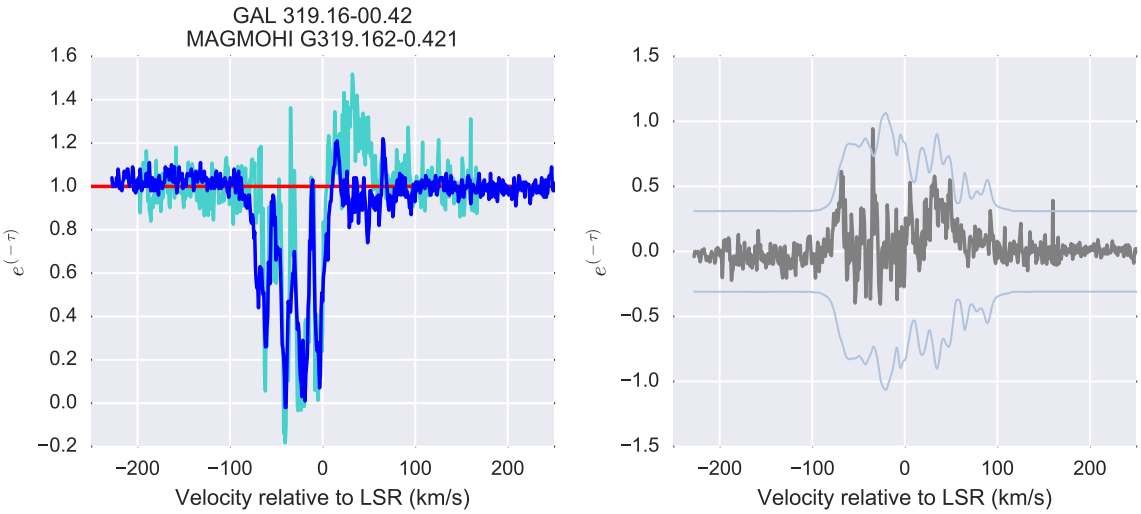

Figure A.8: Comparison of MAGMO and SGPS absorption spectra for GAL 319.16-00.42

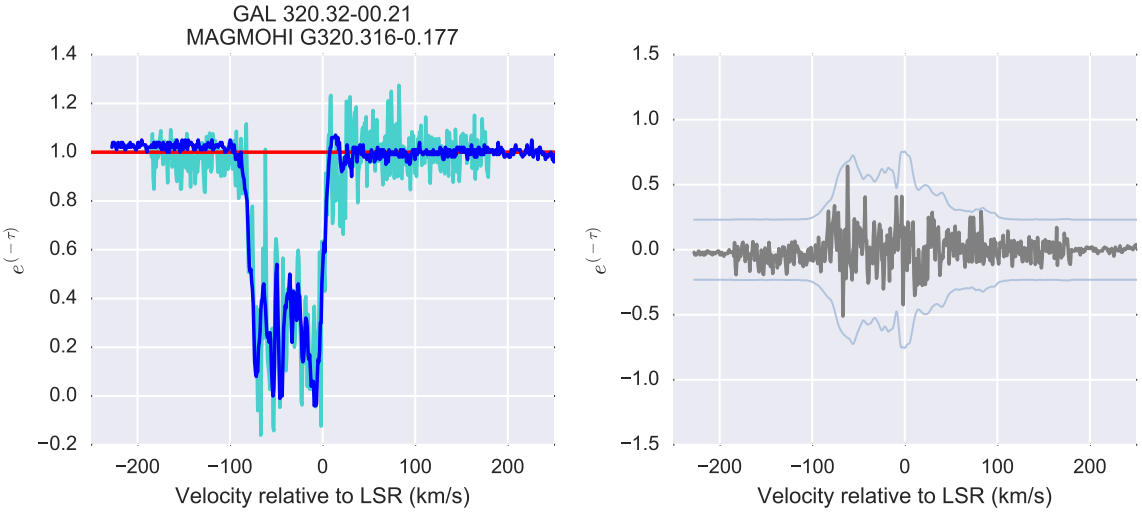

Figure A.9: Comparison of MAGMO and SGPS absorption spectra for GAL 320.32-00.21

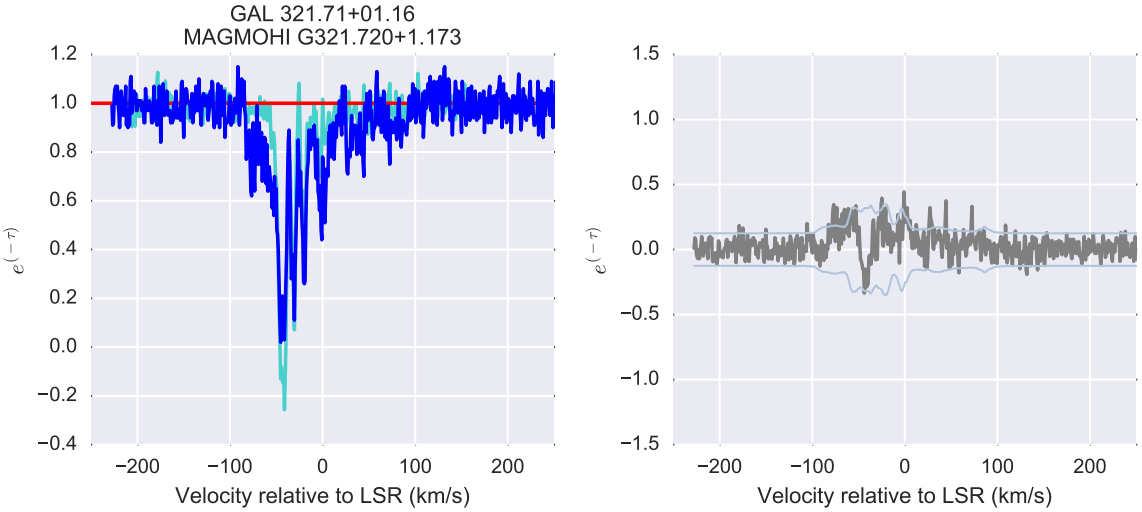

Figure A.10: Comparison of MAGMO and SGPS absorption spectra for GAL 321.71+01.16

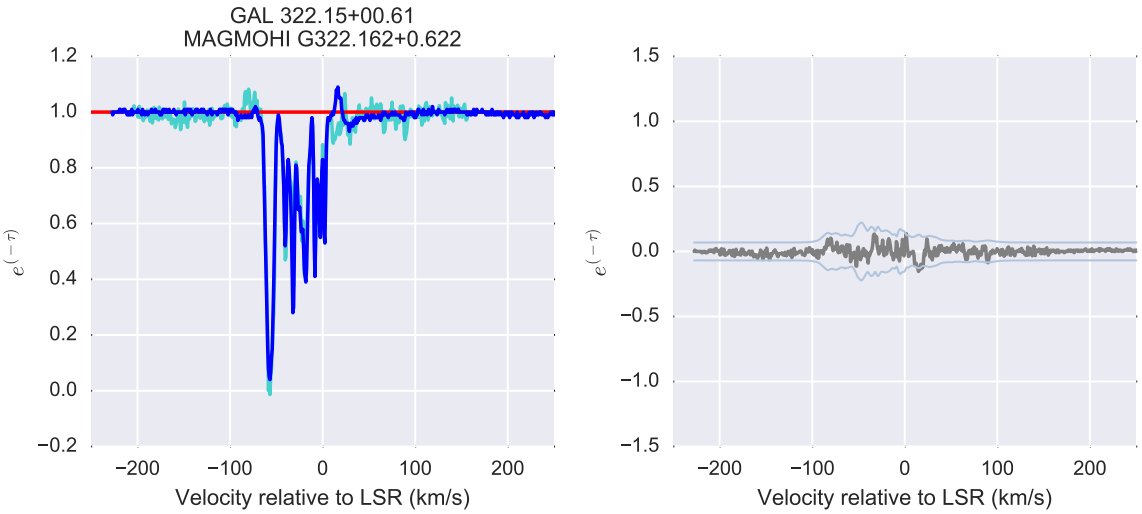

Figure A.11: Comparison of MAGMO and SGPS absorption spectra for GAL 322.15+00.61

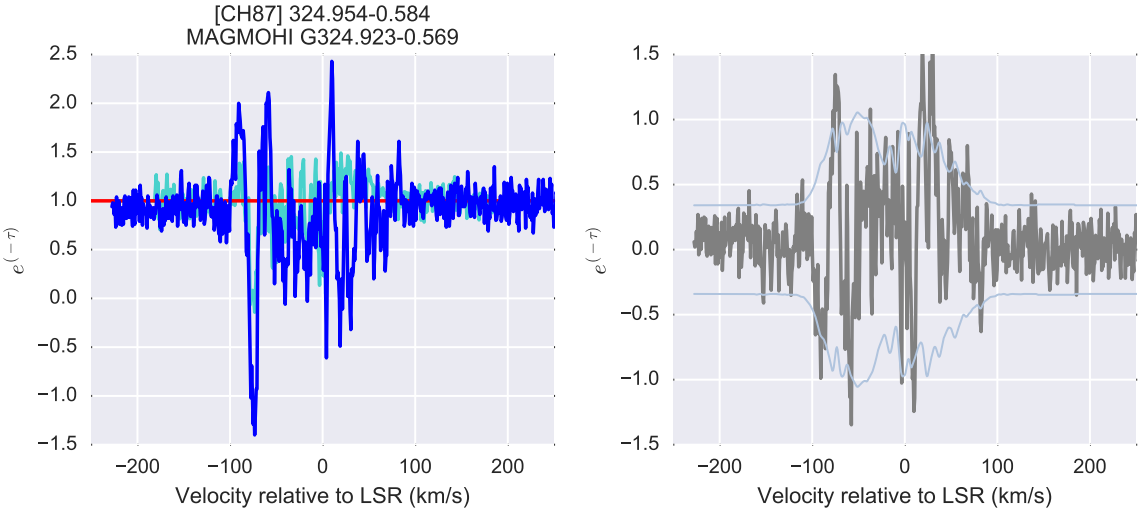

Figure A.12: Comparison of MAGMO and SGPS absorption spectra for [CH87] 324.954-0.584

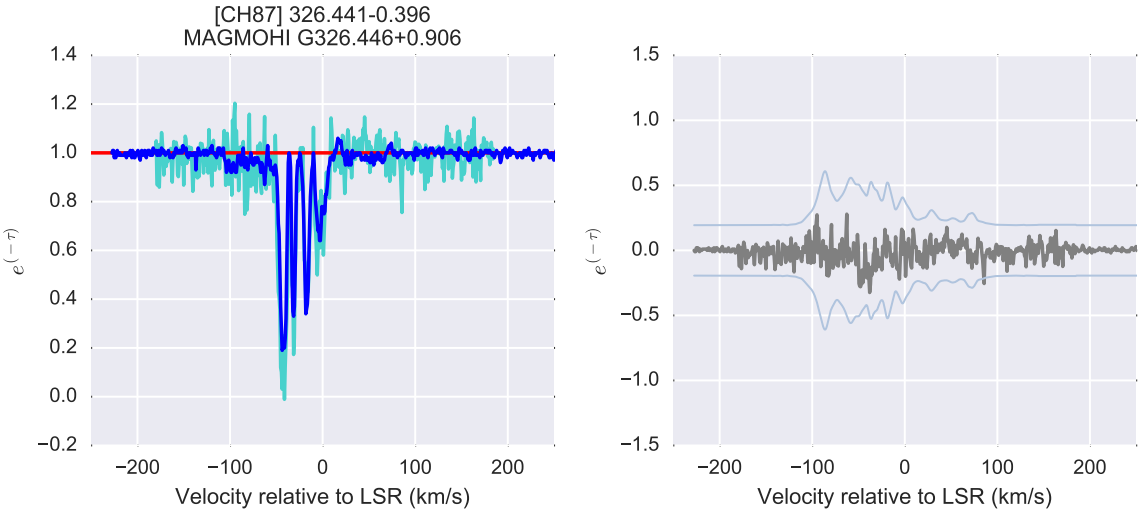

Figure A.13: Comparison of MAGMO and SGPS absorption spectra for [CH87] 326.441-0.396

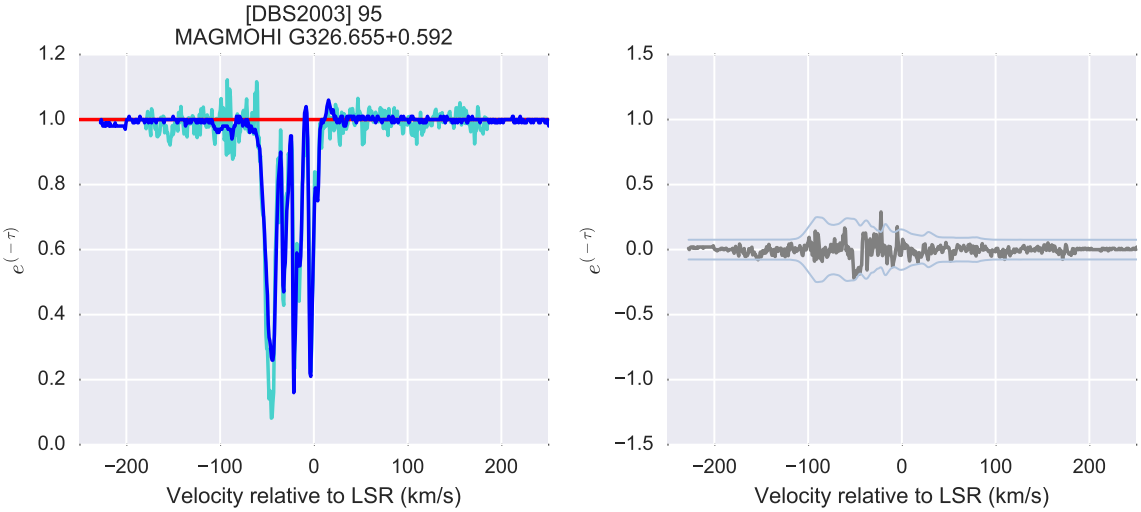

Figure A.14: Comparison of MAGMO and SGPS absorption spectra for [DBS2003] 95

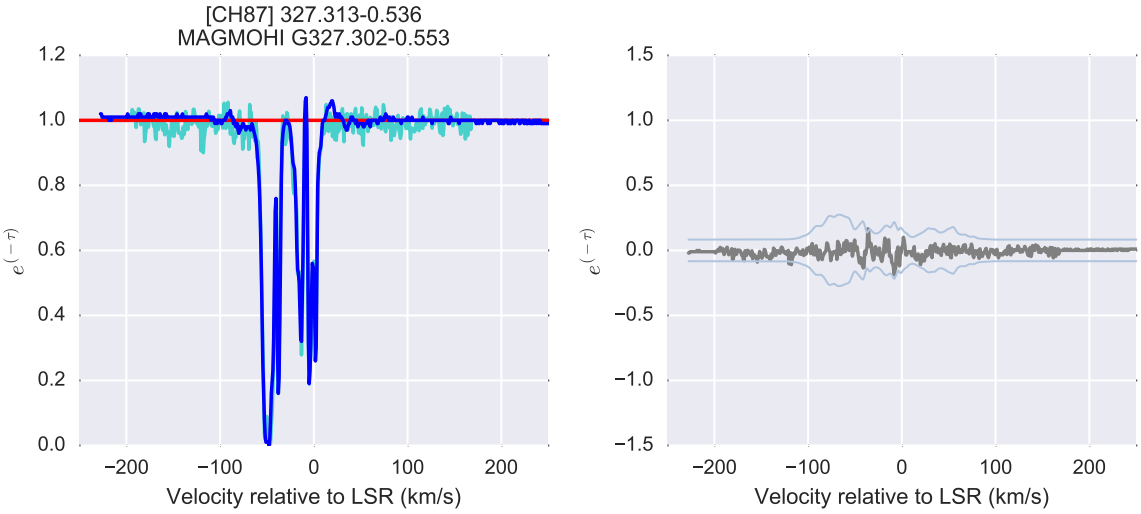

Figure A.15: Comparison of MAGMO and SGPS absorption spectra for [CH87] 327.313-0.536

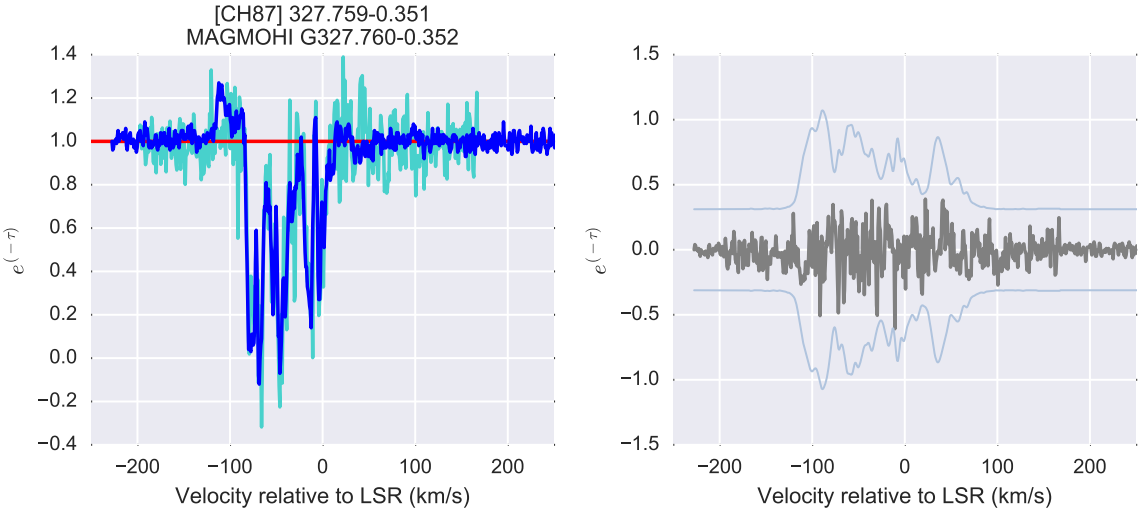

Figure A.16: Comparison of MAGMO and SGPS absorption spectra for [CH87] 327.759-0.351

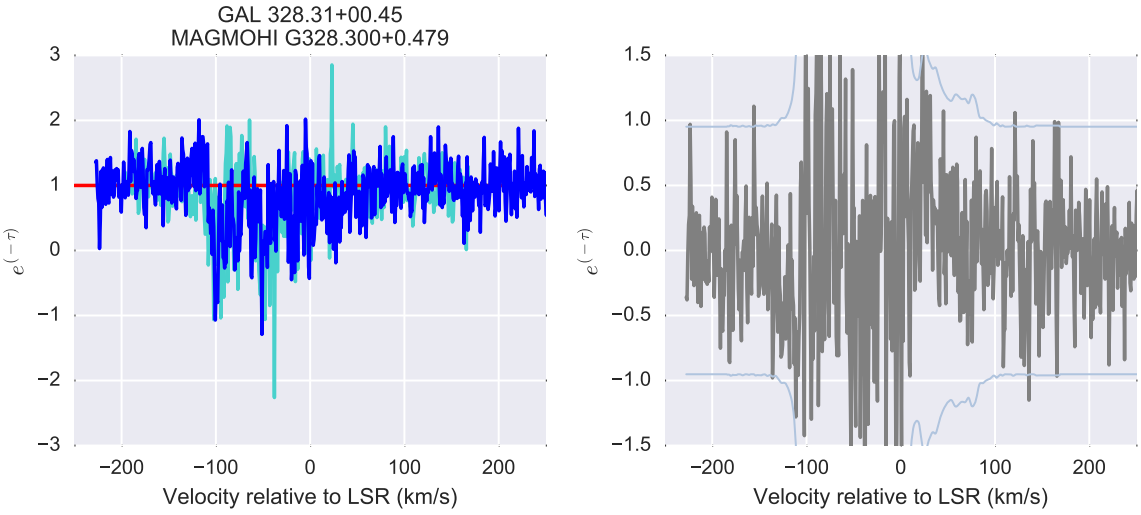

Figure A.17: Comparison of MAGMO and SGPS absorption spectra for GAL 328.31+00.45

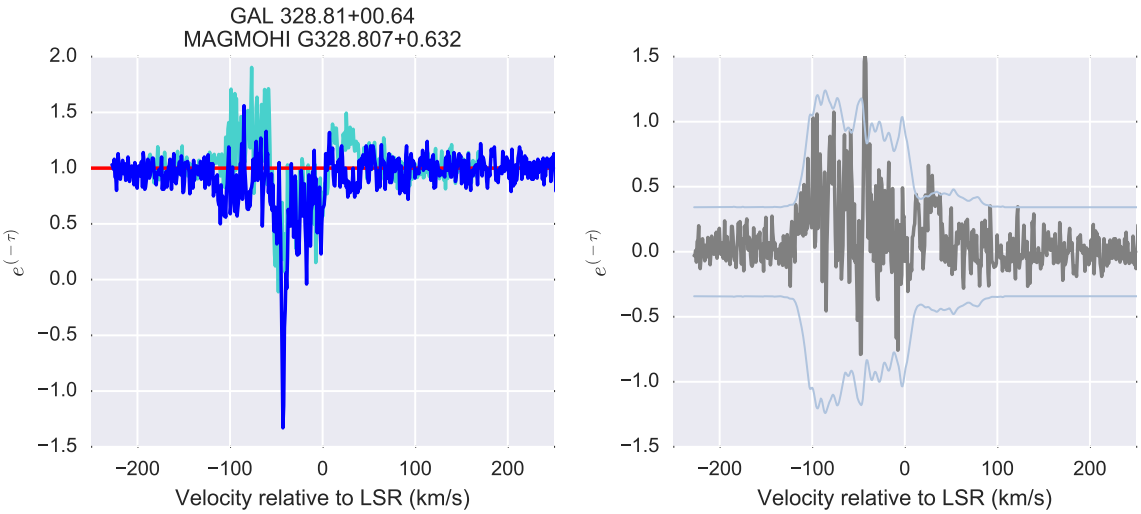

Figure A.18: Comparison of MAGMO and SGPS absorption spectra for GAL 328.81+00.64

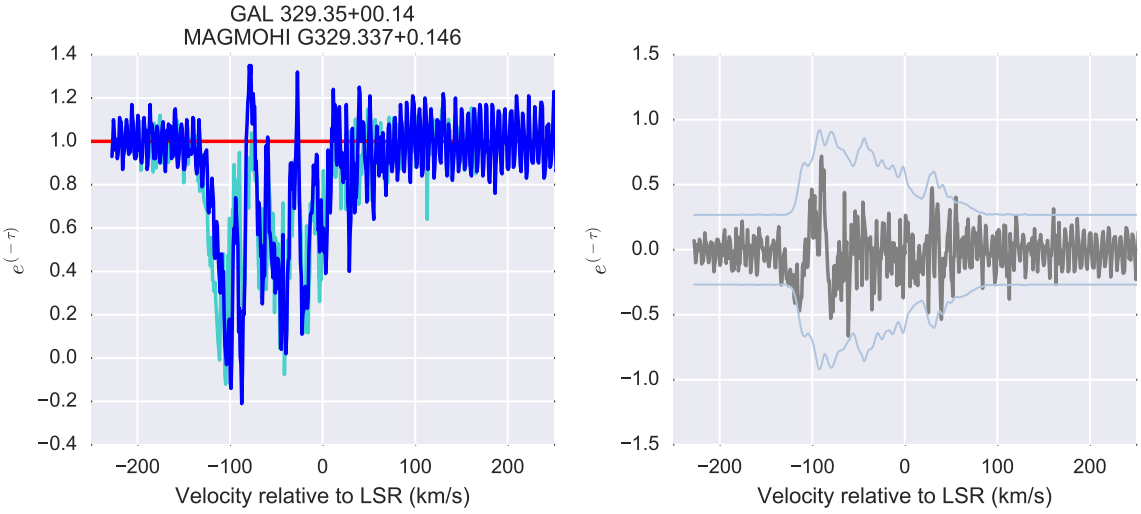

Figure A.19: Comparison of MAGMO and SGPS absorption spectra for GAL 329.35+00.14

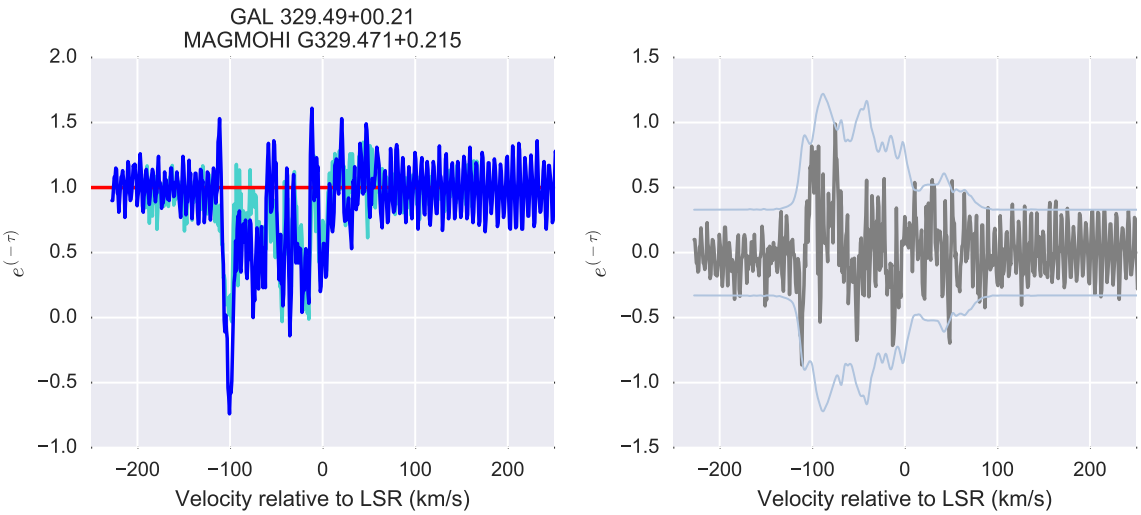

Figure A.20: Comparison of MAGMO and SGPS absorption spectra for GAL 329.49+00.21

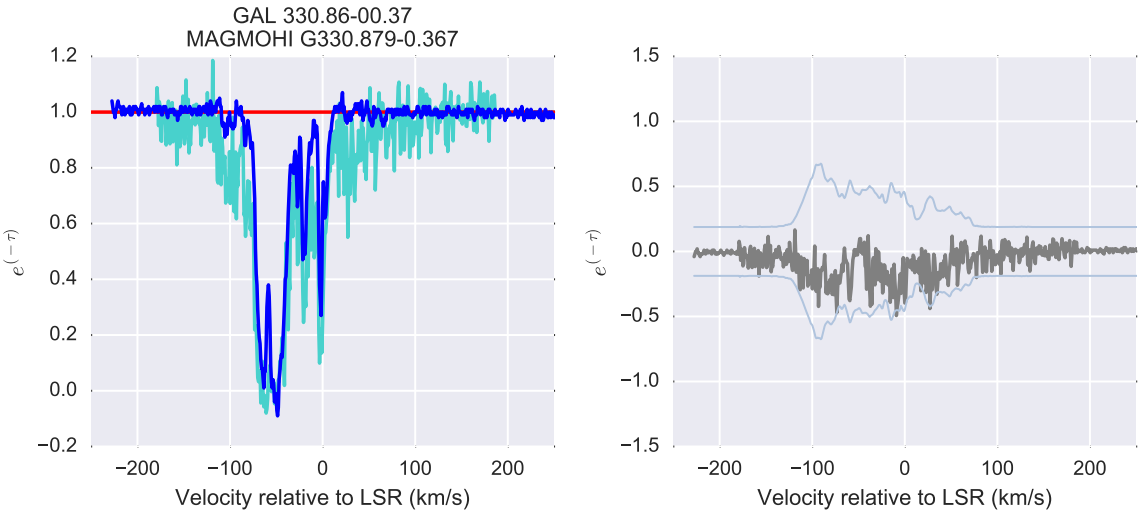

Figure A.21: Comparison of MAGMO and SGPS absorption spectra for GAL 330.86-00.37

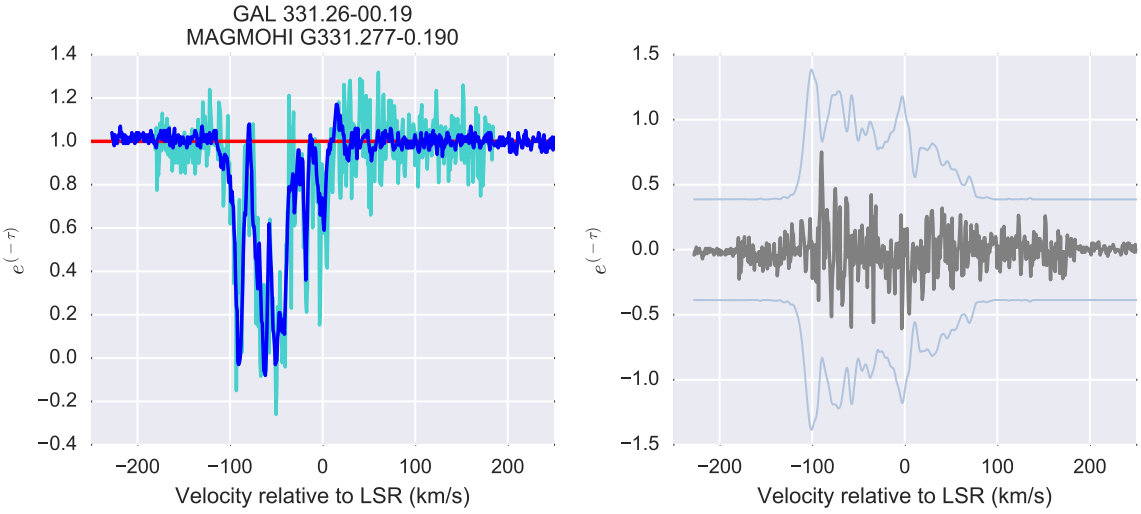

Figure A.22: Comparison of MAGMO and SGPS absorption spectra for GAL 331.26-00.19

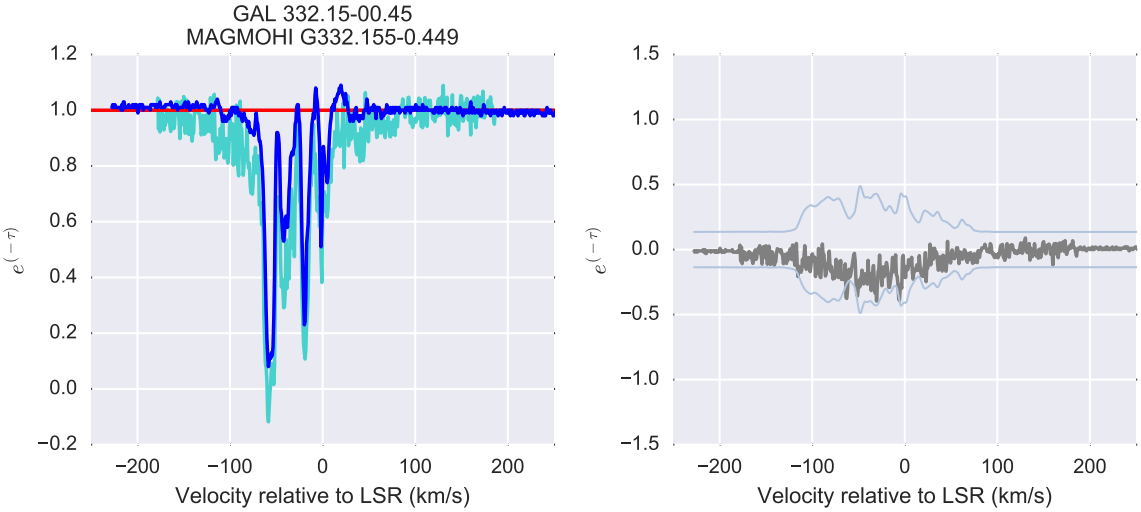

Figure A.23: Comparison of MAGMO and SGPS absorption spectra for GAL 332.15-00.45

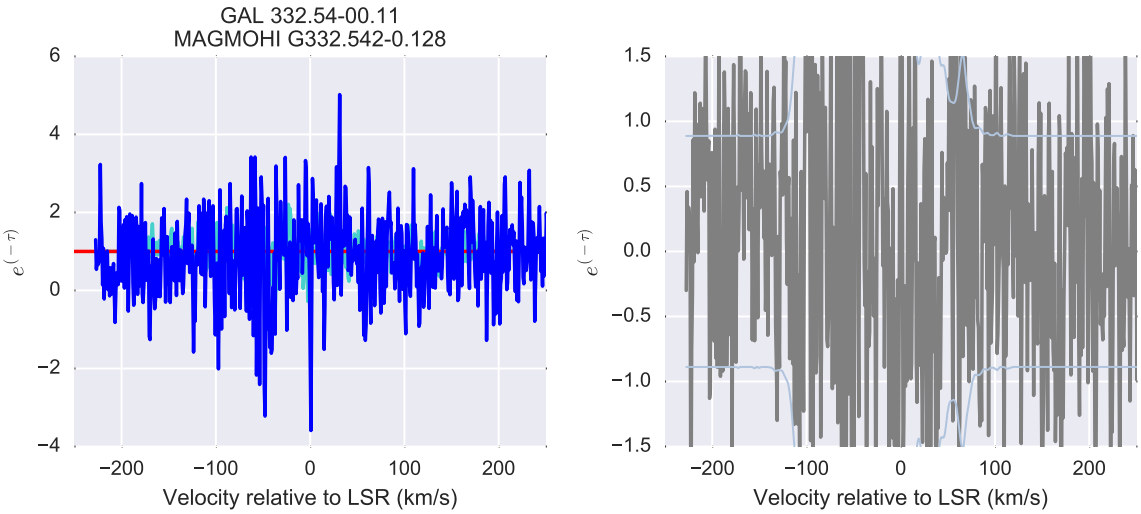

Figure A.24: Comparison of MAGMO and SGPS absorption spectra for GAL 332.54-00.11

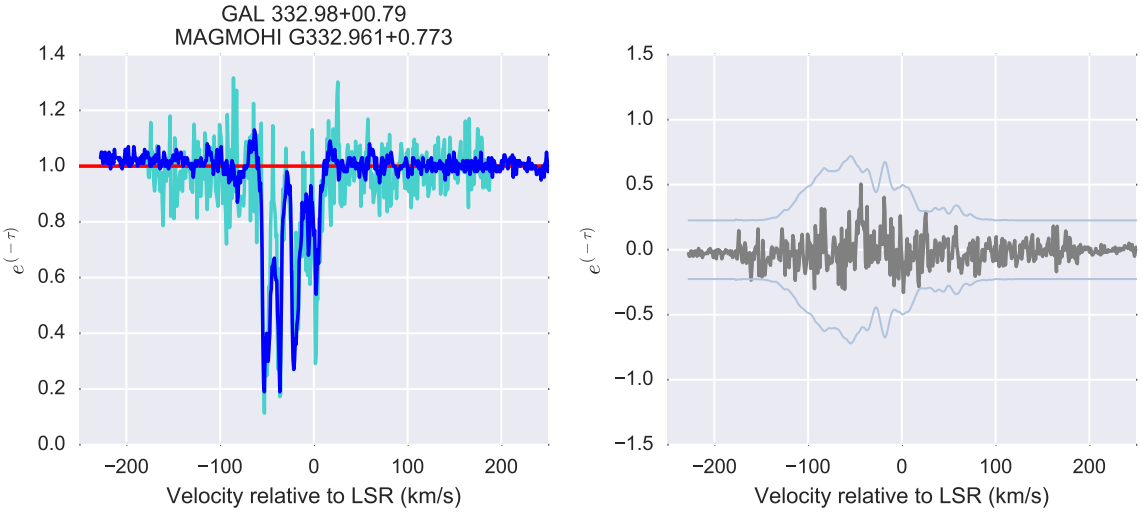

Figure A.25: Comparison of MAGMO and SGPS absorption spectra for GAL 332.98+00.79

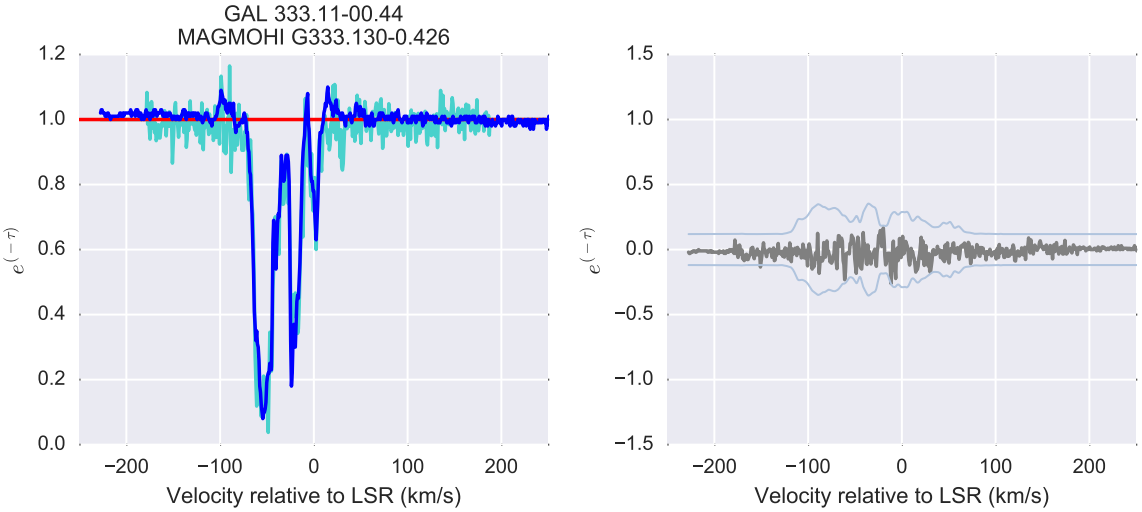

Figure A.26: Comparison of MAGMO and SGPS absorption spectra for GAL 333.11-00.44

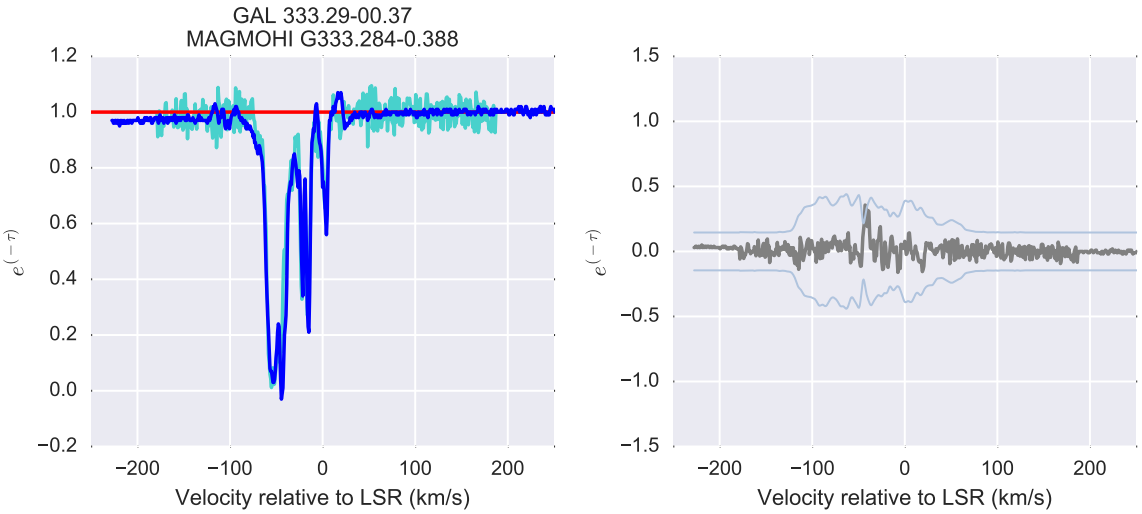

Figure A.27: Comparison of MAGMO and SGPS absorption spectra for GAL 333.29-00.37

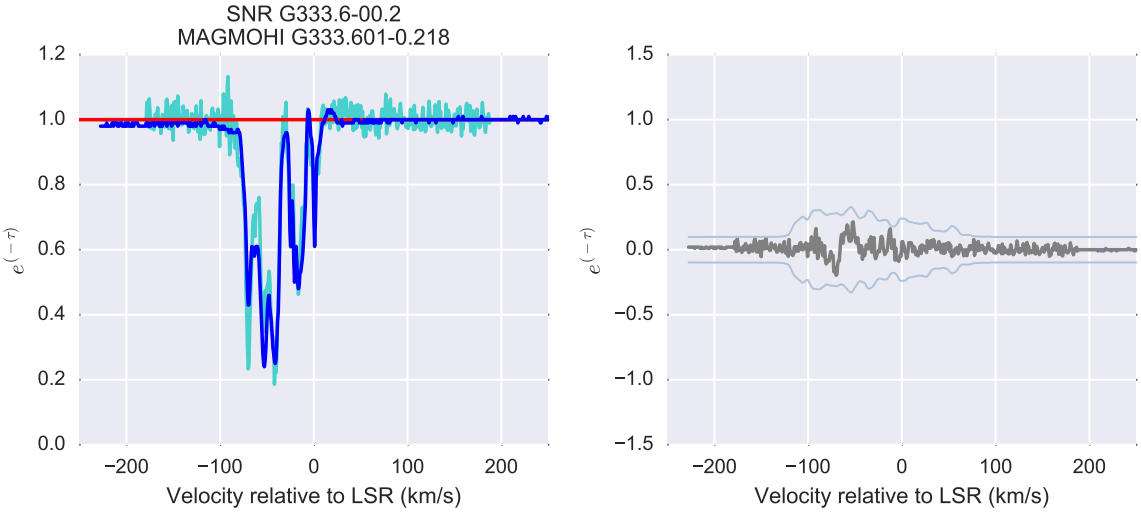

Figure A.28: Comparison of MAGMO and SGPS absorption spectra for SNR G333.6-00.2

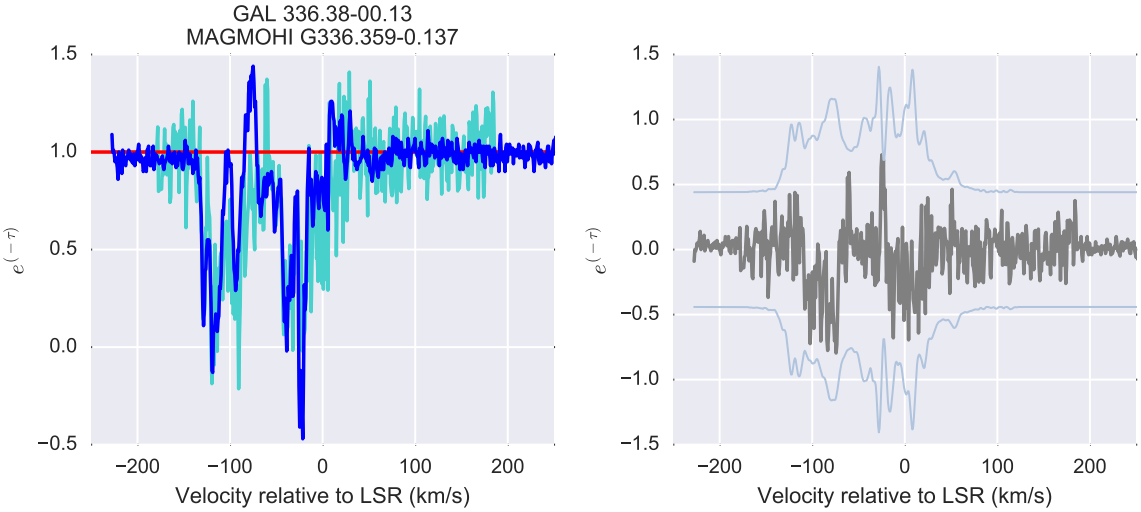

Figure A.29: Comparison of MAGMO and SGPS absorption spectra for GAL 336.38-00.13

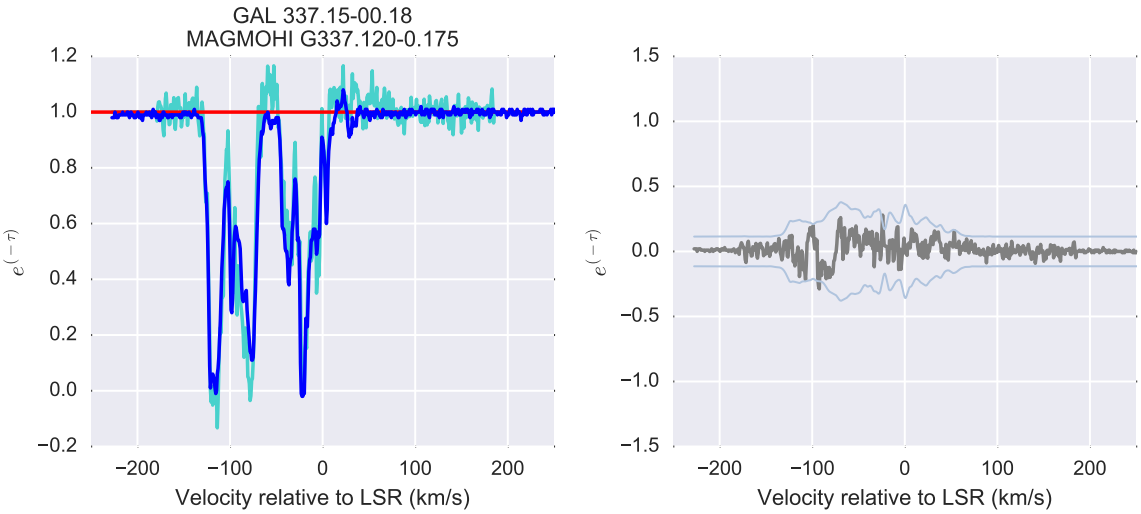

Figure A.30: Comparison of MAGMO and SGPS absorption spectra for GAL 337.15-00.18

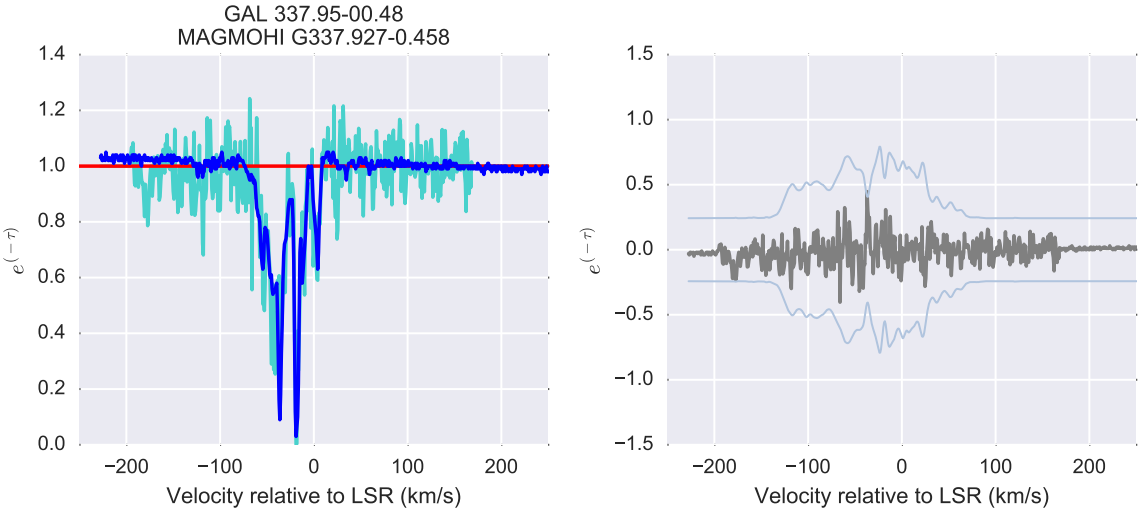

Figure A.31: Comparison of MAGMO and SGPS absorption spectra for GAL 337.95-00.48

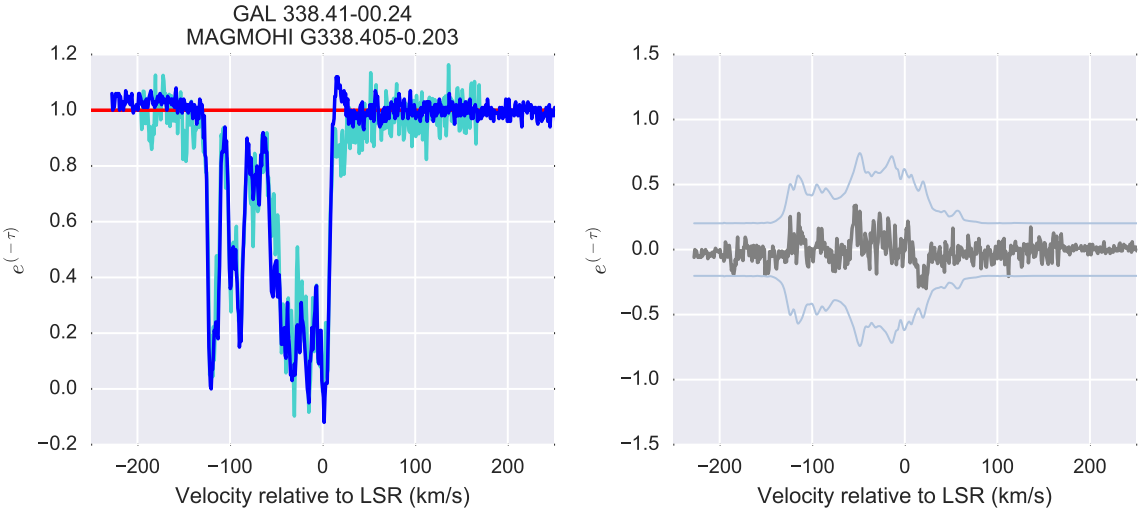

Figure A.32: Comparison of MAGMO and SGPS absorption spectra for GAL 338.41-00.24

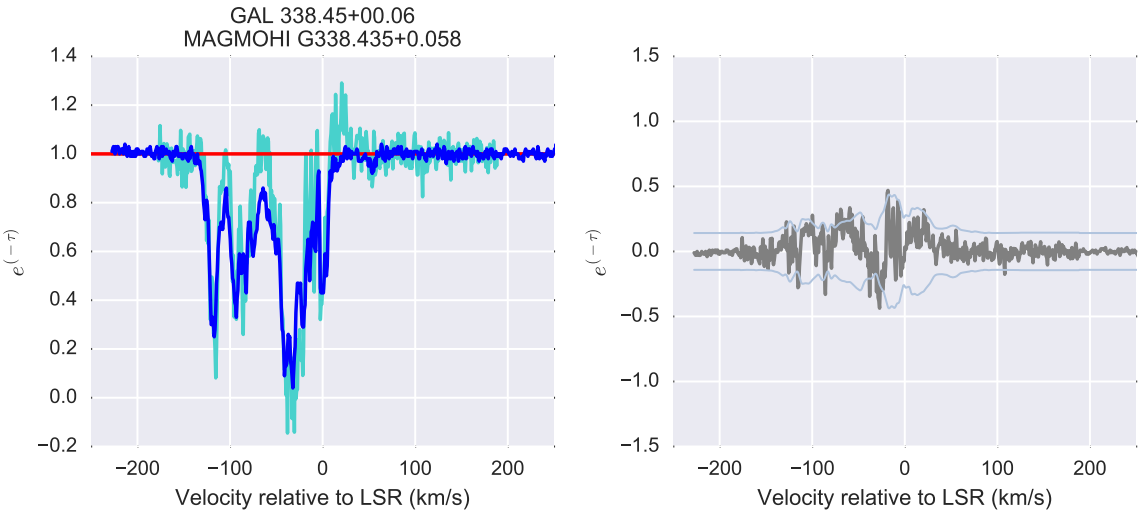

Figure A.33: Comparison of MAGMO and SGPS absorption spectra for GAL 338.45+00.06

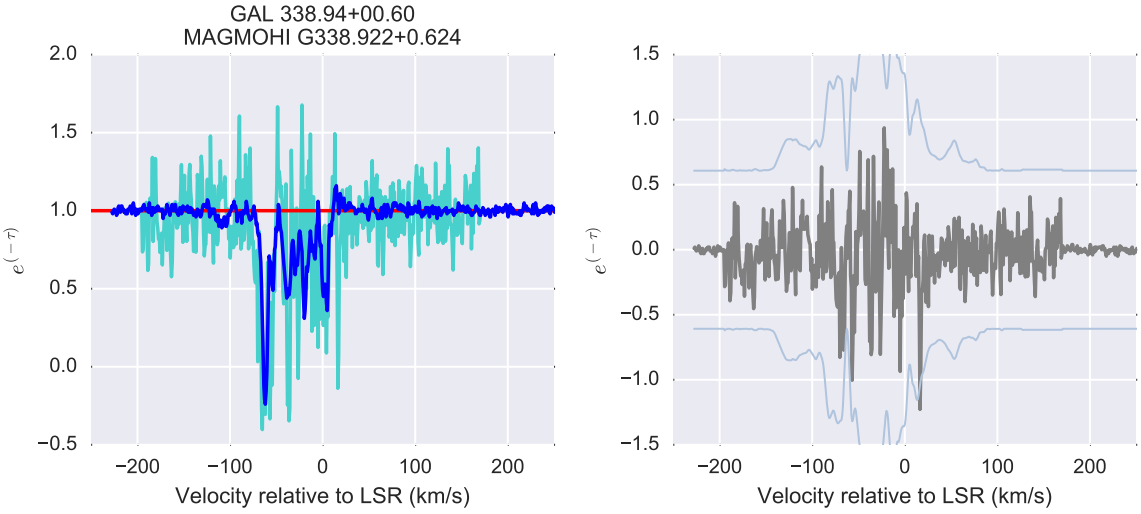

Figure A.34: Comparison of MAGMO and SGPS absorption spectra for GAL 338.94+00.60

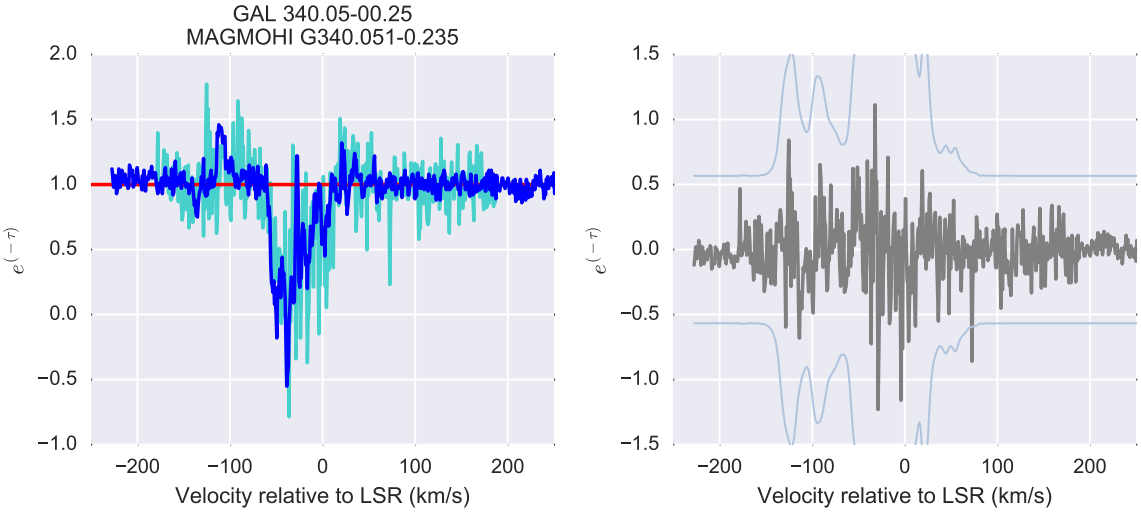

Figure A.35: Comparison of MAGMO and SGPS absorption spectra for GAL 340.05-00.25

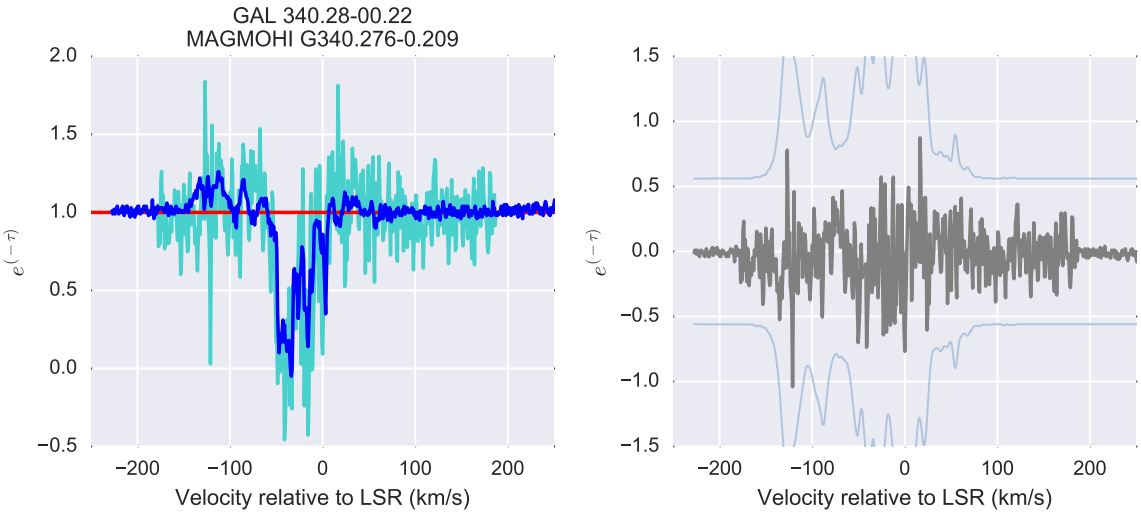

Figure A.36: Comparison of MAGMO and SGPS absorption spectra for GAL 340.28-00.22

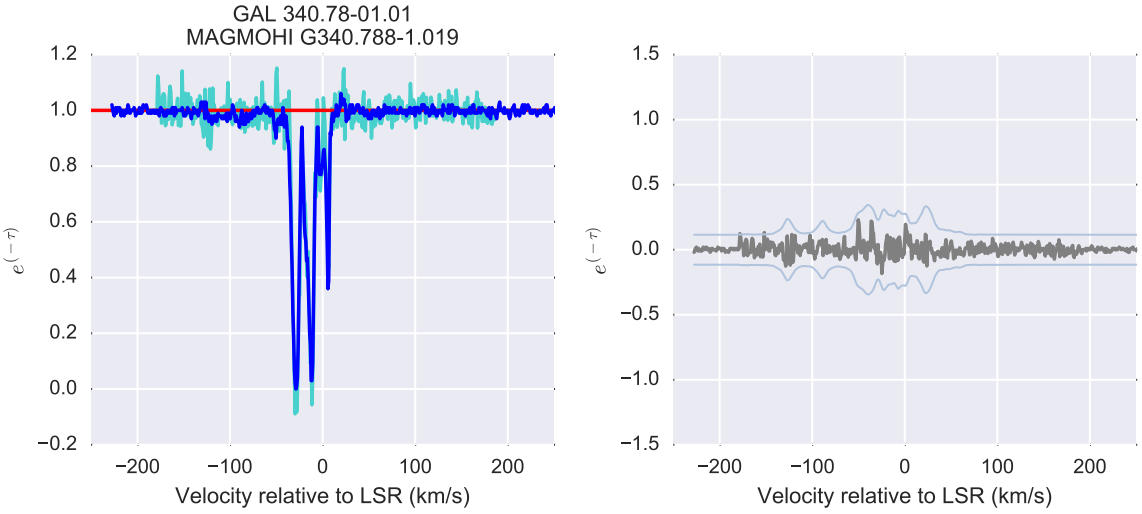

Figure A.37: Comparison of MAGMO and SGPS absorption spectra for GAL 340.78-01.01

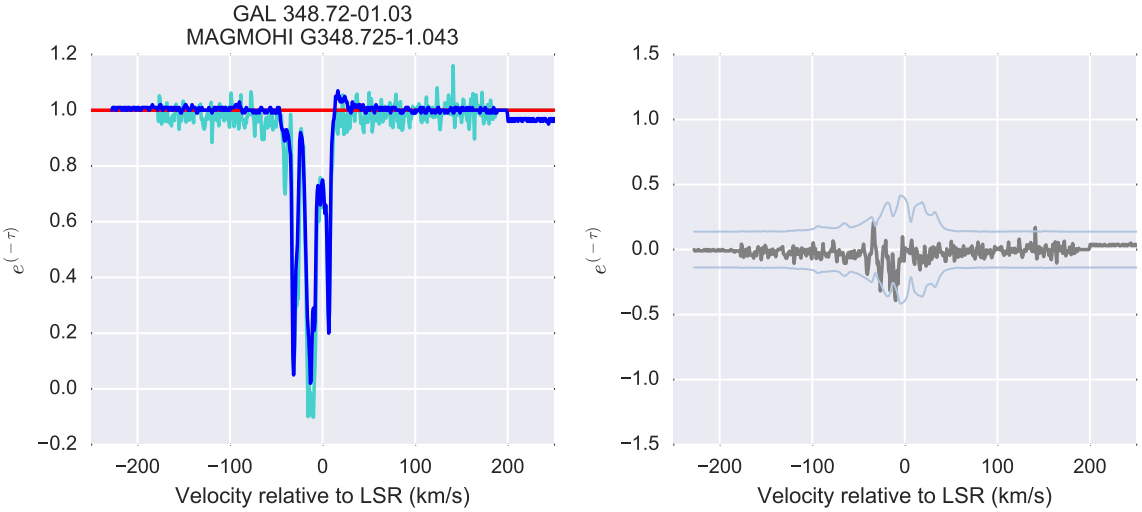

Figure A.38: Comparison of MAGMO and SGPS absorption spectra for GAL 348.72-01.03

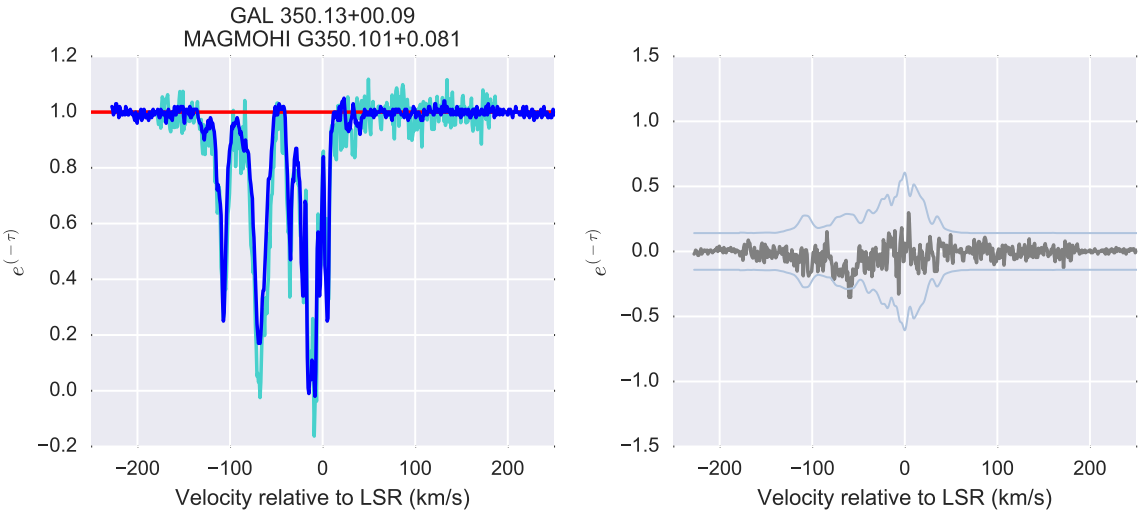

Figure A.39: Comparison of MAGMO and SGPS absorption spectra for GAL 350.13+00.09

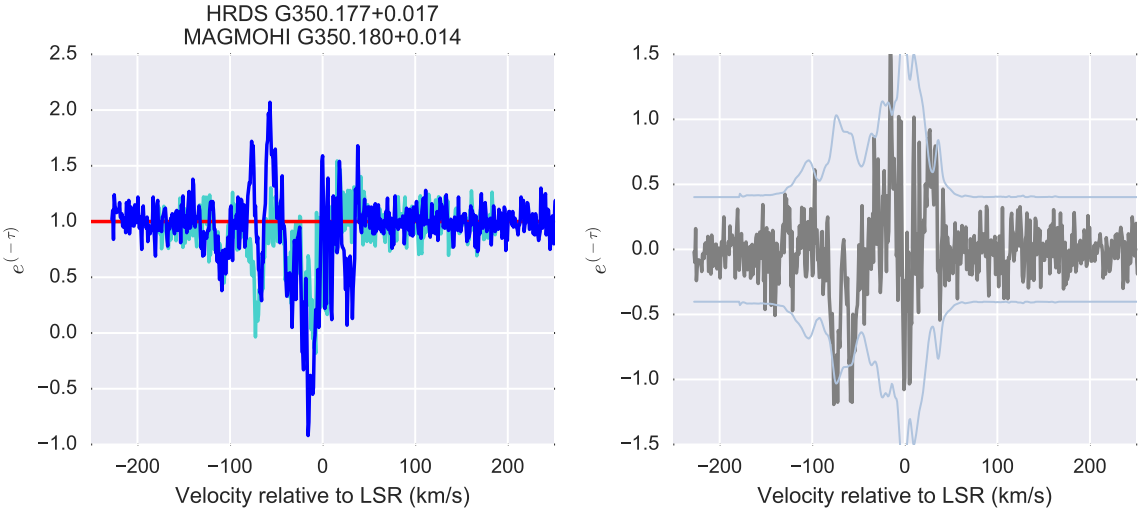

Figure A.40: Comparison of MAGMO and SGPS absorption spectra for HRDS G350.177+0.017

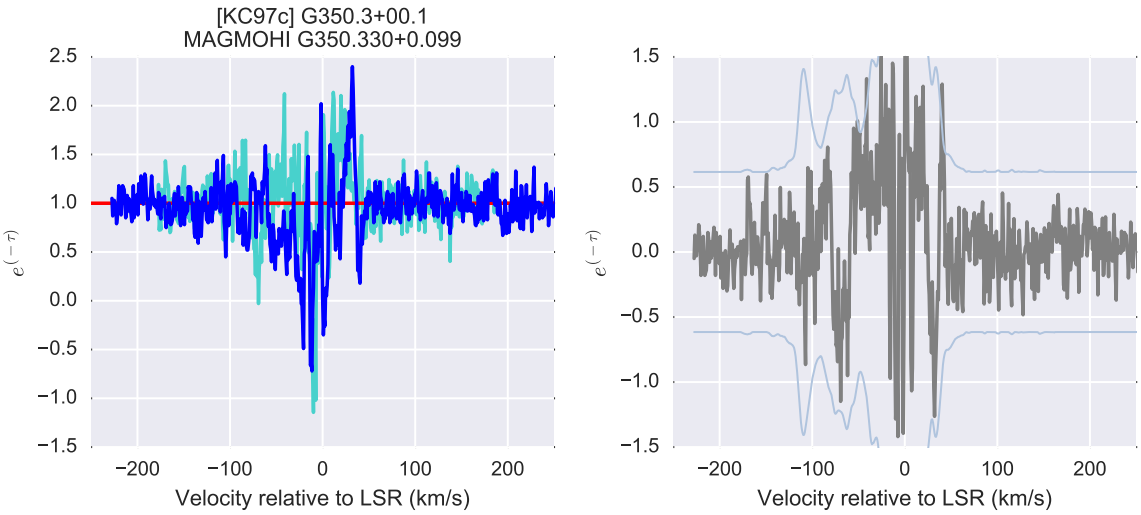

Figure A.41: Comparison of MAGMO and SGPS absorption spectra for [KC97c] G350.3+00.1

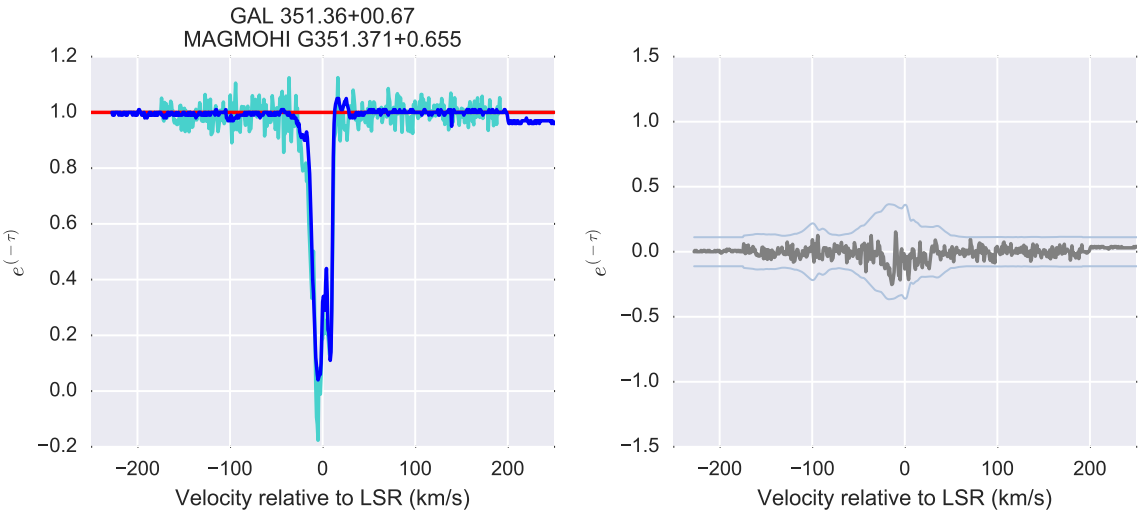

Figure A.42: Comparison of MAGMO and SGPS absorption spectra for GAL 351.36+00.67