



HAL
open science

The HI structures of the Small Magellanic Cloud periphery with high-resolution ASKAP data

Frances Buckland-Willis

► **To cite this version:**

Frances Buckland-Willis. The HI structures of the Small Magellanic Cloud periphery with high-resolution ASKAP data. Astrophysics [astro-ph]. Université Paris-Saclay, 2023. English. NNT : 2023UPASP135 . tel-04430625

HAL Id: tel-04430625

<https://theses.hal.science/tel-04430625>

Submitted on 1 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The HI structures of the Small Magellanic Cloud periphery with high-resolution ASKAP data

*Les structures HI de la périphérie du Petit Nuage de
Magellan avec les données ASKAP à haute-résolution*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 127 Astronomie et Astrophysique d'Île de France (AAIF)
Spécialité de doctorat : Astronomie et Astrophysique
Graduate School : Physique. Référent : Faculté des Sciences d'Orsay

Thèse préparée dans l'unité de recherche **Astrophysique, Instrumentation et
Modélisation de Paris-Saclay** (Université Paris-Saclay, CNRS, CEA), sous la direction de
Marc-Antoine MIVILLE-DESCHÊNES, Directeur de recherche

Thèse soutenue à Paris-Saclay, le 26 Octobre 2023, par

Frances BUCKLAND-WILLIS

Composition du jury

Membres du jury avec voix délibérative

Patrick HENNEBELLE Directeur de recherche, CEA, AIM, CEA-Saclay	Président
Caroline BOT Astronome Adjointe (HDR), Université de Strasbourg	Rapporteuse & Examinatrice
Mary PUTMAN Professeure, Columbia University	Rapporteuse & Examinatrice
Claire MURRAY Chercheuse associée, Johns Hopkins University	Examinatrice
Enrico DI TEODORO Professeur associé, Università degli Studi di Firenze	Examineur
Suzanne MADDEN Directrice de recherche, CEA, AIM, CEA-Saclay	Examinatrice

Titre : Les structures HI de la périphérie du Petit Nuage de Magellan avec les données ASKAP à haute-résolution

Mots clés : formation de structures, milieu interstellaire, 21 cm, analyse de données, nuages de magellan

Résumé : Dans ce travail, je présente les résultats d'une analyse de l'émission HI du Petit Nuage de Magellan (SMC) avec le nouvel interféromètre SKA pathfinder, l'Australian Square Kilometre Array Pathfinder (ASKAP).

Ce travail se place dans le cadre de la collaboration Galactic ASKAP (GASKAP) dédié à un vaste relevé de l'émission à 21 cm de l'hydrogène atomique neutre, le HI, d'une partie du plan galactique et du système de Magellan. L'une des premières zones observées dans le cadre de ce projet fut le SMC. Les études préliminaires ont permis de découvrir l'émission 21 cm de nouvelles structures en périphérie du SMC, ainsi qu'une variabilité de la fraction du HI froid (CNM) à l'aide d'observations en absorption sur un grand nombre de sources radio. Dans ce travail, j'utilise le fait que l'émission à 21 cm permet de cartographier l'information sur les phases du HI (du CNM au WNM). J'analyse les données hyper-spectrales de l'émission de nuages individuels autour du SMC pour révéler la distribution des phases du HI dans l'espace physique et dans l'espace des vitesses.

La première partie de ce travail s'est concentrée sur trois structures ayant des tailles de plusieurs centaines de pc, toutes situées dans la périphérie nord de la galaxie, et précédemment identifiées dans les données de mise en service obtenues par GASKAP. Ces structures, bien résolues par le faisceau de 30" d'ASKAP, ont été analysées à l'aide de ROHSA, un algorithme de décomposition en gaussienne qui est une technique couramment utilisée pour analyser les spectres d'émission de HI. Dans ce travail, j'utilise cet algorithme pour modéliser ces trois structures et reconstruire des cartes de chaque phase. Je trouve que chaque nuage a de grandes variations internes de leurs fractions CNM et dans certains cas une réduction de la fraction CNM à mesure qu'ils s'éloignent de la galaxie. De même, la vitesse radiale change à mesure que l'on s'éloigne de la galaxie, ce qui indique que les différentes phases se déplacent à des vitesses différentes.

La deuxième partie de ce projet s'est concentrée sur la recherche de nuages HI plus petits autour du SMC qui étaient décalés d'au moins 250pc ou 20 km/s de toute autre structure dans l'espace position-position-vitesse. J'ai trouvé 31 nuages en utilisant une combinaison de méthodes de recherche automatique et de vérification humaine. Pour analyser ces petits nuages, j'ai de nouveau utilisé l'algorithme ROHSA pour isoler la structure du CNM en leur sein. En faisant l'hypothèse que ces nuages sont filamentaires, j'ai obtenu des limites inférieures sur la densité de ces nuages ce qui, combiné aux températures dérivées des largeurs de raies, m'a permis d'estimer les conditions physiques de l'environnement où résident ces nuages.

Cette thèse présente les premiers résultats de la décomposition en phase des données 21 cm du SMC, la première partie de l'étude pilote du projet GASKAP. Elle met en évidence les détails avec lesquels nous pouvons maintenant analyser les structures interstellaires HI et donne un aperçu des processus dans l'environnement extrême de la périphérie du SMC. Le SMC est un système extrêmement perturbé sur le plan dynamique, car il interagit fortement avec son partenaire, le Grand Nuage de Magellan (LMC). Du fait de cette interaction, le SMC subit d'importantes forces de marée et il n'est donc pas surprenant que nous trouvions une pléthore de nuages neutres, petits et grands, à des distances significatives de la galaxie. L'analyse effectuée dans le cadre de ce travail nous aide à mieux comprendre la dynamique inhabituelle du système et les forces qui l'ont façonnée. D'autres observations prévues du SMC avec ASKAP, combinées à des études sur d'autres traceurs de gaz et sur la dynamique stellaire, permettront de dresser un tableau plus complet de la structure de la galaxie en 3D et dans l'espace des vitesses, afin d'améliorer notre compréhension de l'un des voisins les plus proches de la Voie lactée.

Title : The HI structures of the Small Magellanic Cloud periphery with high-resolution ASKAP data
Keywords : structure formation, interstellar medium, 21 cm, data analysis, Magellanic clouds

Abstract : In this work I present the results of an analysis of the HI emission spectra survey of the Small Magellanic Cloud (SMC) with the new SKA pathfinder interferometer the Australian Square Kilometre Array Pathfinder (ASKAP).

The Galactic ASKAP collaboration (GASKAP) is conducting a large HI emission survey of the 21cm line to map part of the Galactic plane and the Magellanic system. One of the first areas observed as part of the survey was the SMC. HI emission captures both phases of the neutral atomic medium, the cold and the warm, and previous surveys of the SMC have uncovered new structures in the periphery of the SMC. Previous absorption studies of the SMC have found differing fractions of cold to warm neutral medium (CNM and WNM) so in this work I analyse the emission spectra of individual clouds around the SMC to uncover the distribution of the two phases in physical and velocity space.

The first part of this work focused on three large structures previously identified in the commissioning data obtained by GASKAP, all residing in the northern outskirts of the galaxy. These structures are 100s of pc long, well resolved by the 30" synthesised beam of ASKAP and were analysed using the ROHSA algorithm. ROHSA is a Gaussian decomposition algorithm, which is a common technique used to analyse emission spectra for HI. In this work I use this algorithm to model these three structures and reconstruct maps of each phase for analysis. I find that each cloud has large internal variations of their CNM fractions and in some cases a reduction of CNM fraction as they get further from the galaxy. Similarly the radial velocity changes as we move away from the galaxy with indication that the different phases travel

at different velocities.

The second part of this project focused on searching for smaller clouds of HI around the SMC that were offset by at least 250pc or 20 km/s from any other small-scale structure. I found 31 clouds by using a combination of automated searching methods and human verification. To analyse these small clouds, I again used the ROHSA algorithm to isolate the CNM structure in them and characterised their shape using filament defining packages. This way I obtained lower limits on the density of these clouds. This combined with temperatures derived from the linewidths allowed me to estimate the extragalactic conditions in the areas these clouds reside.

This thesis presents the first results of phase decomposition on the GASKAP SMC dataset, the first part of the pilot survey. It highlights the detail in which we can now analyse structures in HI and gives insight into the processes in the extreme environment of the SMC periphery. The SMC is an extremely dynamically disturbed system as it is interacting strongly with its partner the Large Magellanic Cloud (LMC). From this interaction, the SMC is experiencing large tidal forces so it is no surprise that we find a plethora of gas on large and small scales at significant distances from the galaxy. The analysis done in this work help us better understand the unusual dynamics of the system and the forces that shaped them. Further planned observations of the SMC with ASKAP, combined with surveys in other gas tracers and stellar dynamics will be able to build a more complete picture of the galaxy structure in 3D and velocity space to increase our understanding of one of the Milky Way's nearest neighbours.

This work is dedicated to all the women in Astronomy who have gone before me, both those who were credited, and those who were un-credited. Your passion, dedication, and perseverance, inspired me to get to where I am today.

Acknowledgements

I first want to thank my supervisor Marc-Antoine for guiding me through a thesis during an unprecedented global pandemic, which was no easy feat. His guidance on scientific subjects has helped me expand my horizons of scientific knowledge, and given me the passion to continue to do so going forward. I would also like to thank him for teaching me my first French expressions when I was a clueless beginner!

I would like to thank the members of the GASKAP collaboration for all their insights at our monthly meetings. These discussions helped me interpret my results and gave me more ideas on where to go from there. I would particularly like to thank the Radio Astronomy group at the Research School of Astronomy and Astrophysics in Canberra, Australia for hosting me twice. It was great to talk with old and new members of the team who are doing amazing, innovative things with the the GASKAP data. I am sure there will be many discoveries coming from the survey team.

I would like to thank all those who attended the Interstellar Institute over the course of my thesis. These meeting were invaluable to me in gaining confidence presenting my work in-front of experts in the field. The discussions had in these meetings and collaborations they began will be extremely helpful to me as I begin my academic career.

I would like to acknowledge the Wajarri Yamatji People as the traditional owners of the Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory, where the ASKAP telescope is situated. I also acknowledge the Wiradjuri People as the traditional owners of the Parkes Observatory, where the Murriyang (Parkes) telescope is situated. The data used in this work was collected on these lands.

I would finally like to thank my friends and family back home in Australia. Your support for me making such a big move to the other side of the world was vital. The many lockdown zoom calls and care packages always helped during the ups and downs of living overseas!

Contents

1 Introduction	13
1.1 Neutral Hydrogen (HI)	13
1.1.1 Observational history	13
1.2 HI and the Interstellar Medium (ISM)	15
1.2.1 The different phases of HI	16
1.2.2 Observing the HI	17
1.2.3 Simulating the HI	19
1.3 Gaussian decomposition	20
1.3.1 Fitting emission and absorption spectra	24
1.3.2 Current Gaussian fitting algorithms	26
1.4 The Magellanic System	29
1.4.1 The Large Magellanic Cloud	32
1.4.2 The Small Magellanic Cloud	32
1.4.3 The Magellanic Bridge, Stream and Leading Arm	34
1.4.4 Magellanic Modelling	36
1.5 High/Intermediate velocity clouds	39
1.5.1 Observations	39
1.5.2 Properties	41
1.5.3 Origins	43
1.5.4 HVCs in other galaxies	45
1.5.5 Simulations	46
2 Gaussian decomposition and model selection	48
2.1 Data	49
2.1.1 Noise estimation	49
2.2 ROHSA decomposition testing	50

2.2.1	Hyperparameter tuning	52
2.2.2	Differentiating between solutions	54
2.2.3	Isolating the cloud emission	59
2.2.4	Determining the best fit	62
2.3	Measuring the solution uncertainty	66
3	Phase relationship of HI in 3 large clouds of the SMC	74
3.1	Best fit models	75
3.2	Phase distribution	78
3.3	Velocity structure	80
3.4	Comparison with CO measurements	84
3.4.1	Alpha cloud	85
3.4.2	Hook cloud	85
3.5	HI mass	87
3.6	Morphologies	88
4	HVC/IVC catalogue of the SMC periphery	94
4.1	Identifying further anomalous clouds	94
4.1.1	SMC	99
4.1.2	Other Magellanic fields	102
4.2	Gaussian decomposition of anomalous clouds	104
4.3	Comparison with absorption measurements	105
4.4	Measuring cloud HI density	108
4.4.1	Model selection	109
4.4.2	Measuring cloud depths	113
5	The conditions of the Interstellar Medium around the SMC	117
5.1	Modelling the HI phase diagram	117
5.1.1	Measuring the thermal properties of clouds	120
5.1.2	Measuring the thermal properties of the three large clouds	125
5.1.3	Caveats on ISM environment	126
5.2	Further chemical modelling	127
5.3	All cloud properties	127
6	Conclusions and perspectives	130
6.1	Conclusions	130

6.2 Perspectives	131
A Column density maps of all small clouds	133
Résumé (en Français)	144

List of Figures

1.1	Figure 3(a) and 3(b) from Wolfire et al. (1995) that shows the phase diagram (top) and the heating and cooling rates with dashed and solid lines respectively (bottom).	17
1.2	Excerpt from Figure 7 from Hennebelle et al. (2007) . Left: the density and velocity along the line of sight as a function of the distance along that line of sight. Right: the resulting emission (top) and absorption (bottom) spectra for the same line of sight as a function of velocity (v). Note the absorption spectra is shown in units of $1 - \exp(-\tau(v))$ not $\tau(v)$, hence why the absorption features are positive.	19
1.3	Figure 10 from Heiles and Troland (2003a) showing the emission (top) and absorption (bottom) profiles for source 3C225b. The dashed lines in the emission spectrum represent the individual and total WNM Gaussians fit to the emission. The Dotted lines in both spectra represent the individual and total CNM Gaussians fits to the emission and absorption. The solid lines above zero represent the data and residuals, below zero is the noise.	25
1.4	Figure 17a (left) and 17d (right) from Riener et al. (2019) comparing the same section of GRS CO data fit with GAUSSPY (left) and GAUSSPY+ after both refitting stages (right).	27
1.5	Figure 15 from Marchal et al. (2019) that shows a random 4x4 grid of the fit to observed HI data. The colour of each component indicates its order.	29
1.6	Figure 9 from Nidever et al. (2010) which shows the optical extent of the whole sky with the HI column density of the Magellanic system overlaid in red.	30
1.7	Figure 4 from McClure-Griffiths et al. (2009) that shows the total HI column density of consecutive velocity bins along the Magellanic system. Velocity bins are indicated by the colourscale.	31
1.8	Figure 1 from Staveley-Smith et al. (2003) that highlights the main features of the LMC in relation to the Milky Way and larger Magellanic system.	33
1.9	Figure 1 from McClure-Griffiths et al. (2018) that shows the peak HI emission in the SMC from ASKAP commissioning data. It highlights the two main features of the SMC; the bar, enclosed by the solid rectangle, and the wing, enclosed by the dashed rectangle. It also indicates the directions to the Stream and LMC/Bridge.	34

1.10 Figure 4a and 4b from Bruens et al. (2005) which show the total column density of the Magellanic system (left) and the different sections of the system (right).	35
1.11 Figure 6 from Besla et al. (2012) which shows the final gas distribution for the two different models considered for the evolution of the Magellanic system. The orbits of the LMC and SMC in each model are indicated by the solid and dotted lines respectively.	37
1.12 Figure 2 from Diaz and Bekki (2011) showing the HI observations of Putman et al. (2003) and Bruens et al. (2005) (left) and the final distribution of the test particles from the Diaz and Bekki (2011) model (right).	38
1.13 Figure 12-1 from Wakker and van Woerden (2013) which shows the distribution of all HI emission with $v_{\text{LSR}} > 90 \text{ km s}^{-1}$ based on the Hulsbosch and Wakker (1988) and Morras et al. (2000) results. The colourbar gives the deviation velocity.	40
1.14 Figure 2 from Westmeier (2018) which shows the distribution of all high velocity HI from the HI4PI data (HI4PI Collaboration and et al. 2016). The colourbar gives the LSR velocity.	41
1.15 Left: Figure 4 from Kalberla and Haud (2006) which shows the distribution of the fitted linewidths of all HVCs with the solid red line and the subset of linewidths from HVCs that had only one component profiles with the dotted blue line. Right: Figure 1c from Brüns et al. (2000) that shows a typical HI spectrum from a compact HVC (HVC 125+41-207).	42
1.16 A selection of two head-tail clouds from Figure 1 from Putman et al. (2011), they are cloud #501 (left) and #759 (right) from Putman et al. (2002).	42
1.17 Figure 2 from Barbani et al. (2023) which is a cartoon showing the process of gas being ejected at a rate \dot{M}_{out} falling back onto the disc at a rate of $\dot{M}_{\text{in, fount}}$ with extra inflow from the interaction of the gas with the corona $\dot{M}_{\text{in, int}}$. They also include general accretion from the corona not motivated by the galactic fountain $\dot{M}_{\text{in, cor}}$.	44
1.18 Figure 5(a) from Westmeier et al. (2008) which shows the locations of the HVCs (filled circles) identified in HI around M31. These are overlaid on the total HI column density of M31. The dashed open circles are known satellite galaxies of M31 and the solid open circles are HVCs that were previously known but were not detected in the study.	45
1.19 Figure 1 (top panel) from Heitsch et al. (2022) which shows the spatial distribution of column density, centroid velocity, velocity dispersion, cold gas fraction and metallicity, in that order left to right, of a cold cloud travelling through an ambient medium at an age of 200 Myr.	46
1.20 Figure 6 from Jung et al. (2023) which shows the magnetic field with red arrows overlaid on the density structures in a magnetised, initially clumpy cloud, with the column density indicated by the colourbar.	47

2.1	The trend of χ_{red}^2 of for each hyperparameter as the values are varied for 8 component solutions. The lines indicate the mean χ_{red}^2 and the windows of the corresponding colour indicate the standard deviation.	53
2.2	The distribution of f_{CNM} and χ_{red}^2 for each 8 component solution, with a histogram of χ_{red}^2 above and f_{CNM} to the right.	54
2.3	The distribution of the effective n and χ_{red}^2 for each 8 component solution. The distribution of the effective n at 80% and effective n at 95%, in green and pink respectively, are shown in the top panels.	56
2.4	The trend of χ_{red}^2 for λ_a , λ'_σ , and σ_{init} as the values are varied for 5 component solutions. The lines indicate the mean χ_{red}^2 and the windows of the corresponding colour indicate the standard deviation.	57
2.5	Different σ - v diagrams for the solutions with the lowest (left) and highest (right) χ_{red}^2 values of the exploration with 5 components fit to the data.	58
2.6	σ - v diagram of the unregularised fit to Field 1 to decide the regions for subtraction. The mean velocity and σ of components that were subtracted are marked in red, components kept are marked in pink.	60
2.7	Middle: Mean spectrum of Field 1 before (orange) and after (blue) subtraction of main body. Top: example spectra of three individual sightlines before and after subtraction. Bottom: Mean % difference from original spectrum along the spectral axis. Dotted black line indicates where the emission from the feature of interest begins.	61
2.8	Bottom: Mean spectrum of the Field 2 (left) and Field 3 (right) before (orange) and after (blue) subtraction of main body. Top: example spectra of two individual sightlines before and after subtraction for Field 2 (left) and Field 3 (right).	62
2.9	Column density integrated over 120-131 km s ⁻¹ , before (left) and after (right) removal of clumps that are present in Field 1.	63
2.10	σ - v diagram of solution with $\lambda'_\sigma = 1$ (left) and $\lambda'_\sigma = 100$ (right), the other hyperparameters are kept constant.	64
2.11	σ - v diagram of solution with $\lambda_a = 1$ (left) and $\lambda_a = 100$ (right), the other hyperparameters are kept constant.	64
2.12	Field 2 histogram of the residuals in Brightness Temperature (left), with the skew of the distribution shown in top right corner and the spatial map of χ_{red}^2 values (right).	65
2.13	Field 1 σ - v diagram (left) and the spatial map of χ_{red}^2 values (right).	65
2.14	Field 3 σ - v diagram (left) and the spatial map of χ_{red}^2 values (right).	66
2.15	Full solution set for Field 1. Top: Component 1, Middle: Component 2, Bottom: Component 3. Left: Brightness Temperature Amplitude (a), Middle: Central Velocity (v), Right: Velocity Dispersion (σ). The red and black circles indicate the beamsize.	67
2.16	Full solution set for Field 2. Panels the same as Figure 2.15.	68

2.17 Full solution set for Field 3. Panels the same as Figure 2.15]	69
2.18 Full uncertainty map for Field 1. Top: Component 1, Middle: Component 2, Bottom: Component 3. Left: Brightness Temperature Amplitude (a), Middle: Central Velocity (v), Right: Velocity Dispersion (σ).	70
2.19 The relative strengths from Equation 2.14 of each source of uncertainty in Field 1 (left), Field 2 (middle), Field 3 (right) for each component of each Gaussian as well as the mean of all components.	71
2.20 The total uncertainty from all three sources across the first component of the Field 3 fit. Left: uncer- tainty in the amplitude, Middle: uncertainty in the central velocity, Right: uncertainty in the velocity dispersion.	72
3.1 Column density of all three clouds (Left: Alpha, Middle: Hook, Right: Gamma) integrated over the velocity range indicated in each panel. The blue arrows indicate the direction of the dynamical centre of the SMC from Di Teodoro et al. (2019b).	74
3.2 Spatial maps of the fitted properties of the Alpha cloud from ROHSA, within a $3 \times 10^{19} \text{ cm}^{-2}$ column density contour for each phase. Top: CNM, Bottom: WNM, Left: Column Density, Centre: Central Velocity, Right: Velocity Dispersion. The grey ellipse in the column density maps indicates the beamsize.	76
3.3 Same as Figure 3.2 but for the Gamma cloud.	77
3.4 Same as Figure 3.2 but for the Hook cloud.	78
3.5 CNM fraction for all three clouds (Left: Alpha, Middle: Hook, Right: Gamma) within a $5 \times 10^{19} \text{ cm}^{-2}$ total column density contour.	79
3.6 First moment map of the SMC, ranging from velocities of 75-225 km s^{-1} .	81
3.7 Left: Velocity for the different phases for two different paths through the Alpha cloud (top and bottom). The WNM is shown in red, the CNM in blue with the darker colours indicating higher column density. The error bars in the top panels show the 1σ uncertainties from Section 2.3. The error bars in the bottom panels show the same uncertainties at the 3σ level. Right: Path through the cloud for respective velocity relations. S denotes the start of the path, F denotes the end of the path.	82
3.8 Left: Velocity for the different phases for a path through the Gamma cloud. The parameters in this Figure are the same as for Figure 3.7]	83
3.9 Top Left: Velocity for the different Gaussian components in the field. Component 1 (WNM) is shown in red, Component 2 (CNM) is shown in blue, and Component 3 (CNM) is shown in green with the darker colours indicating higher column density. Bottom Left: The offset between Component 2 and 1 is shown in black and the offset between Component 3 and 1 is shown in orange. Right: Path through the cloud, the same as shown in Figure 3.8]	84
3.10 Left: Velocity for the different phases for a path through the Hook cloud. The parameters in this Figure are the same as for Figure 3.7]	85

3.11 Left: Offset of the identified CO clump velocities with the fitted CNM velocities of the Alpha cloud from ROHSA (indicated by colourbar) shown at the position of each clump against the column density of the CNM (in greyscale, darker indicating higher column density). Right: Offset of the identified CO clump velocities with the fitted WNM velocity of the Alpha cloud from ROHSA (indicated by colourbar) shown at the position of each clump against the column density of the WNM. The size of each circle scales with the strength of the CO flux.	86
3.12 Same as Figure 3.11 but for the Hook cloud.	86
3.13 Figure 17 from Pingel et al. (2022) that shows the intensity weighted velocity (left) and the intensity weighted velocity dispersion (right) of the Alpha cloud.	89
3.14 CNM column density of the Alpha cloud from ROHSA with the shell identified in Staveley-Smith et al. (1997) shown in red, and the shell from Pingel et al. (2022) shown in green.	90
3.15 H α emission for the field of the Alpha cloud from the MCELS data (Winkler et al. 2015). The total HI column density from ROHSA is shown with the cyan contours.	91
3.16 H α emission for the field of the Alpha cloud from the MCELS data (Winkler et al. 2015). The total HI column density from ROHSA is shown with the cyan contours. The green marker shows the position of HW 32, a young stellar cluster of the SMC.	93
4.1 Left: Brightness Temperature map of a single channel of the SMC, at 144.5 km s ⁻¹ , colour scale ranges from 0 to 105 K. Right: Ratio, $\sigma_{XY}(v)/m_{XY}$ of the same channel, for dispersion measured at the 10 pixel scale, colour scale ranges from 0 to 10.	96
4.2 Left: A clump candidate that was accepted using by-eye verification. Right: A clump candidate that was rejected using by-eye verification. Colour scale represents brightness temperature at the channel velocity specified in the top left corner of each panel.	99
4.3 The distribution of the clumps around the SMC with the colourbar representing the velocity of the peak brightness temperature of the clump. The total column density of the SMC is shown in greyscale, with white indicating maximum column density.	101
4.4 Clumps identified for the Magellanic Stream (left) and the Magellanic Bridge (right) with their respective velocity shown with the colourbar. The total column densities of the fields are shown in greyscale, with white indicating maximum column density.	103
4.5 Clumps identified for the LMC with the velocity shown with the colourbar. The total column density of the LMC is shown in greyscale, with white indicating maximum column density.	104
4.6 The absorption source locations for clumps 30, 26, 9 (top row), 17, 6, and 2 (bottom row) with non-detections in pink, overlaid on clump column density. Bottom: location of absorption source detection in white for clump 24. Red point indicates point of high column density for clump 24.	106

4.7	Top: Emission spectra for absorption source J011134-711414 (black), emission spectra for nearby point, positions shown with white and red crosses in Figure 4.6 respectively. The grey and red envelopes indicate the 3σ uncertainty level. Bottom: Absorption spectra for the same source with the 3σ uncertainty level in grey. The blue line is the peak of the absorption (115 km s^{-1}), the green line is the CNM velocity fitted with ROHSA at absorption source (139.6 km s^{-1}), and the purple line is the CNM velocity fitted with ROHSA at the red cross (144.5 km s^{-1}).	107
4.8	Left: Column density of clump 31 CNM component fitted with ROHSA. Middle: Column density modelled as a sphere of constant density (top), column density modelled as a 2D Gaussian profile (bottom). Right: residuals (model-data) for the respective models.	111
4.9	Areas derived for filaments on clump 22 (left) and clump 30 (right) when the filament area is modelled as a sum of rectangles along the filament spine.	113
4.10	Left: The spine constructed along the densest part of clump 24. Right: The fit produced by radfil (Zucker and Chen 2018) showing the background level fit area in green (top) and Gaussian fit to are highlighted in blue (bottom).	114
5.1	Figure 6(b) from Wolfire et al. (1995) showing how the phase diagram changes as Z_g and Z_d change.	120
5.2	Reconstructed phase diagram in this work as the values of Z_g and Z_d are varied together over the same range as in Figure 5.1.	121
5.3	Phase diagram as G'_0 is varied from Wolfire et al. (1995) (left) and reconstructed in this work (right).	121
5.4	Distribution of the best fit values for G'_0 when $\mathcal{M} = 0$ (left) and when $\mathcal{M} = 1$ (right), overlaid on the total density map of the SMC.	123
5.5	Distribution of the best fit values of all clumps and large clouds for G'_0 when $\mathcal{M} = 0$ (blue) and when $\mathcal{M} = 1$ (orange).	124
5.6	Location of the position at which the peak column density was measured (red marker) from the CNM column density map of each large cloud; Alpha (left), Hook (centre), and Gamma (right).	125
5.7	Distribution of the best fit values of all clumps (circles) and large clouds (pentagons) for G'_0 when $\mathcal{M} = 0$ (left) and when $\mathcal{M} = 1$ (right), overlaid on the total column density map of the SMC.	126
5.8	Distribution of the deviation velocities, indicated by the colourscale, overlaid on the total column density map of the SMC, shown in greyscale with white indicating maximum column density.	128
A.1	Left: Integrated column density of Clump 1. Right: Column density weighted, mean spectrum across the field of Clump 1. Pink window indicates the velocity range the column density is integrated over.	133
A.2	Same as Figure A.1, but for Clump 2.	134
A.3	Same as Figure A.1, but for Clump 3.	134
A.4	Same as Figure A.1, but for Clump 4.	134

A.5 Same as Figure A.1, but for Clump 5.	135
A.6 Same as Figure A.1, but for Clump 6.	135
A.7 Same as Figure A.1, but for Clump 7.	135
A.8 Same as Figure A.1, but for Clump 8.	136
A.9 Same as Figure A.1, but for Clump 9.	136
A.10 Same as Figure A.1, but for Clump 10.	136
A.11 Same as Figure A.1, but for Clump 11.	137
A.12 Same as Figure A.1, but for Clump 12.	137
A.13 Same as Figure A.1, but for Clump 13.	137
A.14 Same as Figure A.1, but for Clump 14.	138
A.15 Same as Figure A.1, but for Clump 15.	138
A.16 Same as Figure A.1, but for Clump 16.	138
A.17 Same as Figure A.1, but for Clump 17.	139
A.18 Same as Figure A.1, but for Clump 18.	139
A.19 Same as Figure A.1, but for Clump 19.	139
A.20 Same as Figure A.1, but for Clump 20.	140
A.21 Same as Figure A.1, but for Clump 21.	140
A.22 Same as Figure A.1, but for Clump 22.	140
A.23 Same as Figure A.1, but for Clump 23.	141
A.24 Same as Figure A.1, but for Clump 24.	141
A.25 Same as Figure A.1, but for Clump 25.	141
A.26 Same as Figure A.1, but for Clump 26.	142
A.27 Same as Figure A.1, but for Clump 27.	142
A.28 Same as Figure A.1, but for Clump 28.	142
A.29 Same as Figure A.1, but for Clump 29.	143
A.30 Same as Figure A.1, but for Clump 30.	143
A.31 Same as Figure A.1, but for Clump 31.	143

List of Tables

2.1 Field parameters	49
2.2 Best solution ROHSA parameters for each field with their corresponding χ^2_{red}	66
2.3 Measured uncertainties for the CNM and WNM HI phases for each field.	73
3.1 Fit results, A, v, σ_v , for the components fit to each cloud and the corresponding mean maximum kinetic temperature of each component.	75
3.2 The HI phase masses and total masses for each cloud. The uncertainties reported as specified in Section 3.5.	88
4.1 Field centres and peak velocities of 31 identified clumps from the SMC	100
4.2 Clump parameters and measured properties, with $\lambda_v = \lambda_\sigma = 10$. For each cloud is listed: Clump number, number of components fitted, $\lambda_A, \lambda'_\sigma, \chi^2_{\text{red}}$ value for the field, skew measure of the residuals after model subtraction, mean velocity dispersion of coldest component fit with ROHSA, width of cloud, uncertainty in the width, maximum column density of coldest component fit with ROHSA, number density.	116
5.1 Further measured properties for each clump. For each cloud is listed: number density, uncertainty in the number density, maximum kinetic temperature of coldest component fit with ROHSA, uncertainty in the maximum kinetic temperature, thermal pressure, uncertainty in the thermal pressure.	122
5.2 Measured properties for each large cloud. For each cloud is listed: width of cloud, uncertainty in the width, maximum column density of coldest component fit with ROHSA, number density, uncertainty in the number density, maximum kinetic temperature of coldest component fit with ROHSA, uncertainty in the maximum kinetic temperature, thermal pressure, uncertainty in the thermal pressure.	125

5.3	Properties of all clouds, compiled from Tables 2.1, 3.1, 4.1, 4.2, 5.2, 5.1	Columns: (1) Cloud name/number,
		(2) Field centre Right Ascension, (3) Field centre Declination, (4) Peak velocity of cloud, (5) Deviation ve-
		locity of cloud from the SMC, (6) Mean velocity dispersion of coldest component fit with ROHSA, (7)
		Width of cloud as fit in Section 4.4.2 (8) Uncertainty in the width, (9) Maximum column density of cold-
		est component fit with ROHSA, (10) Number density as calculated in Section 4.4.2 (11) Uncertainty in
		the number density, (12) Maximum kinetic temperature of coldest component fit with ROHSA from
		3.1, (13) Uncertainty in the maximum kinetic temperature, (14) Thermal pressure from 5.16 (15) Un-
		certainty in the thermal pressure. Peak velocities for the large clouds are calculated as the closest
		velocity channel to the mean of the velocities reported for each component related to each cloud in
		3.1. Deviation velocities (v_{DEV}) are defined by Equation 5.19. 129

Chapter 1

Introduction

1.1 Neutral Hydrogen (HI)

Neutral Hydrogen, often referred to as just HI, is one form of hydrogen in the interstellar medium (ISM). Typically, neutral hydrogen can remain neutral between temperatures of 50-8000K (McKee and Ostriker 1977, Wolfire et al. 1995), below which it condenses to form molecular hydrogen, needed to form stars, or above which it becomes ionised and observable at optical and X-ray wavelengths. Before the first observation of HI, observations had been made of ionised hydrogen (HII) towards star forming regions, confirming the presence of hydrogen in these areas, but the true amount of hydrogen in the ISM could not be determined by ionised hydrogen alone. The detection of the neutral hydrogen radio line in 1951 revolutionised the field of radio astronomy.

1.1.1 Observational history

In their neutral state, hydrogen atoms emit a small amount of radiation from a spin-flip transition, at a rest frequency of 1.420 GHz, which can be detected by ground based radio-telescopes. This emission line was first theorised to be possible, but improbable to observe in 1944 by H.C. van de Hulst. It was then looked at by Shklovskii in 1949, who gave a more favourable judgement to possibility of observation, his probability calculation was only off the now accepted value by a factor of 4 (Ewen and Purcell 1982). Finally, it was observed in the early 1950s independently by three teams: Ewen and Purcell in The United States, Muller and Oort in The Netherlands, and Christiansen and Hindman in Australia. Ewen and Purcell and Muller and Oort published their discovery in Nature in 1951 (Ewen and Purcell 1951, Muller and Oort 1951), which was accompanied by a cable from Joseph Pawsey in Sydney, Australia confirming the detection of Ewen and Purcell with observations conducted by Christiansen. Christiansen and Hindman published their results the next year (Christiansen and Hindman 1952) which built upon the existing evidence for the detectability of the HI line. They produced a comprehensive map of multiple lines of sight in the southern

sky with two main takeaways. First, that the intensity of the HI emission varies as you move along the galactic plane and that this was likely an indication of the inhomogeneous distribution of the HI gas. Second, that there were multiple, spatially consistent, detections of double-peaked profiles in one quadrant of the galaxy. This was proof that there was more than one gas profile, and in their work they suggested this was coming from the spiral arms. These early detections were incredibly important in proving to the astronomy community that it was viable to observe the neutral atomic gas directly.

The advancement of this field did not slow down, with the first extragalactic observation of HI being reported by Kerr and Hindman in Australia in 1953 (Kerr and Hindman 1953) with their observations of the Magellanic clouds. With a physical resolution of 1° , they were able to observe that the HI gas content of the two galaxies were very similar, despite large differences in the optical emission of each galaxy. They also managed to show that the HI extent of each galaxy is larger than the optical one. Just as quickly as the HI line was observed in emission, it was then observed in absorption in 1954 by Hagen, McClain and Hepburn (Hagen et al. 1954) on a number of galactic lines of sight.

This rapidly emerging field of radio astronomy demanded the construction of newer and more powerful telescopes. There were a number of iconic radio telescopes constructed in the years following, particularly the 1960s; Parkes (Australia, 1961), 300 Foot Telescope (USA, 1962), Arecibo (Puerto Rico, 1963), and Nançay (France, 1967). Even after this boom, the following were constructed; Effelsberg (Germany, 1971), Very Large Array (USA, 1980), Great Metre Radio Telescope (India, 1995), Greenbank Telescope (USA, 2000). All of these telescopes, some single dishes, some interferometers, have been used to observe the 21cm line and have provided the field of neutral hydrogen with the tools to map the structure of galaxy and beyond, in hydrogen. Single dish telescopes have done a lot of work on mapping large fields of the sky in big surveys e.g. Effelsberg-Bonn HI survey (EBHIS, Kerp et al. (2011)), the Parkes Galactic All-Sky Survey (GASS, McClure-Griffiths et al. (2009)), and the Arecibo Legacy Fast ALFA (ALFALFA, Giovanelli et al. (2005)). These large dishes have good sensitivity due to their collecting area being that of the entire dish, however, they are often limited in angular resolution compared to interferometers. The angular resolution of a telescope (θ) is proportional to the wavelength of the radiation being observed (λ) and inversely proportional to the dish diameter (D) such that the relation looks like this:

$$\theta = 1.22 \frac{\lambda}{D} (\text{radians}). \quad (1.1)$$

So, even the largest dish observing today, the Five-hundred-meter Aperture Spherical radio Telescope (FAST), which has a dish diameter of 500m would only be able to resolve $1.8'$. On the other hand, interferometers consist of multiple dishes, typically with dish diameters on the order of 10s of metres, that form one telescope. The signals are correlated together from multiple dishes and the resolution is determined by the distances between the different dishes (referred to as the baselines) rather than diameters of the dishes themselves, which instead determines the

field of view. So, to get a resolution of 1" with two dishes, they would have to be placed 53 km apart, which is much more practical than building a 53 km diameter single dish. Modern interferometers consist of multiple dishes all at different distances from each other, forming an array.

The processing of interferometric data for imaging requires Fourier transforming the visibility as a function of the baseline coordinates to the sky brightness distribution as function of the sky coordinates, as per the van Cittert-Zernike theorem (Thompson et al. 2017). The baselines are represented in the u-v plane by their position in the array in the east-west and north-south directions. The more of the u-v plane that is covered by an observation, the better the imaging quality, as gaps in the u-v plane coverage create sidelobes in the point spread function, which shows the way a point source would be imaged by the array. Gaps in the u-v plane also reduce the sensitivity to certain spatial scales on the sky. Single dishes are sensitive to all spatial scales up until the limiting scale imposed by the diameter of the dish, which is referred to as the resolution. This is because the u-v plane coverage of a single dish telescope is a solid circle whose radius is the dish diameter on the u-v plane. The goal of an interferometric array is to have an arrangement such that during an observation most of the u-v plane becomes covered. While each baseline pair is just a point on the u-v plane, as the Earth rotates during an observation the position of that baseline on the u-v plane changes as its orientation changes relative to the source it is observing. This means that each baseline creates a track on the u-v plane as the Earth rotates, filling up more of it. So by using the rotation of the Earth and the configuration of the array, the u-v plane will become as filled in as possible. One inherent limitation of the u-v plane coverage with interferometry is that the centre of the u-v plane cannot be filled as the baselines cannot be smaller than the diameter of a single dish and is referred to as the "short-spacing" problem. This missing coverage means observations are insensitive to large spatial scales, which can be advantageous for some studies, but in general it needs to be rectified. Typically, some form of data combination with observations from a single dish telescope are done to complete the missing u-v coverage.

Together, single dish telescopes and interferometers have allowed us to conduct anything from large-scale surveys to high resolution targeted observations. Astronomers have taken the field of neutral hydrogen a long way in the past almost 80 years, starting with observations of our own galaxy and extending further and further out into the universe, to where we are in the current day to a detection of the HI signal at a redshift of 1.3 (Chakraborty and Roy 2023).

1.2 HI and the Interstellar Medium (ISM)

The interstellar medium (ISM) consists of all the material between stars within a galaxy. The overwhelming majority of this material is gas but it also includes dust and radiation. Of the gas, around 90% of it is hydrogen while the rest is mostly helium and other metals and molecules. Neutral hydrogen makes up part of this hydrogen fraction and is seen throughout galaxies. Looking at the relationship between neutral hydrogen and all the other species in the ISM

is crucial to understanding how the HI forms, changes phase, or condenses into molecular hydrogen. In particular, the cooling of HI gas to eventually form molecular hydrogen (H_2) is of particular interest, as this is a key step towards star formation in a galaxy. To understand the role of HI in the ISM, we need to have some understanding of the phases of neutral hydrogen.

1.2.1 The different phases of HI

Neutral hydrogen can only exist under specific conditions and if it exceeds a certain temperature, will become collisionally ionised. In general, neutral hydrogen is thought to exist as a two-phase medium, the cold neutral medium (CNM) and the warm neutral medium (WNM). However, many studies have pointed to the existence of a phase in-between these two, often called the unstable neutral medium (UNM) or sometimes the lukewarm neutral medium (LNM). The conditions that these exist under are set by the balance of heating and cooling processes that occur in the ISM, but under galactic conditions in equilibrium the CNM is normally defined as having a temperature of between 50-500K and the WNM is defined as having a temperature of 5000-8800K (Heiles and Troland 2003b). In looking at the balance of heating and cooling processes, it is found that under most conditions there is a window of pressures in which CNM and WNM can both exist, as well as a thermally unstable region in temperature occupied by the UNM. The UNM exists in the temperature range between the CNM and WNM, so typically 500-5000 K in galactic conditions. By distinguishing between these phases in temperature, measurements of the physical distribution of the two phases can be made from observations and simulations.

The seminal paper of Field (1965) comprehensively described the ISM thermal instability that was then vigorously explored in the important papers of Wolfire et al. (1995) and Wolfire et al. (2003), which detail the various heating and cooling processes of the HI in the ISM. The dominant heating and/or cooling process changes with the density of the gas, this is shown in Figure 1.1. At all densities photoelectric heating from dust grains dominates out of all the heating processes, however at low densities X-ray heating becomes important and at high densities photoionisation of CII becomes important. For the cooling processes, cooling by CII fine structure is dominant at high densities whereas at low densities recombination and $Ly\alpha$ transitions are most important. The balancing of all these different processes leads to the phase diagram also seen in Figure 1.1. The phase diagram in Figure 1.1 is calculated using standard local values and is representative of the local ISM, but even when different values are used, the phase diagram still retains this distinctive 'wiggle'. This 'wiggle', seen in Figure 1.1 is where the gradient of the phase diagram changes from positive to negative and then back again, the negative gradient is seen where $-0.2 < \log(n) < 0.4$. A negative gradient indicates the gas is in an unstable phase, any change to its equilibrium state and it will shift to a stable area of the phase diagram and become either WNM or CNM. This 'wiggle' also means that there is a pressure window in which CNM and WNM can both exist in equilibrium. For the standard conditions shown in in Figure 1.1 this is $3.0 < \log(P/k) < 3.6$. This means that CNM and WNM can be observed for HI structures in equilibrium within

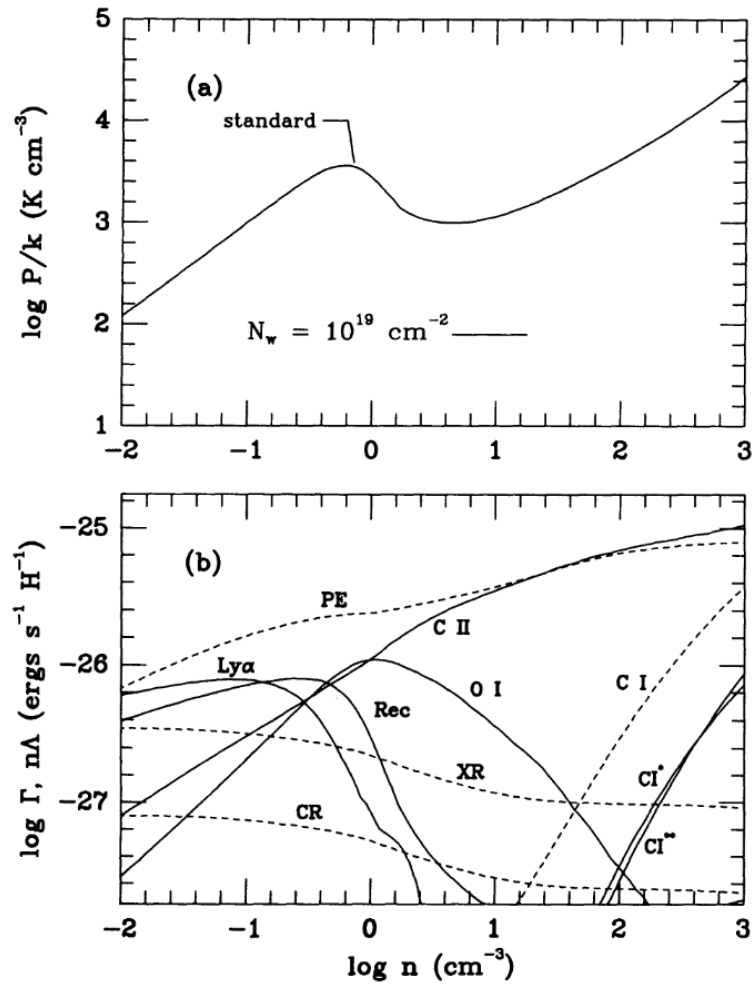


Figure 1.1: Figure 3(a) and 3(b) from [Wolfire et al. \(1995\)](#) that shows the phase diagram (top) and the heating and cooling rates with dashed and solid lines respectively (bottom).

this range. The presence or absence of the both phases can therefore constrain the thermal pressure of the ISM environment. However, if the gas is not in equilibrium, then the presence of the different phases cannot be used to well-constrain the conditions of the ISM.

1.2.2 Observing the HI

As detailed in Section [1.1.1](#) the HI has been widely observed over the decades. The two ways to observe the HI are in emission and absorption. Emission observations of the HI give the brightness temperature of the gas a function of velocity $T_B(v)$. Absorption measurements show the absorption of the HI gas of continuum emission from radio sources, or emission produced by HI behind the absorbing cloud. The former is typically just referred to as absorption, whereas the latter is called HI self-absorption (HISA). The HI gas emission is Doppler shifted by the speed at which it travels, allowing for multiple HI signals travelling at different velocities to an emission or absorption spectrum. It is often thought that CNM is measured in absorption and WNM is measured in emission, however that is not

necessarily the case. The brightness temperature spectrum ($T_B(v)$, i.e. emission) is parametrised by the equation:

$$T_B(v) = T_s(1 - \exp(-\tau(v))) \quad (1.2)$$

where T_s is the spin temperature and τ is the optical depth. For a spin temperature of 50 K and an optical depth of just 0.5 the brightness temperature would be 20 K, well above the observational limits of most HI observations. Additionally, while WNM components have small optical depths, sensitive absorption surveys have detected at least the UNM (Murray et al. 2018) and if sensitivity can be increased, the WNM could also be observed in absorption.

Thus, theoretically both phases can be observed with both methods, however both the measurements are necessary to fully constrain the HI. If the HI is optically thick, then the emission from that HI is saturated and the true column density does not then scale linearly with the brightness temperature. Measurement of the optical depth spectrum from absorption and brightness temperature spectrum from emission can be used in Equation 1.3 to give the total column density (N_H), with the assumption that the gas is isothermal:

$$N_H = C \int \frac{\tau(v)T_B(v)}{(1 - \exp(-\tau(v)))} dv \quad (1.3)$$

where $C = 1.823 \times 10^{18} \text{ cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$. In the case that the gas is optically thin, that is $\tau \ll 1$, Equation 1.3 is approximated to:

$$N_{H,\text{thin}} = C \int T_B(v) dv. \quad (1.4)$$

This optically thin column density can be obtained directly from just emission spectra.

While CNM can be seen in absorption, extracting its properties either requires Gaussian decomposition (detailed in Section 1.3) or for the emission spectrum to be simplistic enough that the CNM emission does not overlap with any other emission signal in velocity. Simplistic emission spectra are more common at higher galactic latitudes and so the CNM distribution of the galactic plane are best constrained by absorption. McClure-Griffiths et al. (2023) summarise the current state of knowledge about the relative fraction of the phases. They highlight that from the Millennium Arecibo survey (Heiles and Troland 2003b), which sampled all galactic latitudes, most sightlines had a CNM fraction below 0.3. They also then point to the recent study that found low CNM fractions of Complex C in Murray et al. (2021), with a median CNM fraction of 0. The 21-SPONGE survey (Murray et al. 2015a), a survey of high-latitude sightlines, find a median CNM fraction of 0.2. Clearly, it is not possible to define a universal CNM fraction and in all of these studies, there are some sightlines that have very high CNM fraction, near 1. Even producing a definitive map of the CNM fraction is difficult, as absorption samples very small scales and does not uniformly cover the galaxy, due to the background source density limitation. However maps have been made with machine learning methods (Murray et al. 2020). Murray et al. (2020) found by training a convolution neural network on emission-absorption pairs in the high latitude sky, the algorithm could predict the CNM fraction and column density correction factor (to account for

difference between column densities from Equations 1.3 and 1.4). They find some areas with high CNM fractions, but most have fractions below 0.2, similar to what was found in Heiles and Troland (2003b).

1.2.3 Simulating the HI

Evidently, there are areas with enhanced CNM fractions, caused by the processes acting on the gas. To understand the processes at play and compare to observations, simulations are needed. Audit and Hennebelle (2004) explore the impact of turbulent flows on the condensation of WNM into CNM. They found that in the case of colliding flows with weak turbulence, CNM structures were able to fragment and form quickly. As the strength of the turbulence increases, the fraction of the gas in the unstable phase does as well and there is less CNM produced. They also found that the CNM structures that were formed were in pressure equilibrium with the WNM, and had a typical temperature of 80 K and width of 0.1 pc.

Comparing the results of the simulations to what can be seen in observations, is crucial to understanding the HI. Hennebelle et al. (2007) produced emission and absorption spectra for multiple lines of sight. They comment on the broadening of the emission signals due to turbulence and argue that it is not due to the turbulence of the individual CNM structures, but rather the extra velocity dispersion is introduced by multiple unrelated CNM structures at similar velocities, so a blending problem. Figure 1.2 shows a example for a sightline where there are four main density enhancements along the line of sight, of which three travel at velocities near 0 km s^{-1} and one travels at a velocity closer to 5 km s^{-1} . This results in a spectrum with two components, with the one near 0 km s^{-1} broadened compared to the one at 5 km s^{-1} .

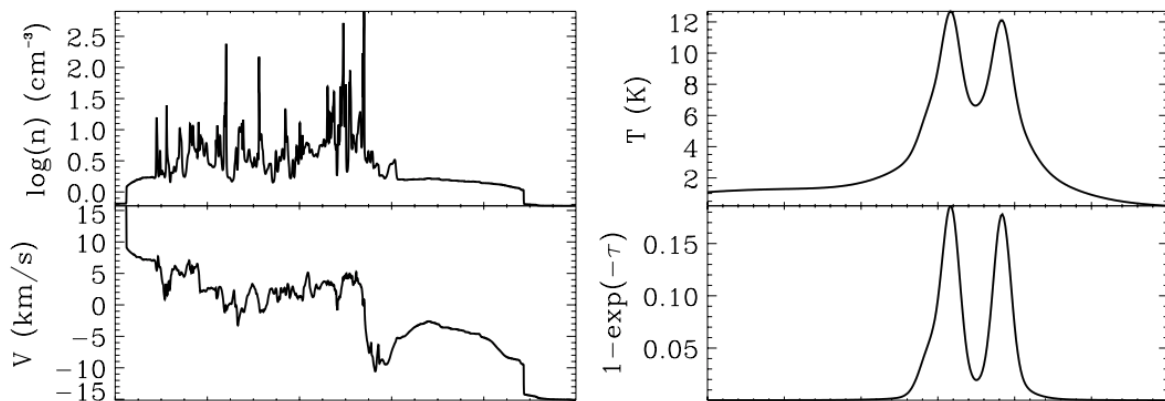


Figure 1.2: Excerpt from Figure 7 from Hennebelle et al. (2007). Left: the density and velocity along the line of sight as a function of the distance along that line of sight. Right: the resulting emission (top) and absorption (bottom) spectra for the same line of sight as a function of velocity (v). Note the absorption spectra is shown in units of $1 - \exp(-\tau(v))$ not $\tau(v)$, hence why the absorption features are positive.

A study of the transition of WNM into CNM by Saury et al. (2014) showed a WNM density above 1 cm^{-3} was required to cause condensation of the WNM into CNM. In the low density cases where CNM was not being formed, increasing the turbulence did nothing to help produce CNM. In the higher density cases, where CNM was forming,

the higher the initial WNM density, the more CNM was produced. For this defining threshold of density 1 cm^{-3} , the ratio of compressive to solenoidal turbulence was varied to investigate its effects on CNM formation. They found that the CNM only forms when the majority of turbulence comes from the compressive mode. This indicates that CNM preferentially condenses from the WNM when experiencing compressive forces, like supernovae, outflows, shocks etc.

Simulations including ideal MHD by [Hennebelle \(2013\)](#) show that magnetic fields help form dense filaments and keep them together. They do not investigate the phases of the gas in this study, but dense structures indicate cold gas. So this study suggest that magnetic fields aid in the formation and survival of dense gas in the ISM, which typically forms CNM. [Inoue and Inutsuka \(2016\)](#) investigated the role of magnetic fields in shocked gas and found that the shock front induces a thermally-unstable, magnetised layer behind it which forms CNM fragments. These fragments in turn form CNM filaments which are prevalent throughout their simulation box. They also look at the alignment for the resulting filamentary structure that arises in their 3D MHD simulations. They find that filaments are typically aligned parallel or perpendicular to the magnetic field when there is no turbulence injected into the gas. The ratio of parallel to perpendicular depends on the orientation of the shock with respect to the magnetic field. However, when turbulence is injected, the alignment is preferably parallel, but there is greater scatter in the offset angle between the filament and magnetic field directions. This again shows the importance of magnetic fields in forming dense CNM structures and the disruptive role increasing turbulence can play in the ISM.

1.3 Gaussian decomposition

Gaussian decomposition of HI spectra was pioneered in the late 1950s with the work of [Heeschen \(1955\)](#) in emission and [Muller \(1957\)](#) in absorption. [Heeschen \(1955\)](#) used the known formalism of the following equations:

$$T_B(\nu) = T_K \left(1 - e^{-\tau(\nu)}\right) \quad (1.5)$$

where $T_B(\nu)$ is the emission spectrum , brightness temperature (in K) as a function of velocity (ν), T_K is the kinetic temperature and:

$$\tau(\nu) = \tau_0 f(\nu) \quad (1.6)$$

where τ_0 is the peak optical depth of the HI absorption line and $f(\nu)$ represents the function describing the broadening of the 21cm line. They explored a Gaussian distribution for the broadening function as well as another function of the form that was posited by [van de Hulst et al. \(1954\)](#):

$$f(\nu) = \frac{1}{2\sigma} \exp\left(-\left|\frac{\nu - \nu_0}{\sigma}\right|\right). \quad (1.7)$$

This type of distribution was proposed to account for the presence of emission at higher velocities than would be expected from a Gaussian distribution, producing what was described as "wings". However, after analysis of both profiles they decided the Gaussian profile better describes the signal they are seeing for their 9 profiles fitted. They note the presence of these "wings" in their data as well, but maintain that overall, the Gaussian distribution is a better fit. Muller (1957) agreed with this conclusion as well in their work, noting the presence of the "wings" in their absorption profiles, but arguing that the Gaussian profile captures most of the important signal. They suggested observations with smaller bandwidths may eliminate the issue of the wings. So from the consensus of these studies, the representation of the absorption spectrum was:

$$\tau(\nu) = \tau_0 \exp\left(-\frac{(\nu - \nu_0)^2}{2\sigma^2}\right). \quad (1.8)$$

If more than one component is needed to fit the signal, the total spectrum can be expressed as the sum of N components:

$$\tau(\nu) = \sum_{n=1}^N \tau_{(0,n)} \exp\left(-\frac{(\nu - \nu_{(0,n)})^2}{2\sigma_n^2}\right). \quad (1.9)$$

With the ability of Gaussian decomposition to easily encode the HI line, the physical characteristics that the fitted parameters represent started to be used to uncover the 3D HI distribution of the local neighbourhood and beyond. If we look at the Gaussian distribution shown in Equation 1.8 we can assign some physical meaning to the fitted parameters. The value at which the Gaussian peaks (ν_0) would represent the bulk velocity of this HI and the dispersion of the Gaussian (σ) would represent the dispersion due to the random motions of the gas in this HI. The intrinsic width of the 21cm line is extremely small (Condon and Ransom 2016), so the dispersion is entirely due to the thermal and turbulent motions of the gas.

In the years following the first Gaussian decomposition efforts, further studies were done using the same technique. Using the velocity information they could get from Gaussian decomposition, Dieter (1965) catalogued the velocity populations of the HI in the north and south galactic poles. In this work they also determined that the lower limit to the dispersion they could fit was 1.2 km s^{-1} . This same limit in dispersion was echoed in the results from van Woerden et al. (1962) and Takakubo and van Woerden (1966). We can use this dispersion measure to estimate the temperature of the emitting gas. The maximum temperature possible from this dispersion measurement would be in the case where the dispersion is due entirely to thermal motions with no turbulence. This is likely unphysical but allows for an upper limit. The known relation between the kinetic temperature of the gas (T_k) and the dispersion (σ) is:

$$T_k = 121\sigma^2 \quad (1.10)$$

from Draine (2011). For the smallest dispersions measured in Dieter (1965) this would give a maximum kinetic temperature of 174K. Already at this early stage, CNM was being observed in emission spectra using Gaussian de-

composition.

While dispersion was an interesting property, many of these early studies remarked on the bulk velocities of the measured HI gas. Muller (1959) used Gaussian fitting on HI absorption to find cold HI in the Orion and Perseus arms of the MW towards Cas A. Dieter (1965) catalogued the bulk velocities of HI in the North and South galactic poles finding that there are two different velocity streams in the north, while there was only one in the south. They also identified several components in the south that had very large negative velocities, suggesting that these were high velocity clouds (HVCs), which will be detailed in Section 1.5. A comprehensive catalogue of HI components was compiled in Lindblad (1966) with the completion of the Dwingeloo Atlas survey which surveyed parts of the galactic plane and surrounding regions. In Takakubo and van Woerden (1966) they took a very methodical approach to Gaussian fitting. They used the combined emission and absorption equations (i.e. Equations 1.5 and 1.9) to fit their emission profiles. They had many sightlines where there were multiple, equally good solutions, so they had to use selection criteria. They note that the fitting process for a single line of sight took on the order of 5 hours! First, they looked at the residual spectra and rejected any solutions for which the residuals were evidently non-random. Second, they looked at the least squares values of the residuals and compared it to their expected values and rejected or tried to improve solutions that differed greatly from the expected value. Third, they looked at the number of components in each solution and opted for the solution with less components. As a last resort, if two or more solutions can not be differentiated between with the aforementioned criteria then they considered which solution was most similar to neighbouring sightline solutions. This paper was one of the first to clearly outline a set of criteria for distinguishing between degenerate solutions.

The very next year in Takakubo (1967) in a study of emission at intermediate galactic latitudes, divided the fitted components up into groups based on their fitted dispersions. In this work they divided them into groups S, M and, L corresponding to $\sigma < 3 \text{ km s}^{-1}$, $3 < \sigma < 7 \text{ km s}^{-1}$ and $\sigma > 7 \text{ km s}^{-1}$ respectively. They identified these 3 groups of HI signals and came to the conclusions that the S and M populations had similar properties and thus the M population was likely made up of blended S group components that were smeared in the beam. This type of division based on dispersion measurements is also seen in Heiles (1967), making out a background component as well as large and compact cloud components. In Mebold (1972) when they fit their profiles with Gaussian decomposition they found when they looked at column density as a function of velocity dispersion there was a bimodal distribution. This was another indication that the velocity dispersion measures have a physical significance for the properties of the gas. They make a comment at the end of the paper about the "Ambiguity of the Gaussian Analysis" that when they compare their dispersion distributions with that of Dieter (1965) and Takakubo and van Woerden (1966) that whilst they do not see the bimodal distribution in each case, they do see peaks at similar values. They argued that at that point with many studies, albeit at varying galactic latitudes, the fact that there were clear groupings of components in velocity dispersion that show different spatial distributions which suggest we were seeing the gas in different states. They praised the ability of Gaussian decomposition to provide solutions which can give physical information

about the state of the gas.

Evidently, Gaussian decomposition became accepted as a very convenient way to model a spectrum and analyse some physical characteristics of the gas being observed. However, it is not without its issues. One of the biggest problems facing Gaussian decomposition, even in the present day, is solution degeneracy. For a spectrum, especially one with many signals that have similar bulk velocities such that their signals overlap, there may be multiple solutions that fit the spectra equally well by whatever metric is being used. This ambiguity around what the correct solution is can be troublesome as we try to derive physical characteristics from these Gaussians. For example, [Shuter and Verschuur \(1964\)](#) found in their absorption observations that they were able to fit Gaussians with dispersions of less than 1 km s^{-1} with their high spectral resolution. However, [Dieter \(1965\)](#) commented on dependence of fitted velocity dispersions on the number of components that were fit. They point out that one of the spectra fitted by [Shuter and Verschuur \(1964\)](#) was also analysed by [Barrett \(1964\)](#) and each team came to a different solution. [Shuter and Verschuur \(1964\)](#) were only considering the HI spectra when fitting and fitting 3 components with dispersions of 4.0, 2.3 and 2.2 kc s^{-1} (note this is not km s^{-1} , but frequency units), whereas [Barrett \(1964\)](#) fit two components of 4.9 and 4.2 kc s^{-1} as they were using the indications from the OH spectral lines to inform their fit.

This ambiguity is often cited as the main weakness of Gaussian decomposition. Signals coming from different clouds of HI that are travelling at very similar velocities can be indistinguishable as separate components and may be interpreted as a single component, which could have a larger dispersion fit to it. Conversely, a negative dip in a broad emission feature caused by noise could give the illusion of a double component when it is not the case. This is particularly a problem in data with either poor spectral resolution, where components have to be quite separate from each other in velocity to be resolved, or poor physical resolution where components with a wide range of velocities are caught in the same beam and are blended together. This frustration with the ambiguity was highlighted in [Dickey and Lockman \(1990\)](#) where they stated "It is not generally useful to try to decompose emission spectra into Gaussian components because a profile's shape is determined as much by the vagaries of galactic rotation, velocity crowding, and streaming, modified by opacity and variations in temperature, as it is by the superposition of discrete elements." This is quite a pessimistic view of the usefulness of Gaussian decomposition, but a fair summary of the issues faced, especially when looking through the galactic plane, with emission spectra. They do follow up with a slightly more positive statement regarding anomalous emission signals with "A Gaussian analysis provides unambiguous information only when an emission feature has a very odd velocity (such as high-velocity clouds) or is very much brighter than its surroundings, so that it is not blended with most other emission." However, despite this warning on the limited usefulness of Gaussian decomposition in the 1990s, it has continued to be used, and can be a very powerful tool when used in conjunction with other measurements.

For example, the comparison of the fits in [Shuter and Verschuur \(1964\)](#) and [Barrett \(1964\)](#) by [Dieter \(1965\)](#) highlights the difference it can make to a solution when it is informed by data from other measurements. Other measurements can make all the difference in HI Gaussian decompositions as long as they are tracing the same structures

as the HI. So we can use other molecular emission lines like OH, or CO, but we can also use other HI measurements. Absorption measurements have already been mentioned, but I have not discussed the emission-absorption studies that have been done over the last few decades.

1.3.1 Fitting emission and absorption spectra

Emission-absorption studies have the ability to calculate the optical depth along a line of sight, which can help to recover the true column density of the HI and give clear information about the CNM. A study that detailed a formalism around the joint decomposition of absorption and emission source is [Heiles and Troland \(2003a\)](#). They fit the full radiative transfer equations, similar to Equations [1.5](#) and [1.9](#) but with an extra term (\mathcal{F}) that represents the fraction of WNM gas in front of the absorbing CNM cloud, thus $(1 - \mathcal{F})$ is the fraction of WNM gas behind the absorber. This introduces more variables into the equation, potentially increasing the chance of degenerate solutions, but provides a more physical model. In their work they did not allow this variable to vary continuously, but to only be 0, 0.5 or 1. The first step in their fitting process is to fit the absorption profile (aka the opacity profile) with the following equation:

$$\tau(\nu) = \sum_0^{N-1} \tau_{0,n} \exp\left(-\frac{(\nu - \nu_{0,n})^2}{\sigma_n^2}\right). \quad (1.11)$$

Then they fit a full emission spectrum using the CNM values that were fit. The emission spectra is a combination of emission from CNM and WNM, represented by the separate equations:

$$T_{B,CNM}(\nu) = \sum_0^{N-1} T_{S,n} (1 - \exp(-\tau_n(\nu))) \exp\left(-\sum_0^{M-1} \tau_m(\nu)\right). \quad (1.12)$$

$$T_{B,WNM}(\nu) = \sum_0^{K-1} [\mathcal{F}_k + (1 - \mathcal{F}_k) \exp(-\tau(\nu))] T_{0,k} \exp\left(-\frac{(\nu - \nu_{0,k})^2}{\sigma_k^2}\right) \quad (1.13)$$

So that they fit N components to the absorption, N CNM components to the CNM emission with M representing the number of CNM components that lie in-front of the n th cloud, as well as fitting K WNM components. This whole equation is fit together:

$$T_B(\nu) = T_{B,CNM}(\nu) + T_{B,WNM}(\nu). \quad (1.14)$$

There are a lot of variables in these equations and the number of components that were fit were subject to the judgement of the authors, who preferred to fit with the least number of Gaussians possible. Even with the most experienced eyes, this can sometimes lead to inconsistencies within datasets. Figure [1.3](#) shows an example of the emission-absorption spectrum from [Heiles and Troland \(2003a\)](#). This example has multiple absorptions features

which can then be seen to contribute to the emission spectrum at the same velocities, informing the fit.

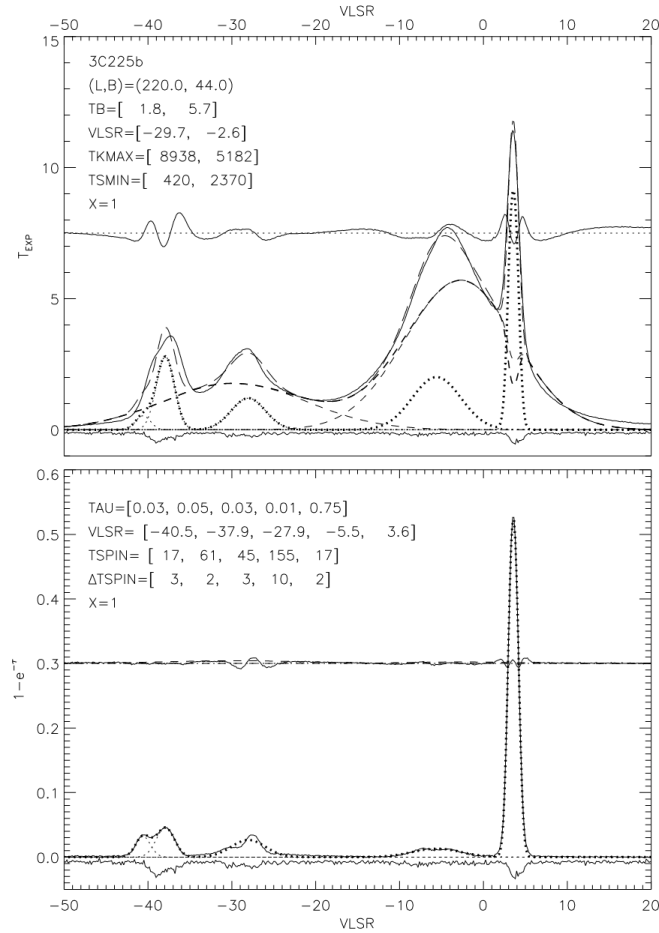


Figure 1.3: Figure 10 from [Heiles and Troland \(2003a\)](#) showing the emission (top) and absorption (bottom) profiles for source 3C225b. The dashed lines in the emission spectrum represent the individual and total WNM Gaussians fit to the emission. The Dotted lines in both spectra represent the individual and total CNM Gaussians fits to the emission and absorption. The solid lines above zero represent the data and residuals, below zero is the noise.

In a more recent emission-absorption study ([Murray et al. 2015a](#)) they use the same formalism as is described in [Heiles and Troland \(2003a\)](#) and above. They also include a new way of differentiating between multiple possible fits for a single emission-absorption pair. They test the improvement that an additional Gaussian component makes to the overall fit by calculating a confidence level in the new component, by comparing the reduced chi-squared values. Component addition conditions like these can help allay concerns about the possibility of over-fitting the data and solution reproducibility. The authors in this study remark that in the majority of cases, there is a point in the addition of components to a spectrum where the confidence level sharply declines and this is the indication of where to stop in Gaussian fitting process.

In [Dickey et al. \(2003\)](#) they use a Gaussian fitting technique only for the absorption spectrum. This study was of lines of sight through the Milky Way and had very complex and blended profiles, so it could be argued that Gaussian decomposition is not helpful in this context. Also, in a study of the SMC, [Dickey et al. \(2000\)](#) do not use Gaussian fitting

for the emission spectra, only the absorption, for identifying the central velocities of the absorption components. Instead they opted for an alternate method of extracting the important physical information from the absorption and emission spectra. So even when the spectra are not fit jointly like they are in [Heiles and Troland \(2003a\)](#) and [Murray et al. \(2015a\)](#), Gaussian decomposition can be used as a supporting tool in identifying the absorption lines.

1.3.2 Current Gaussian fitting algorithms

Many studies of HI spectra that used Gaussian decomposition to model spectra, developed their own Gaussian decomposition models for their dataset, but recently a few algorithms have been developed to be used by the community. The main ones are GAUSSPY+ from [Riener et al. \(2019\)](#) (GAUSSPY in [Lindner et al. \(2015\)](#) was an earlier version of this algorithm) and ROHSA from [Marchal et al. \(2019\)](#). These two algorithms take slightly different approaches.

GAUSSPY/GAUSSPY+ is an autonomous algorithm, meaning that the algorithm determines the number of components that are fit to a spectrum, not the user. It selects this number by identifying 'bumps' in the spectrum reliant on meeting various criteria for the values of the spectrum, its second, third and fourth derivatives at each point. This provides the number of components and guess for their central velocities. It then determines the initial guess for the widths from a relationship with the second derivative, and then uses these initial guesses to calculate initial guesses for amplitude. It is important to get the initial guess for spectrum close to the real solution as possible, as it will affect the solution that is found in the fitting process, especially on complex sightlines with many blended components. The key idea behind this algorithm is that the initial guess is the most important part of the Gaussian fitting process, and does the bulk of the work of determining the solution, the fitting process just refines it.

Because of this focus on the initial guess, it is imperative that the high order derivatives of the spectra are usable. One of the difficulties in using high order derivatives are the effects of noise. Noisy data will produce noisy derivatives which will in turn, produce false component detections. They tackle this by using a smoothing equation, for which the smoothing parameters are obtained through machine learning on a training dataset. A Gaussian kernel set by the smoothing parameters is used to de-noise the data before the derivatives are calculated. This extra step of supervised machine learning, requires the user to provide a few hundred well-fit spectra. For HI emission, some areas of the sky are more complex than others in terms of the spectra. This means that the training dataset would have to have a comparable complexity to the dataset to be analysed, for the smoothing parameters to be appropriate. The training dataset they use in [Lindner et al. \(2015\)](#) is synthetic data produced from randomly selecting components from the Millennium Arecibo survey ([Heiles and Troland 2003b](#)) which looks at points at galactic latitudes $|b| > 10^\circ$. The data fit by the GAUSSPY is from the 21 SPONGE survey [Murray et al. \(2015a\)](#) which is also a study of high latitude sightlines. In that work they utilised GAUSSPY to fit absorption spectra from the 21 SPONGE survey. So it is a versatile tool that can fit both emission and absorption spectra separately.

Once an appropriate training set has been produced and the smoothing parameters determined, GAUSSPY is

a very quick algorithm. It fits a single spectrum in roughly a second (the precise time depends on the number of spectral channels and the number of components fit) and each spectrum is fit individually, so parallelisation of the process could lead to very short computation times for large datacubes.

GAUSSPY+ is an enhancement of the original GAUSSPY algorithm, designed specifically with CO spectra in mind. It makes improvements on the original by estimating the noise, determining the smoothing parameters automatically, introducing quality control for Gaussian components, and introducing the option for regularisation after the initial fit. Noise estimation is important to determining the signal to noise level in data. This is achieved in this algorithm by determining the areas that are likely noise and likely not noise in the spectra based on probability. The smoothing parameters can be automatically determined using the dataset that the user wants to fit, by allowing GAUSSPY+ to fit a subset to train on. During the main fitting process each component is tested to see if it meets various acceptability criteria, including looking at neighbouring pixels to see if the number of components fit is consistent. After all of this, there is the opportunity to refit the solutions with variable weighting given to the nearest neighbour pixels, by checking the spatial consistency with the neighbouring sightlines. Figure 1.4 shows an Figure 17a and 17d from Riener et al. (2019) which compares the improvement and more spatially consistent solutions obtained with GAUSSPY+ compared to the original GAUSSPY algorithm, tested on data from the Galactic Ring Survey (GRS) in CO (Jackson et al. 2006).

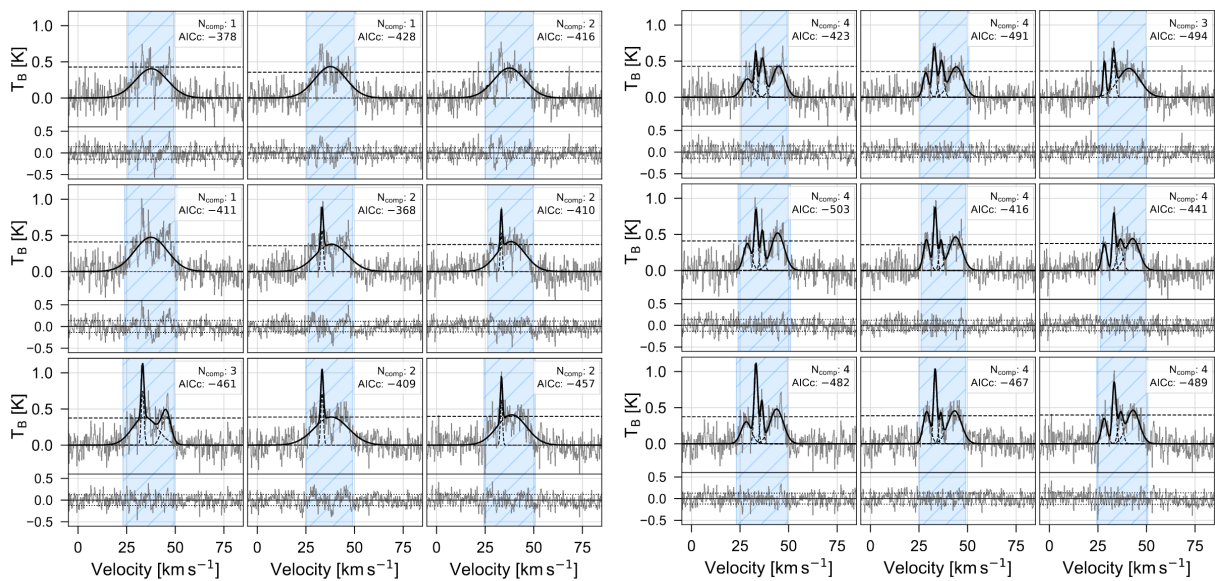


Figure 1.4: Figure 17a (left) and 17d (right) from Riener et al. (2019) comparing the same section of GRS CO data fit with GAUSSPY (left) and GAUSSPY+ after both refitting stages (right).

This is a more complex implementation of Gaussian fitting, which does require the user to set certain thresholds related to signal-to-noise that determine the fitting parameters. The acceptance criteria of fit components relies on these thresholds which will lead to rejections of weak components in noisy data. This would particularly affect weak CNM components, as the rejection criteria is based on integrated column density of the component. It also does

not strictly enforce regularisation in the Gaussian component parameters; amplitude, central velocity and velocity dispersion. However, it does well in estimating noise, smoothing the data and allows for quick decomposition.

Another algorithm that is available to the community is ROHSA (Regularised Optimisation for Hyper-Spectral Data). It takes a top-down approach to the task of Gaussian decomposition. It starts at the highest level, the mean spectrum of the entire datacube, and fits the initial spectrum to this. It then segments the cube into 2^i grids at each level i (ranging from 0 to the value determined by $\log_2(\max(X, Y))$ where X and Y are the number of pixels in the two spatial axes of the datacube). For each newly segmented grid a Gaussian solution is fit, using the solution of the level before it as the initial guess. Additionally, the cost function includes not just the typical least squares term that is often used in this type of fitting, but also additional terms that enforce regularisation. These terms each quantify the offset of one the three terms in each Gaussian between neighbouring solutions. If these terms are large, the minimiser will work to reduce the offsets, pushing towards a solution with higher regularisation. The size of these terms are governed by hyperparameters, which are set by the user. These hyperparameters are multipliers that control the influence of each regularisation terms, one for each of the Gaussian term, amplitude, central velocity and velocity dispersion, and another that controls the deviation around the mean velocity dispersion of each Gaussian components across the whole field. An example of spectra from a regularised solution is shown in Figure 1.5 and highlights how the components retain their order across adjacent lines of sight, as the shapes of each Gaussian also vary slowly over the 4x4 grid.

The number of Gaussians is also a parameter that has to be set by the user before running the fitting algorithm. All spectra in the cube have the same number of Gaussians fit to them, however any Gaussians fit to a spectrum can have an amplitude of zero if that is optimal for the fit, so not all Gaussians will be 'active' in the fit of every spectrum in the datacube. One of the concerns that arises in Gaussian fitting is overfitting, which happens when there are too many Gaussians fit to a spectrum and some may worry about this with the way the ROHSA algorithm works. However, if the regularisation is set to a reasonable level, the restriction on the central velocity means the minimiser is more likely to converge to a solution where Gaussians have amplitudes of zero, rather than a solution where the central velocity changes drastically to over-fit the signal in a different part of the spectrum.

One of the other problems with fitting a large number of Gaussians, other than over-fitting, is the computational time. This method already does multiple levels of fitting and fits every segment in each level simultaneously. This takes a lot of time, and increasing the number of Gaussians only adds to that. Additionally there is no obvious value for the number of Gaussians or for the value of the hyperparameters in the initial fitting stage, as this depends on the physical scales and complexity of the data. A user may be able to make an informed decision on likely values for the hyperparameters based on previous fields analysed with this tool (e.g. Marchal et al. (2019), Vujeva et al. (2023), Taank et al. (2022) and others), however a range of parameter values and number of Gaussians should be assessed before the user can be confident in their solution. This means this method is computationally expensive, but very physically motivated.

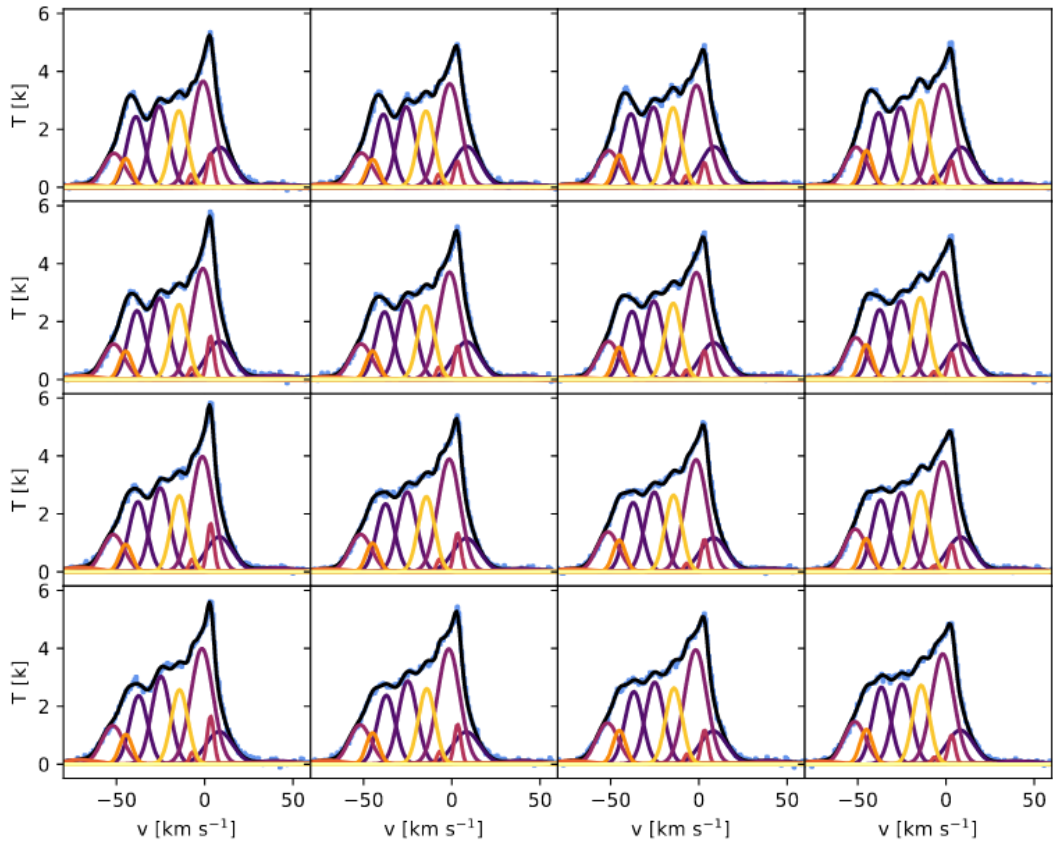


Figure 1.5: Figure 15 from [Marchal et al. \(2019\)](#) that shows a random 4x4 grid of the fit to observed HI data. The colour of each component indicates its order.

Both of these methods have proved their ability to accurately describe synthetic data in their respective introductory papers, and have been used in various studies since 2019. They have their advantages and disadvantages. GAUSSPY+ does most of the difficult work for the user in calculating the noise, smoothing the data, finding the number of initial guesses, and is fast by fitting all the spectra individually. It also has the re-fitting option that looks at points where the fit is inconsistent across the image, however it does not actively enforce regularisation across an image. ROHSA is slower due to the nature of its iterative approach and requires the user to fine-tune the hyper-parameters, but its method helps break the degeneracy of solutions along individual spectra by enforcing spatial regularisation in the cost function. I use ROHSA in this work, detailed in Chapter [2](#).

1.4 The Magellanic System

The Magellanic system consists of the main two clouds, the Large and Small Magellanic Clouds (LMC and SMC) both of which are dwarf galaxies. They are very prominent features of the southern sky, and were known to residents of the southern hemisphere long before their noted observation in the 1500s by Portuguese explorer Ferdinand Magellan (Fernão de Magalhães), for whom they are named. For example, to the Yirrkala communities in Arnhem land,

Australia, the LMC and SMC represent two sisters, the older and younger respectively (Noon 2022) who are periodically separated as the LMC dips below the horizon for part of the year. This separation and reunification of the older sister with the younger signals the change of season for these communities. For the Pitjantjatjara communities in Central Australia, the clouds represent two brothers, the Kungara brothers who decide upon the fate of a dying person's spirit (Orchiston 2000). In the Māori language, from Aotearoa (New Zealand) there are many recorded names for the Magellanic clouds, for example Pātari-rangi and Pātari-kaihou for the LMC and SMC respectively (Orchiston 2000).

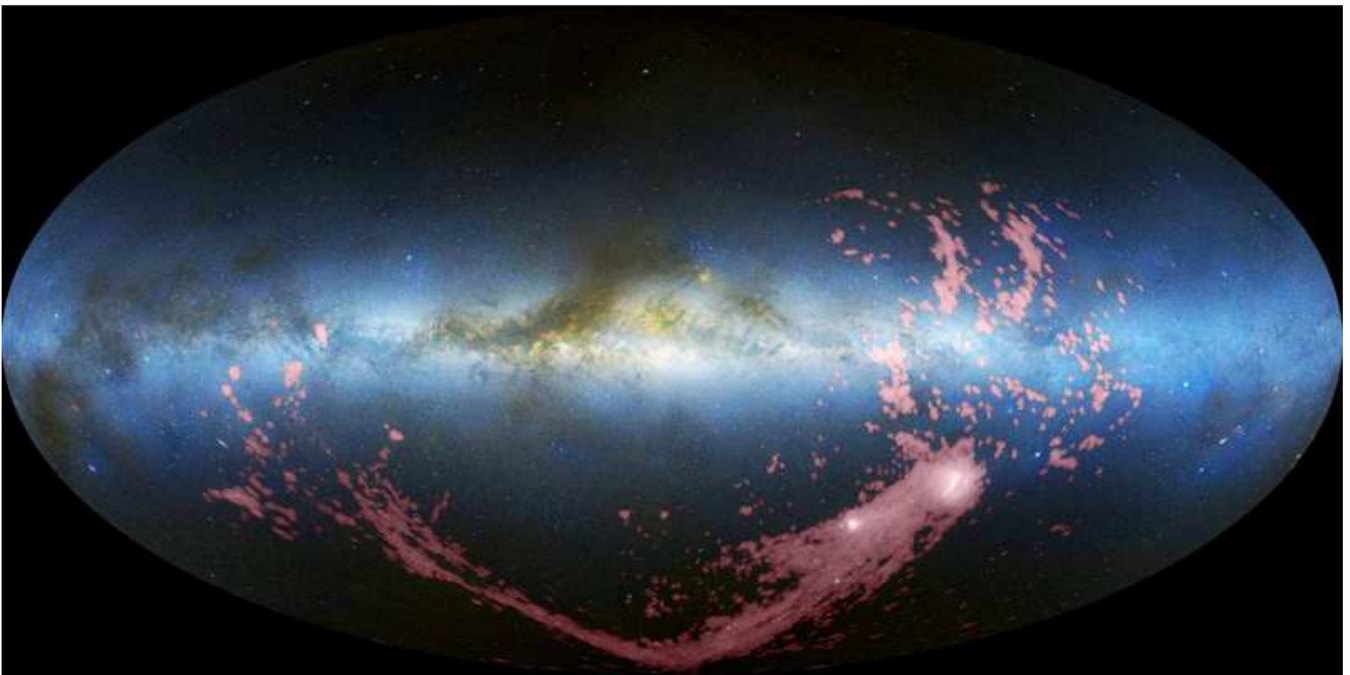


Figure 1.6: Figure 9 from Nidever et al. (2010) which shows the optical extent of the whole sky with the HI column density of the Magellanic system overlaid in red.

Observations made by the the naked-eye or optical telescopes can only observe the stellar population of the Magellanic clouds. The stellar mass of the large and small cloud is $3 \times 10^9 M_{\odot}$ and $3 \times 10^8 M_{\odot}$ respectively (D'Onghia and Fox 2016), a factor of 10 difference. This, as well as the difference in visual diameter of 6° and 2.5° (Bok 1966) points at why they were labelled as large and small based on these initial observations. However, when astronomers began to observe the galaxies in other wavelengths, a wealth of hidden structure was revealed. Most notably, the radio HI content of the Clouds clearly extends a lot further spatially than the optical, shown in Figure 1.6. In Figure 1.6 the Magellanic clouds are the two white blobs seen embedded in the red HI column density, south of the Galactic plane. The SMC is to the west and the LMC to the east. It is evident that not only does the HI structure extend further out from the centre of each Cloud, but there is a lot of structure in the HI that is not traced by dense stellar populations. Generally, the Magellanic system can broken up into 5 distinct structures; The LMC, The SMC, The Magellanic Bridge, The Magellanic Stream and The Leading Arm.

However, it may not be useful to say these structures are 'distinct' because they actually form part of a continuous system. The entire system has a continuous velocity structure which is demonstrated in the detailed HI map of the whole system from the GASS project (McClure-Griffiths et al. 2009) in Figure 1.7 which shows the smooth velocity gradient from -324 km s^{-1} up to 400 km s^{-1} .

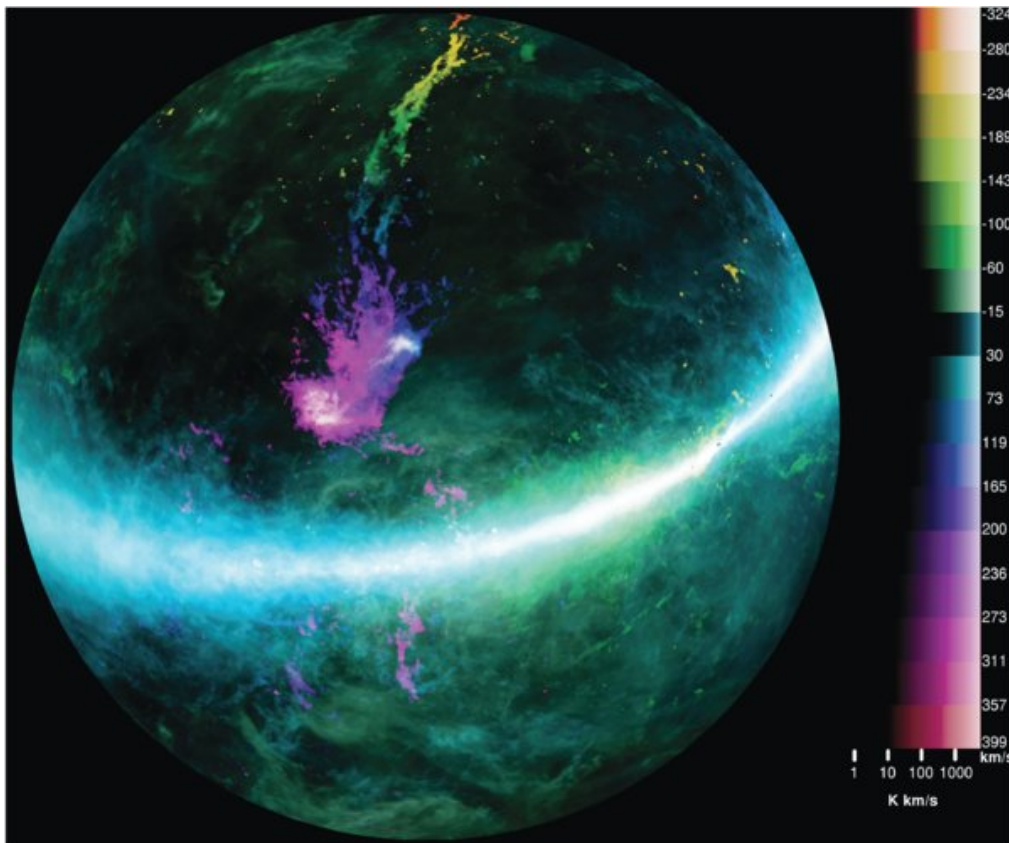


Figure 1.7: Figure 4 from McClure-Griffiths et al. (2009) that shows the total HI column density of consecutive velocity bins along the Magellanic system. Velocity bins are indicated by the colourscale.

This large velocity gradient and extended physical extent on the sky show us that the Magellanic clouds are interacting with one another and that interaction has thrown around a lot of neutral gas its wake. The largest of this is the Magellanic Stream, which was first hinted at with the detection of anomalous clouds around the south galactic pole in Dieter (1965) and then a number of clouds that were found to form a linear structure in Wannier and Wrixon (1972). These studies were limited to what could be observed from the northern US, mainly the northern sky. So a study was conducted with the Parkes (Murriyang) telescope in the Southern Hemisphere (Mathewson et al. 1974) that confirmed the continuous stream of gas that connected the clouds found in the more northern parts to the Magellanic clouds. Mathewson et al. (1974) suggest in their work that the Stream originated from the interaction of the Magellanic clouds and ponder many situations that could have caused the gas to emanate from the SMC in the manner observed.

One of the situations pondered in Mathewson et al. (1974) was the formation of the Stream from the interaction

of the Magellanic Clouds with the Milky Way. Interaction with the gravitational potential of our galaxy, a galaxy much larger than either Cloud, would draw out a lot of gas and stars according to Clutton-Brock (1972), potentially in a manner similar to that of the Stream.

1.4.1 The Large Magellanic Cloud

The Large Magellanic Cloud is the most massive part of the Magellanic system in both stellar and HI mass (D’Onghia and Fox 2016). It sits at a distance of $49.89^{+2.11}_{-2.03}$ kpc from us (de Grijs et al. 2014) and is classified as an irregular galaxy, of Magellanic subtype, a genre it defined, which has a bar and single spiral arm (De Vaucouleurs and Freeman 1972). Sparke and Gallagher (2000) measure the inclination of the stellar disc to be $i = 45^\circ$, Cullinane et al. (2022) measure the inclination to be $i \approx 36.5^\circ$ in the north-eastern, unperturbed outskirts (LMC galactocentric radii of 8° - 11°). It becomes more disturbed in the south-western outskirts due to the fact that at those radii the outskirts of the LMC overlap with the SMC. The metallicity of the galaxy is ≈ 0.5 of the solar value (Rolleston et al. 2002). The typical line of sight velocities for the LMC range from 189.5 km s^{-1} to 359.0 km s^{-1} (Staveley-Smith et al. 2003) with some sparse extended emission up to 425 km s^{-1} and down to 100 km s^{-1} . The peak column density of HI found in this same study was $5.6 \times 10^{21} \text{ cm}^{-2}$ and the diameter of the galaxy when measured from the HI was found to be 9.3 kpc. The most recent HI absorption survey found a cold gas fraction of 0.35 (Marx-Zimmer et al. 2000). The LMC is also connected to two other features of the larger system, the Magellanic Bridge to the south-west and the Leading Arm to the south, when the galaxy is projected in equatorial coordinates, such as in Figure 1.8 from Staveley-Smith et al. (2003) which shows the main features of the LMC.

1.4.2 The Small Magellanic Cloud

The Small Magellanic Cloud is the second most massive part of the Magellanic system by stellar and HI mass (D’Onghia and Fox 2016). It sits at a distance of 63 ± 5 kpc from us (Di Teodoro et al. 2019b). It is also classified as an irregular galaxy. It has a bar, which is the most obvious part of the galaxy in the optical wavelengths, as that is where much of the stellar population resides. It also has a ‘wing’ (Stanimirović et al. 2004) which extends over towards the Magellanic Bridge which sits east of the SMC. The bar and wing are clearly outlined in the solid and dashed rectangles of Figure 1.9. The galaxy itself is highly disturbed, likely by the pull of the LMC, and thus does not behave like an ideal rotating galaxy, making it difficult to measure the orientation of the galaxy. Inclination measured from the HI was calculated in Di Teodoro et al. (2019b) was $i = 51 \pm 9^\circ$. Inclination measured from the distribution of intermediate age stars was $i = 35.4 \pm 1.8^\circ$ (Tatton et al. 2021) but seems to vary with the age of the stellar population, as they also report very shallow inclinations for old stars and higher inclinations for younger stars.

This confusion around the orientation of the SMC is only furthered by the seemingly large line of sight depths through the cloud. In Muraveva et al. (2018) they look at the structure of the galaxy through RR Lyrae stars which

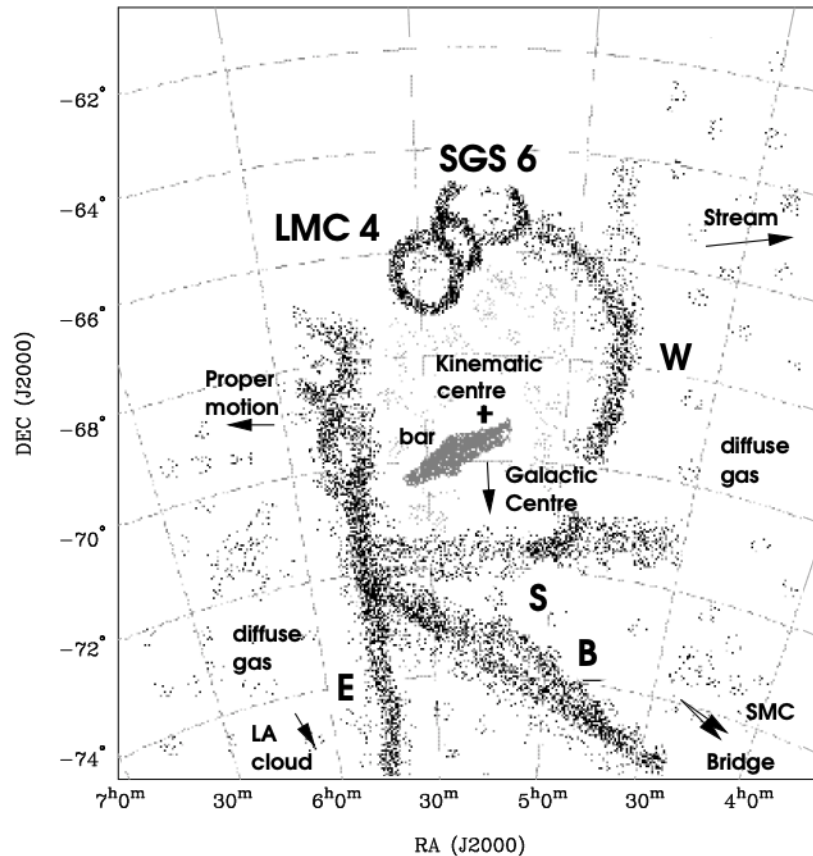


Figure 1.8: Figure 1 from [Staveley-Smith et al. \(2003\)](#) that highlights the main features of the LMC in relation to the Milky Way and larger Magellanic system.

show line of sight depths ranging from 1-10 kpc with the centre of the galaxy having the largest depths. They also conclude that the eastern side of the galaxy is closer than the western side. This is consistent with the notion that the SMC and LMC are physically connected, as the LMC is closer than the SMC and the eastern side of the SMC is closest to the LMC. Another study from the same VMC survey ([Ripepi et al. 2017](#)) looked at the distances measured from Classical Cepheids (CCs) and found a difference in the spread of distances along the line of sight between the older and younger population of CCs segmented at an age of 140Myr. The older population had a larger spread of distances, indicating they have been more affected by the tidal interaction with the LMC. Overall the population of CCs showed an total elongation of 25-30 kpc across the galaxy.

The metallicity of the galaxy is 0.2 solar ([Russell and Dopita 1992](#)). The typical line of sight velocities for the SMC range from $\sim 100 \text{ km s}^{-1}$ to $\sim 200 \text{ km s}^{-1}$ ([Staveley-Smith et al. 1997](#), [Pingel et al. 2022](#)). The peak column density from the latest HI emission survey ([Pingel et al. 2022](#)) is around $7.5 \times 10^{21} \text{ cm}^{-2}$. Absorption studies of the SMC have put constraints on the cold gas fraction of the galaxy and gave values from 0.07-0.2, however it is important to note that these are mean values for the whole SMC. Additionally, two of the three surveys ([Dickey et al. 2000](#) and [Jameson et al. 2019](#)) are targeted surveys which may inflate the value of the cold gas fraction. The latest survey ([Dempsey](#)

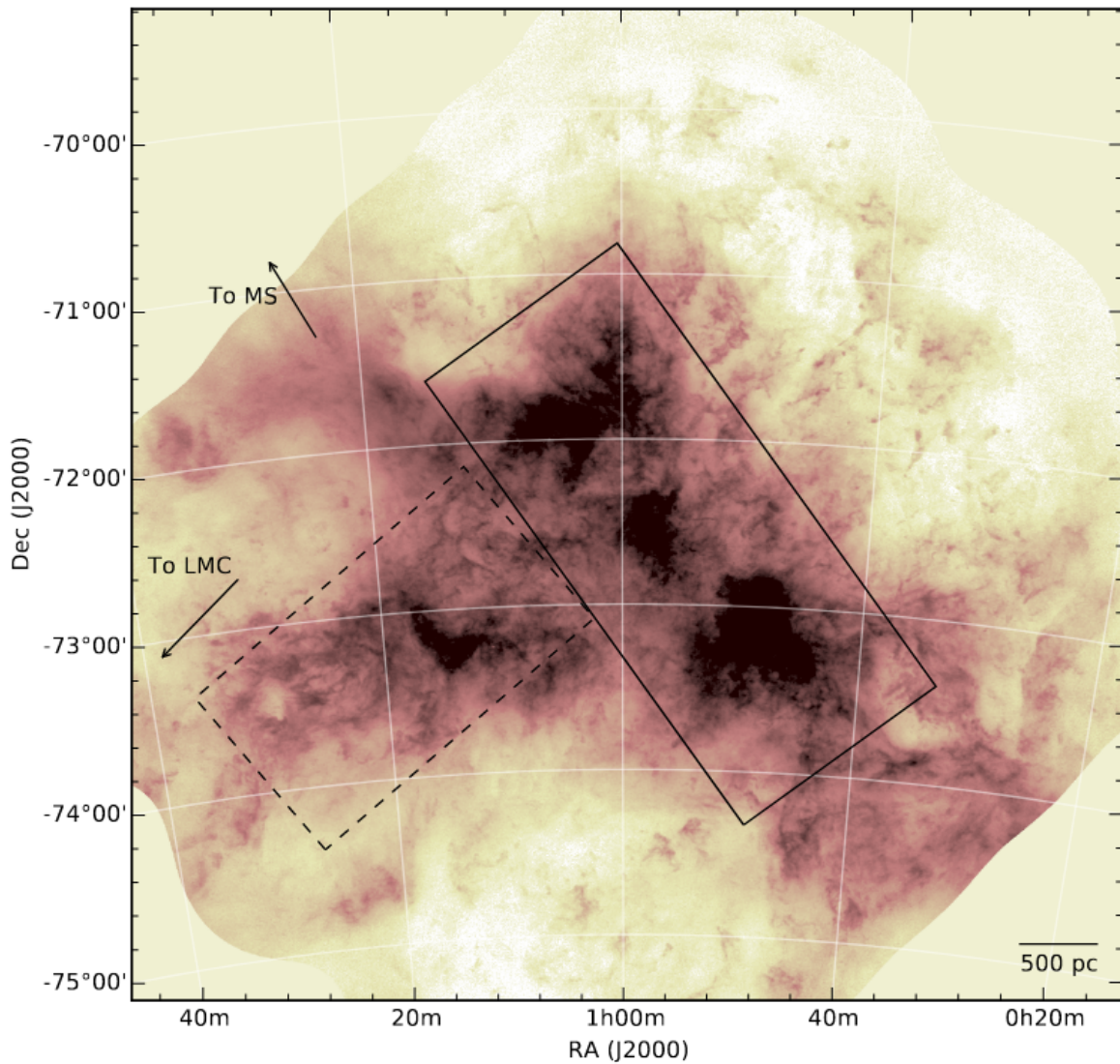


Figure 1.9: Figure 1 from [McClure-Griffiths et al. \(2018\)](#) that shows the peak HI emission in the SMC from ASKAP commissioning data. It highlights the two main features of the SMC; the bar, enclosed by the solid rectangle, and the wing, enclosed by the dashed rectangle. It also indicates the directions to the Stream and LMC/Bridge.

[et al. \(2022\)](#), is an un-targeted search that is limited primarily by source flux and covers more sources than the other surveys combined, they report a 0.11 median cold gas fraction.

1.4.3 The Magellanic Bridge, Stream and Leading Arm

The Magellanic Bridge was first identified in HI emission by [Hindman et al. \(1963\)](#). It is the clear HI link that spans the distance between the two main Clouds, clearly labelled in Figure [1.10](#). From analysis of the HI velocities of the Bridge, [McGee and Newton \(1986\)](#) determined that the Bridge has two separate components travelling at radial velocities of 214 and 238 km s^{-1} . They also found some other components below 200 km s^{-1} that seemed to connect the Bridge to the SMC. The total HI gas mass of the Bridge is $1.84 \times 10^8 M_{\odot}$ ([Bruens et al. \(2005\)](#)) and the stellar mass is most

recently measured from clusters to be $3-5 \times 10^5 M_{\odot}$ (Oliveira et al. 2023). The metallicity of the Bridge is closer to that of the SMC, measured by absorption towards a QSO (Misawa et al. 2009) to be $\sim 0.1-0.3$ solar. Recent observation of the stellar distribution through the Bridge in Oliveira et al. (2023) show an old population of stars with metallicities closer to that of the SMC and a younger population of stars with near constant metallicity of ~ 0.4 solar. Skowron et al. (2014) also find a young population of stars that span the length of the Bridge, with particular over-densities near the SMC and halfway along the Bridge. The interaction of the SMC and LMC that led to the formation of the Bridge is expected to have drawn out stars that belong to both galaxies, but evidently the gas drawn into the Bridge has allowed it form stars in situ, thus why we see these different stellar populations. This star formation activity in the Bridge is supported by the observation of CNM in Kobulnicky and Dickey (1999) and molecular species such as CO (Mizuno et al. 2006; Valdivia-Mena et al. 2020) and HCO+ (Murray et al. 2015b). Overall, the Magellanic Bridge is an interesting tidal feature of the Magellanic system, pulling material between the Magellanic Clouds and forming its own stars in the process.

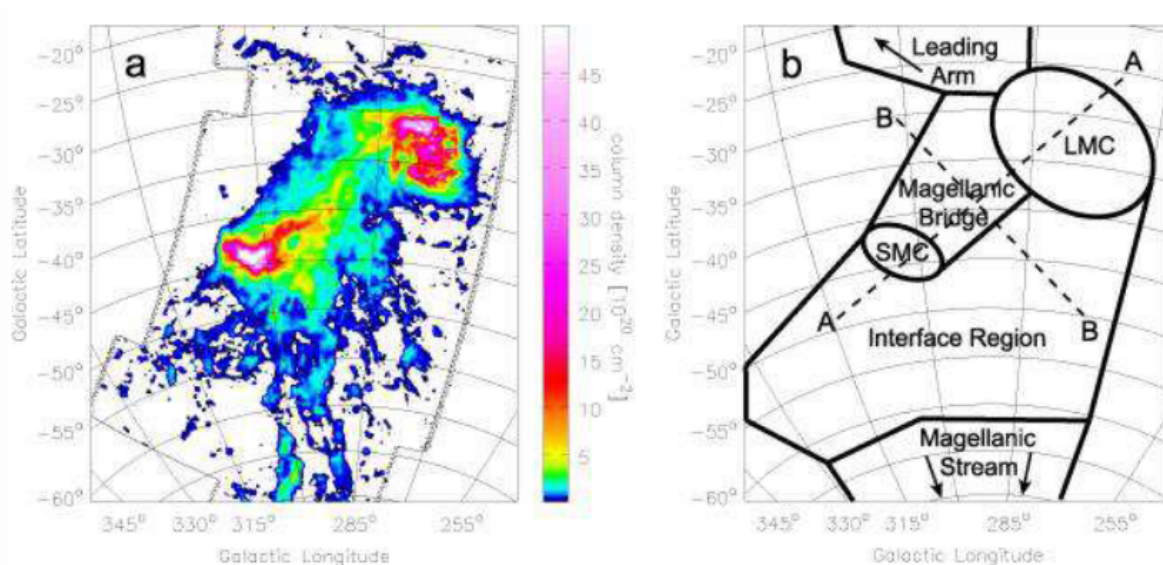


Figure 1.10: Figure 4a and 4b from Bruens et al. (2005) which show the total column density of the Magellanic system (left) and the different sections of the system (right).

The Magellanic Stream is spread over 100s of degrees of sky. The first time it was catalogued in-depth there were six concentrations of gas labelled MS I-VI with increasing distance away from the Magellanic Clouds (Mathewson et al. 1974). Newer surveys with better resolution and sensitivity have revealed more HI connecting these larger density enhancements (Bruens et al. 2005; Putman et al. 2003). Just looking at the projection of the Stream on the plane of the sky, it would appear to be one long filament, but analysis of the HI data by Putman et al. (2003) identified two main filaments that make up the Stream which seem to twist around each other. They suggest that these two filaments originate from different points in the Magellanic Clouds, one from the SMC and one from the Bridge, inheriting the typical velocities of their origin. Kinematic analysis in Nidever et al. (2008) used Gaussian decomposition to trace

one of the two filaments back to the LMC, more precisely the region including 30 Doradus, the main LMC starburst region. Chemical analysis of the typical metallicities in the two different filaments in [Fox et al. \(2013\)](#) and [Richter et al. \(2013\)](#) show a difference in metallicity between the filaments. The filament traced back to the LMC in [Nidever et al. \(2008\)](#) shows a metallicity of 0.5, typical of the LMC and supports the theory of an LMC origin. The other filament shows a metallicity of 0.1 solar, more in line with the typical SMC metallicity of 0.2. The bulk column density decreases linearly along the length of the Stream. Unlike the Bridge, the Stream was thought to have no stellar objects associated with it, however a recent spectroscopic survey in [Chandra et al. \(2023\)](#) has confidently associated 13 stars with the Magellanic Stream, indicating that there is a stellar counterpart to the neutral gas.

The Leading Arm of the Magellanic system is a collection of neutral gas that appears to emanate from the LMC and is the counterpart to the Stream at the other end of the system. It stems from the LMC and crosses the Milky Way, becoming more disrupted in this crossing, breaking up into many smaller clouds. It was first linked to the Magellanic system in [Putman et al. \(1998\)](#) and that fact that it precedes the Magellanic Clouds in their motion around the MW indicated that the structure of the Magellanic system was not created by ram-pressure stripping, but by tidal interactions. It has a total HI mass of $3.0 \times 10^7 M_{\odot}$ ([Bruens et al. 2005](#)). [Venzmer et al. \(2012\)](#) identifies four main sections of the Leading Arm, where the sections approaching the MW Galactic Plane are decelerating. They see head-tail signatures in many of the clouds that make up the Leading Arm, indicating that they are interacting with the Galactic Halo of the Milky Way. There is also evidence of in-situ star formation with a population of young stars identified in [Casetti-Dinescu et al. \(2014\)](#). More recently, [Richter et al. \(2018\)](#) analysed the gas phases abundances of various metals and concluded that the chemical make-up of LA II (the second section of the Leading Arm) indicates that it likely originated from the SMC, but does not completely rule out an LMC origin.

1.4.4 Magellanic Modelling

There have been attempts to model the Magellanic system to attempt to understand how it came to be the massive structure we see today. [Besla et al. \(2012\)](#) model a scenario in which the SMC and LMC are on their first pass around the MW. They set the SMC to have an eccentric orbit around the LMC and allow the SMC to orbit around the LMC twice and thrice, defining two different models that are then placed in the MW potential. The results after letting these systems evolve in the MW potential are shown in [Figure 1.11](#) which do resemble the distribution of gas seen in [Figure 1.6](#).

However, both models fail to reproduce the mass and bifurcation of the Stream, the location of the Leading Arm and the current star formation rate of both the LMC and SMC. The goal of this study was not to precisely reconstruct the entirety of the Magellanic system, as they did not explore the full range of initial conditions to make a determination between multiple models. Instead, what this study showed was that the second model, where the SMC had orbited three times and directly collided with the LMC, better reflected the dynamics observed in the LMC.

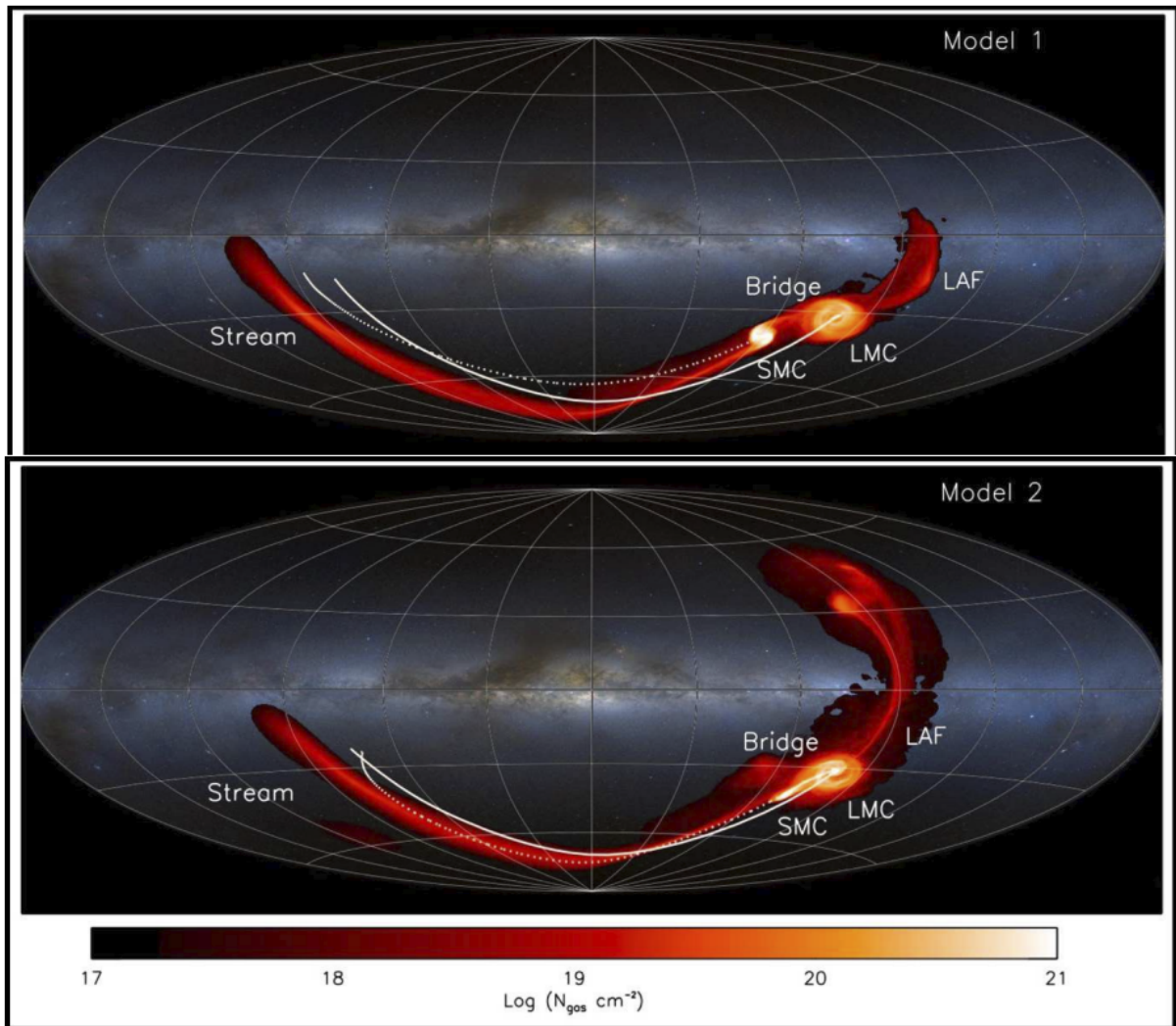


Figure 1.11: Figure 6 from Besla et al. (2012) which shows the final gas distribution for the two different models considered for the evolution of the Magellanic system. The orbits of the LMC and SMC in each model are indicated by the solid and dotted lines respectively.

Diaz and Bekki (2011) proposed a model whereby the LMC and SMC originated as separate MW satellites that were in a bound orbit around our galaxy. They posit that eventually the LMC captured the SMC around 1.6 Gyr ago and they have made two passes since then, resulting the Stream and Leading Arm features. This model, shown in Figure 1.12 reproduces the bifurcation of the Magellanic Stream better than that of Besla et al. (2012). They suggest the potential of the Milky Way plays an important role in elongating the Leading Arm and Magellanic Stream, which is possibly why first-pass models where the LMC and SMC spend less time in the stronger MW potential, cannot replicate these features as convincingly. The passages of the LMC and SMC around each other within the last 1.6 Gyr, suggested in this model, could be correlated to the star formation bursts around 0.5 and 2 Gyr ago.

Ram-pressure stripping as the mechanism responsible for forming the Stream has been investigated and can remove a lot of material from the LMC (Mastropietro et al. 2005). However, the presence of the Bridge really points

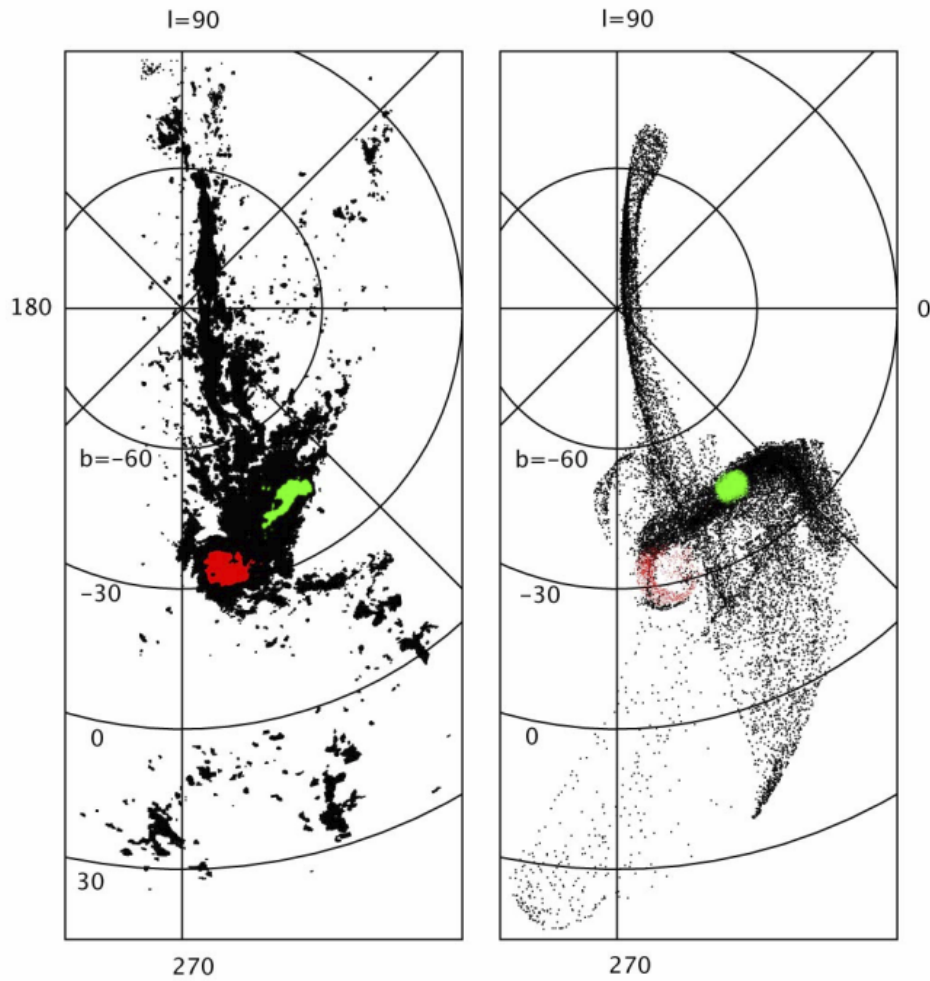


Figure 1.12: Figure 2 from [Diaz and Bekki \(2011\)](#) showing the HI observations of [Putman et al. \(2003\)](#) and [Bruens et al. \(2005\)](#) (left) and the final distribution of the test particles from the [Diaz and Bekki \(2011\)](#) model (right).

to a tidal interaction occurring between the clouds as well. A recent study by [Wang et al. \(2022\)](#) shows that a model which combines the collision of the two clouds with the ram-pressure forces from the passage through the hot Milky Way halo, most accurately reproduces the morphology of the Stream compared to the other tidal models they consider. They considered the model from [Wang et al. \(2019\)](#) and compared with the more recent observational constraints of the Magellanic system, namely the stellar proper motions from Gaia EDR3 ([Luri et al. 2021](#)).

Overall, it seems that combining the various effects of the tidal interaction of the SMC and LMC with each other whilst taking into account their motions around the MW and the effect of its halo, we are getting closer to an understanding of how the Magellanic System formed and where it might be going.

1.5 High/Intermediate velocity clouds

High velocity clouds (HVCs) have been an astronomical puzzle for decades. A large study of stellar spectra towards high-latitude stars from [Münch and Zirin \(1961\)](#) was one of the first times that various components identified in absorption were referred to as high-velocity clouds. At this time, "high-velocity" was defined as components that differed by more than 24 km s^{-1} from the velocity expected from Galactic rotation. In [Spitzer \(1956\)](#) they suggested that neutral hydrogen may be found at great distance and speeds from the Galactic plane and surveys were undertaken with radio telescopes to uncover this. It was not until 1963 that there were detections of neutral hydrogen at radial velocities of -116 to -174 s^{-1} ([Muller et al. 1963](#)). In 1966, studies were published from teams in The Netherlands cataloguing the HVCs ([Hulsbosch and Raimond 1966](#)) and intermediate velocity clouds (IVCs), ([Blaauw and Tolbert 1966](#)). The velocity at which the distinction was made between the two categories was $\pm 70 \text{ km s}^{-1}$.

Typically in the present day, HVCs are defined as clouds with a deviation velocity (v_{DEV}) where $|v_{\text{DEV}}| > 90 - 100 \text{ km s}^{-1}$. IVCs are defined as having a v_{DEV} where $40 < |v_{\text{DEV}}| < 90 \text{ km s}^{-1}$ ([Wakker 2004](#)). The deviation velocity is defined as the difference between the cloud radial velocity and the radial velocity due to the differential rotation of the Milky Way. It is defined in [Wakker \(2004\)](#) as:

$$v_{\text{DEV}} = v_{\text{LSR}} - v_{\text{min}} \quad (1.15)$$

if $v_{\text{LSR}} < 0$, or:

$$v_{\text{DEV}} = v_{\text{LSR}} - v_{\text{max}} \quad (1.16)$$

if $v_{\text{LSR}} > 0$. Where v_{min} and v_{max} are the minimum and maximum velocities compatible with the model of galactic rotation.

These early observations in the 1960s were the beginning of a field of study that continues today both in observational studies that aim to map these clouds in our galaxy and beyond, and simulations that try and constrain how these clouds form in and around galaxies.

1.5.1 Observations

Successive observations have expanded our catalogue of the HVCs around the MW as the sensitivity of surveys has improved. HVCs are easily seen in HI, so the most commonly used line for detection is the 21cm line. This means that large HI surveys are great datasets for observers looking to catalogue these clouds. Given the highly deviant velocities of some HVCs, the surveys need to cover a wide range of velocities. Around the turn of the century, two complementary HI surveys of the northern and southern skies, that of [Hartmann and Burton \(1997\)](#) and [Arnal et al. \(2000\)](#), were conducted, providing a opportunity to expand the often cited catalogue compiled by [Wakker and van Woerden \(1991\)](#) from the HI surveys of [Hulsbosch and Wakker \(1988\)](#) and [Bajaja et al. \(1985\)](#).

de Heij et al. (2002) used the Hartmann and Burton (1997) LDS survey to find 67 compact HVCs. Morras et al. (2000) analysed the data from Arnal et al. (2000) to uncover 6846 high velocity objects in the southern sky (south of -25° declination). Putman et al. (2002) reprocessed the data from the HIPASS survey (Barnes et al. 2001) to recover the large angular scales and uncovered 1956 high velocity objects. A combination of the surveys completed by Hartmann and Burton (1997) and Arnal et al. (2000) were reprocessed by Kalberla et al. (2005) to create the LAB survey, which could be assessed for HVCs.

Many HVCs are part of complexes, which are groups of clouds that form part of a larger structure. These can take up large areas on the sky and many are apparent and labelled in Figure 1.13. Figure 1.14 shows the HVC distribution produced by Westmeier (2018) from the HI4PI (HI4PI Collaboration and et al. 2016) data selected by using a deviation velocity of $|v_{\text{DEV}}| > 70 \text{ km s}^{-1}$.

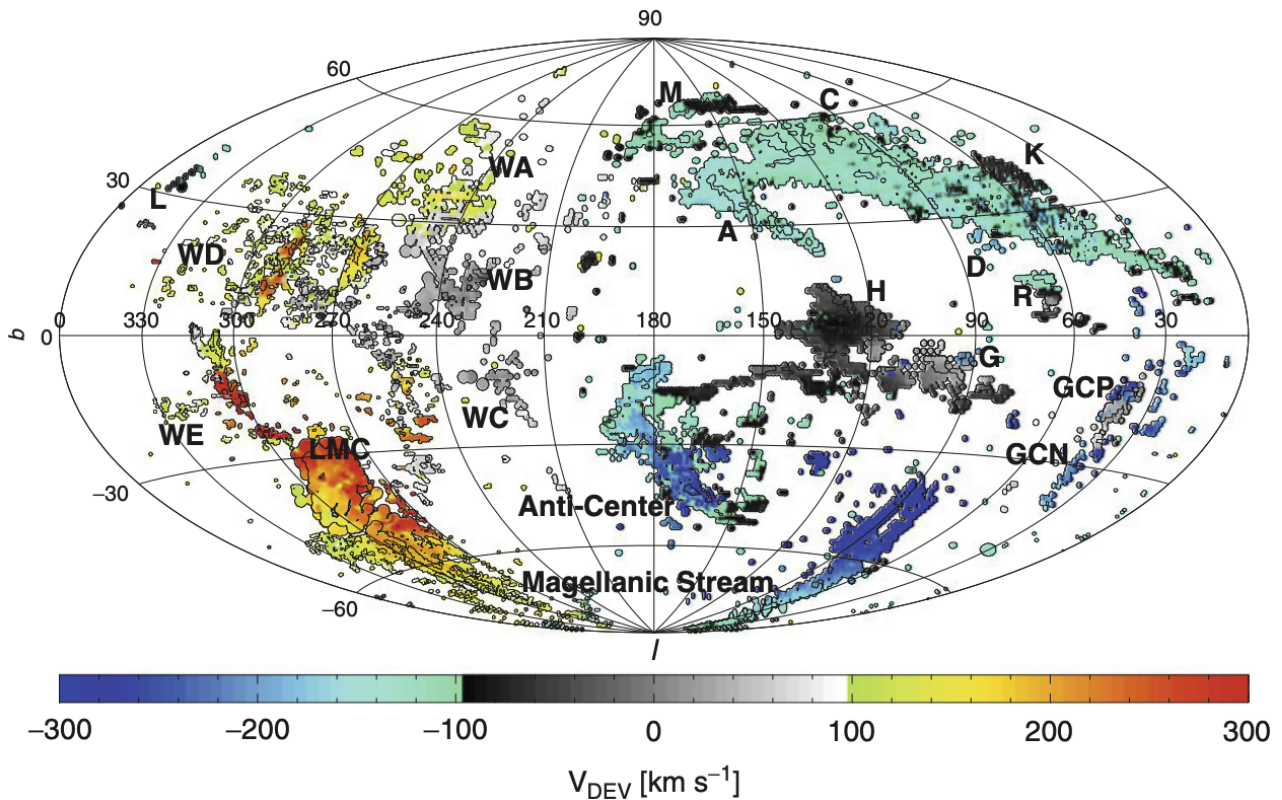


Figure 1.13: Figure 12-1 from Wakker and van Woerden (2013) which shows the distribution of all HI emission with $v_{\text{LSR}} > |90| \text{ km s}^{-1}$ based on the Hulsbosch and Wakker (1988) and Morras et al. (2000) results. The colourbar gives the deviation velocity.

Evidently the Magellanic System, outlined in Section 1.4 dominates a significant fraction of the southern sky. There is also an obvious split between the in the distribution of deviation velocities either side of $l = 180^\circ$, caused by the motion of the local standard of rest. If the velocities are instead converted to the galactic standard of rest, this asymmetry disappears (Wakker 2004). The wealth of HVC/IVC structures that have been catalogued provide a great sample from which to study the properties of the clouds themselves.

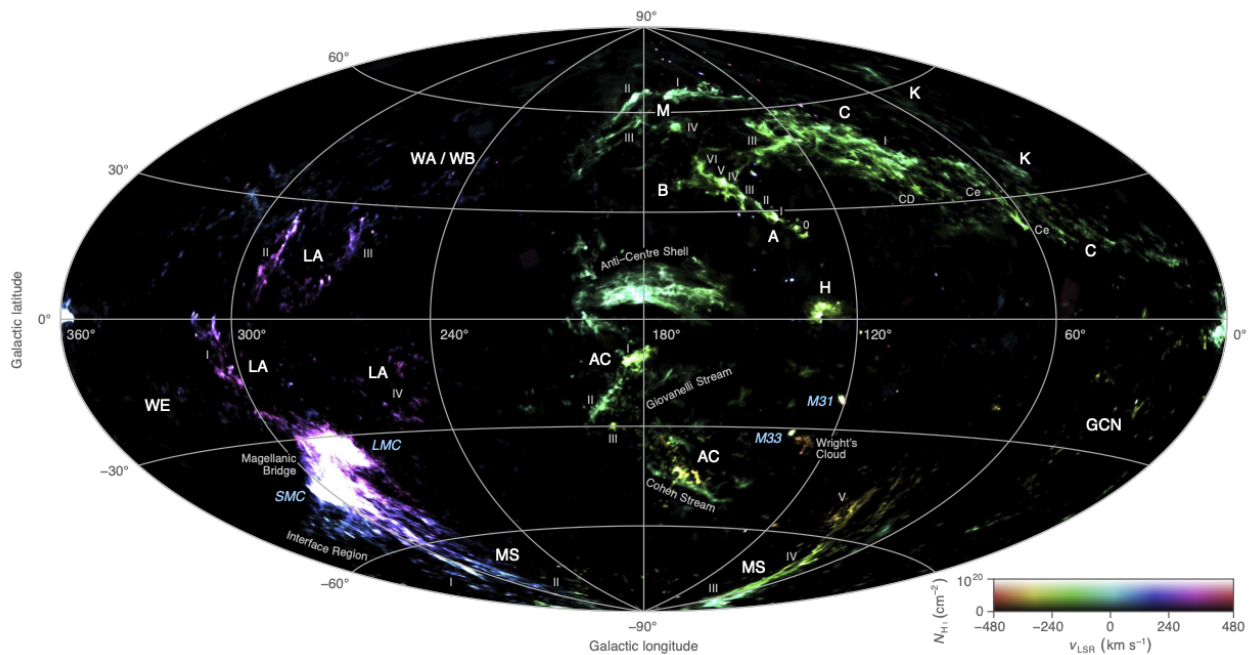


Figure 1.14: Figure 2 from Westmeier (2018) which shows the distribution of all high velocity HI from the HI4PI data (HI4PI Collaboration and et al. 2016). The colourbar gives the LSR velocity.

1.5.2 Properties

The structure of HVCs is something of particular interest, as it may help to constrain the conditions of the clouds. There is a large sample of clouds to consider from the aforementioned surveys and catalogues, so some general properties can be deduced. Giovanelli et al. (1973) show that the HI spectra of a number of HVCs can be fit with a two component Gaussian model with one with a narrow and one with a wide linewidth. They also find that the smaller linewidth components seem to correspond to the bright, compact parts of the clouds, i.e. the cold cores. A study by Haud (2008) found that both IVCs and HVCs had this two component structure as well, with the mean FWHMs for the IVCs being 7.2 and 22.2 km s⁻¹ and for HVCs 7.3 and 27.2 km s⁻¹. They also found that the infall velocity for the cold core of IVCs was higher than that of the warm envelope. Kalberla and Haud (2006) undertook a similar statistical study and fit Gaussians to all HVCs from the LAB survey. They found that 90% of the gas in HVC complexes, excluding the CHVCs, had a two component profile. The distribution that they found is shown in Figure 1.15(a), which demonstrates that for the single component clouds, the gas is predominately in the warm phase. A typical fit of a two component solution to a HVC is shown in Figure 1.15(b) from a study of CHVC (HVC 125+41-207) from Brüns et al. (2000).

In terms of the spatial structure, many clouds exhibit a head-tail morphology, with a dense head of gas at one end of the cloud and a more diffuse tail trailing behind. Brüns et al. (2000) found that 45 of their 252 HVC sample exhibited head-tail structure according to their criteria, with the head generally having the highest radial velocity of the whole cloud. Putman et al. (2011) sampled the population of head-tail clouds in the HIPASS data, totalling 116

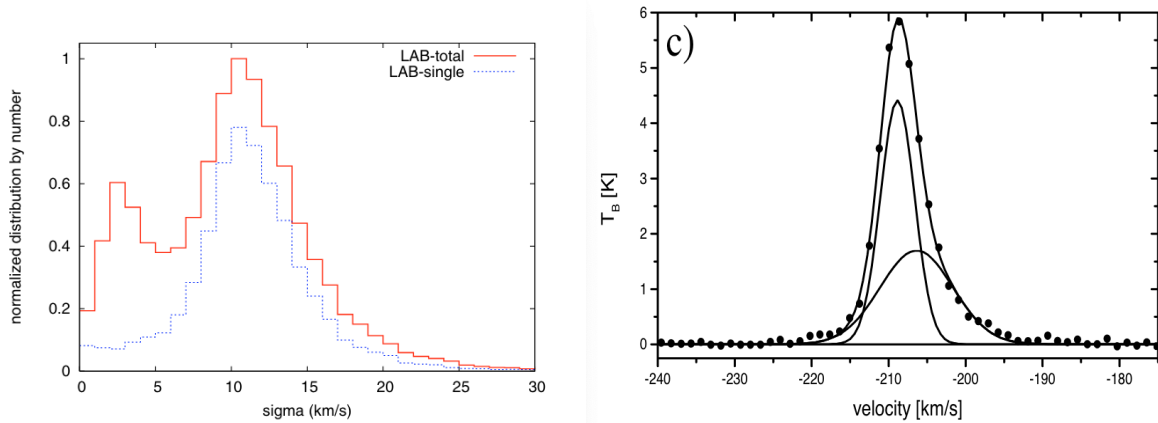


Figure 1.15: Left: Figure 4 from [Kalberla and Haud \(2006\)](#) which shows the distribution of the fitted linewidths of all HVCs with the solid red line and the subset of linewidths from HVCs that had only one component profiles with the dotted blue line. Right: Figure 1c from [Brüns et al. \(2000\)](#) that shows a typical HI spectrum from a compact HVC (HVC 125+41-207).

clouds. Two examples of the spatial extent of head-tails clouds from [Putman et al. \(2011\)](#) are shown in Figure 1.16 clearly showing the dense head with a lower column density tail trailing behind.

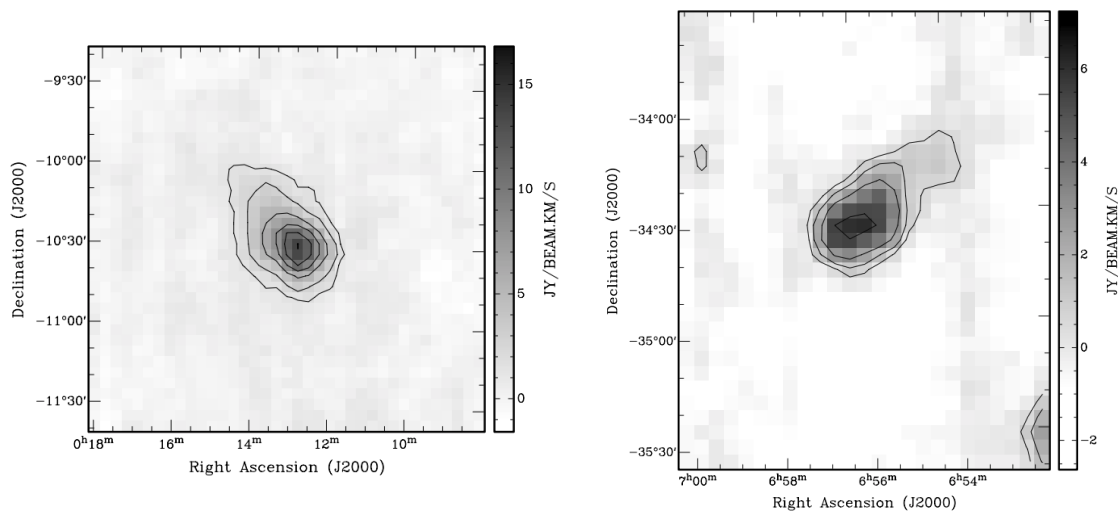


Figure 1.16: A selection of two head-tail clouds from Figure 1 from [Putman et al. \(2011\)](#), they are cloud #501 (left) and #759 (right) from [Putman et al. \(2002\)](#).

Measuring the metallicity of an HVC can help to understand where it originated from. This can be difficult to measure and there can be large variation in metallicity across large HVC complexes. For example in Complex C, there are a range of metallicities, typically from 0.1-0.3 solar but rising nearer to 0.5 at some positions ([van Woerden and Wakker 2004](#)). [van Woerden and Wakker \(2004\)](#) also compiled the measured metallicity ranges of other HVCs, most of which have values below 0.5 solar, except Complex M which has a range of 0.4-1.8 solar metallicity. Work by [Hayakawa and Fukui \(2022\)](#) has used the dust-to-gas ratio as a proxy for the metallicity, as they are expected to scale together, and find maximum metallicities of 0.2 and 0.3 solar for Complexes C and A respectively. The IVCs

analysed had maximums closer to solar metallicity and the Magellanic Stream was measured to have a metallicity of around 0.1 solar, consistent with [Fox et al. \(2013\)](#). The reason that these metallicities are interesting is that they can point to where the gas originated. HVCs with values near solar metallicity likely originate from MW disc, whereas lower metallicity HVCs could point to an extragalactic or Magellanic origin as the LMC and SMC have 0.5 and 0.2 solar metallicity respectively. Metallicity alone is not generally a definitive diagnostic of origin, but in conjunction with other measurements, like distance and mass, it can narrow down the possible formation scenarios.

CO is often used to trace molecular hydrogen (H_2) by use of a CO conversion factor (X_{CO}). However, the absence of CO does not mean that there is no H_2 , as CO is easily dissociated ([Richter and de Boer 2004](#)). This is especially true in low metallicity environments with the ability for CO to trace the H_2 decreasing, leaving a lot of the CO-dark H_2 undetected ([Glover and Mac Low 2011](#), [Madden et al. 2020](#)). CO has been detected towards a handful of IVCs ([Magnani and Smith 2010](#), [Desert et al. 1990](#)). Searches for CO in HVCs have been largely unsuccessful ([Wakker et al. 1997](#), [Dessauges-Zavadsky et al. 2007](#)). Direct observations of H_2 absorption in the FUV has resulted in detection of H_2 in IVCs ([Richter et al. 2003](#)) and in HVCs ([Tchernyshyov 2022](#), [Cashman et al. 2021](#)). So there is evidently CO-dark H_2 present in HVCs.

Dust emission from the Planck survey ([Planck Collaboration 2011](#)) was strongly observed in all IVCs from their sample, with several having higher dust temperatures than the local ISM. They struggled to find dust emission above the uncertainties for HVCs and so reported upper limits. The dust emission in HVCs is expected to be low as the metallicities are typically low. Additionally, the uncertainties in the dust emission are increased by the foreground contamination, as HVCs are likely further from the galactic disc than IVCs. This detection of dust in IVCs but not in HVCs is consistent with previous surveys with IRAS ([Richter and de Boer 2004](#)). A more recent study ([Lenz et al. 2016](#)) with Planck data and improved foreground modelling towards a single HVC still did not detect dust emission above the uncertainties.

1.5.3 Origins

A popular idea for how HVCs originate is the Galactic fountain model. The Galactic fountain model relies on the idea that large amounts of gas can be pushed up out of Milky Way's disc where it eventually slows, cools, and condenses to form clouds that rain back onto the galaxy, this cycle is shown in Figure [1.17](#). One of the main mechanisms for transporting gas out of the disc would be breakout from superbubbles, caused by multiple supernovae. Cosmic rays are also thought to play a role in providing the additional pressure needed to bring material out of the galaxy, possibly even to escape velocities ([Mao and Ostriker 2018](#), [Kahn 1994](#)).

Simulations have shown that the ejection of material from the disc can interact with the corona and funnel material back onto the disc, providing the necessary material for the observed star formation rate ([Li et al. 2023](#), [Barbani et al. 2023](#)). Simulations of our own galaxy have been able to reproduce certain features of the HVC population,

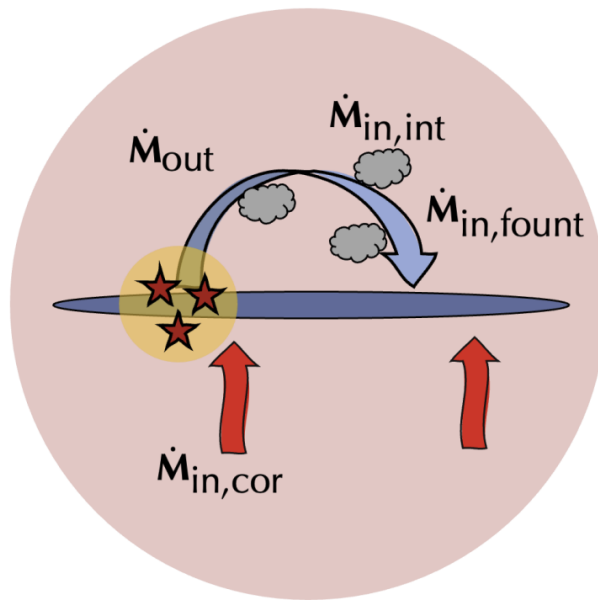


Figure 1.17: Figure 2 from [Barbani et al. \(2023\)](#) which is a cartoon showing the process of gas being ejected at a rate \dot{M}_{out} falling back onto the disc at a rate of $\dot{M}_{\text{in,fount}}$ with extra inflow from the interaction of the gas with the corona $\dot{M}_{\text{in,int}}$. They also include general accretion from the corona not motivated by the galactic fountain $\dot{M}_{\text{in,cor}}$.

such as Complex C in [Fraternali et al. \(2015\)](#) which is shown to have originated from material ejected from the disc. The ejection of gas from the disc also enriches the halo, providing metals that improve the cooling efficiency of the coronal gas ([Fraternali 2017](#)). Figure [1.17](#) illustrates the cycle of gas which is ejected and then brought back to the disc.

Additionally a study by [Marasco et al. \(2022\)](#) posit that there is separation of the inflowing and outflowing gas in their spatial distribution. They suggest that inflowing material can be found well diffused across the galaxy, whereas the outflowing material is distributed in a bi-cone pointing out of the disc. [Di Teodoro et al. \(2018\)](#) studied a collection of HI clouds around the Galactic Centre whose distribution can be reproduced by a bi-conical wind travelling at 330 km s^{-1} . Follow up studies of three of these clouds in HI and CO ([Noon et al. 2023](#) and [Di Teodoro et al. 2020](#) respectively) uncovered the presence of molecular as well as the HI gas being accelerated out of the galactic centre. They show three clouds at different distances from the galactic centre which appear to be in different stages of their journey with the molecular fraction of the clouds decreasing with the distance away from the galactic centre. This demonstrates that strong forces from a galaxy may be able to push molecular material out of the disc, but it struggles to stay molecular as it travels.

Another theory on the origin of some HVCs is gas stripping from interactions with dwarf satellite galaxies. The presence of the Magellanic Stream and the large number of HVCs that it contains shows that the interaction of dwarf galaxies with each other and the Milky Way galactic halo can disturb a lot of gas. [Bland-Hawthorn et al. \(1998\)](#) also highlight the possible association of the Smith cloud with the Sgr dwarf galaxy that appears to be merging with the Milky Way ([Ibata et al. 1994](#)). They find it unlikely that the cloud is part of Sgr as it lies well out of the tidal radius, but

the interaction of the galaxy with the Milky Way could have dislodged the Smith Cloud.

There was also suggestion in [Blitz et al. \(1999\)](#) that there is a population of clouds in the Local Group that eventually fall onto Local Group galaxies, accounting for at least some of the HVCs around the Milky Way. This hypothesis would place some HVCs at distances of around a Mpc. It would also suggest that there are similar clouds for other Local Group galaxies, but as stated in [Bregman \(2004\)](#), surveys more sensitive than the mass limit predicted by the [Blitz et al. \(1999\)](#) model fail to find any new objects like this in the Local Group.

1.5.4 HVCs in other galaxies

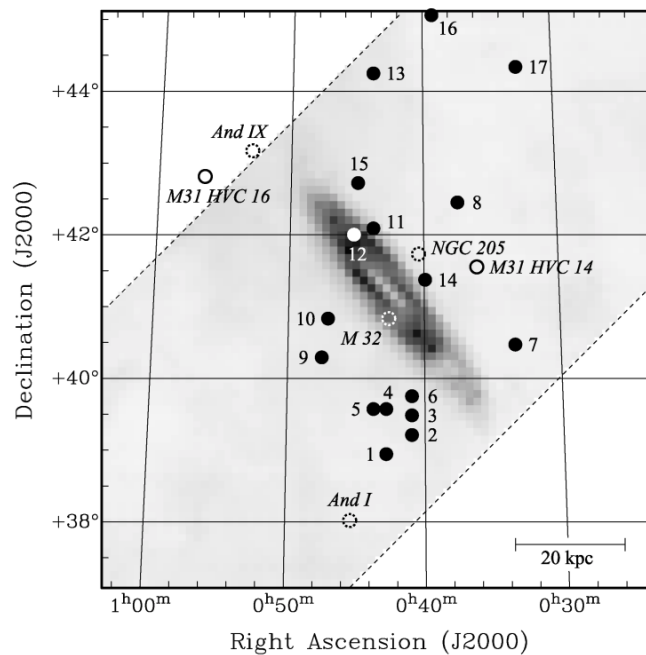


Figure 1.18: Figure 5(a) from [Westmeier et al. \(2008\)](#) which shows the locations of the HVCs (filled circles) identified in HI around M31. These are overlaid on the total HI column density of M31. The dashed open circles are known satellite galaxies of M31 and the solid open circles are HVCs that were previously known but were not detected in the study.

HVCs are unlikely a phenomenon unique to the Milky Way, in fact, anomalous gas has been detected in other galaxies. [van der Hulst and Sancisi \(1988\)](#) studied the HI distribution of M101 and found two high velocity features in disturbed regions of the disc of the galaxy. Studies by [Fraternali et al. \(2001\)](#) and [Fraternali et al. \(2002\)](#) uncovered HI structures in NGC2403 that were not consistent with the rotation model of the galaxy. They found that this anomalous HI, which made up 10% of the total HI, was rotating $25\text{-}50 \text{ km s}^{-1}$ slower than the galaxy and was inflowing. They posit that this gas may be the analogue of our on HVCs in the Milky Way.

M83 was analysed in [Miller et al. \(2009\)](#) and was found to have an anomalously rotating disc, similar to that in [Fraternali et al. \(2002\)](#). With deep HI observation they uncovered eight distinct HI clouds with masses ranging from

$10^5 - 10^7 M_\odot$ with deviant velocities up to 200 km s^{-1} . They explain that they expect this is representative of the high mass end of HVCs in M83 and further observations will likely uncover more. M31 was similarly studied in [Westmeier et al. \(2008\)](#) where they found 17 clouds associated with the galaxy out to distances of 50 kpc in projected distance. These clouds have masses of a few $10^5 M_\odot$ and are distinct from the known satellite galaxies of M31 due to their lower typical mass. [Figure 1.18](#) shows the distribution of the identified clouds on the total HI column density of M31.

Clearly HVCs exist in other galaxies, likely IVCs too, just as they do in the Milky Way. A limiting factor in extragalactic studies is the angular resolution, which is particularly poor at radio wavelengths, which may explain why only clouds of relatively high mass have been detected so far. Higher resolution studies of other galaxies may uncover smaller clouds and IVC populations similar to what we see in the Milky Way.

1.5.5 Simulations

Simulations of HI clouds in a fast, low-density wind were done by [Quilis and Moore \(2001\)](#). They were able to reproduce the head-tail structure seen in many HVCs as long as the wind density was above 10^{-4} cm^{-3} . They found that the lifetime of each neutral tail was around 10^9 years until it dropped to undetectable levels. [Konz et al. \(2002\)](#) investigated role of magnetic fields in the passage of a neutral HVC through the ionised plasma medium and found that the magnetic field drapes around the cloud and protects it from mixing and instabilities in the hot plasma. [Heitsch and Putman \(2009\)](#) conducted hydrodynamical simulations of HVCs either being subjected to a hot wind

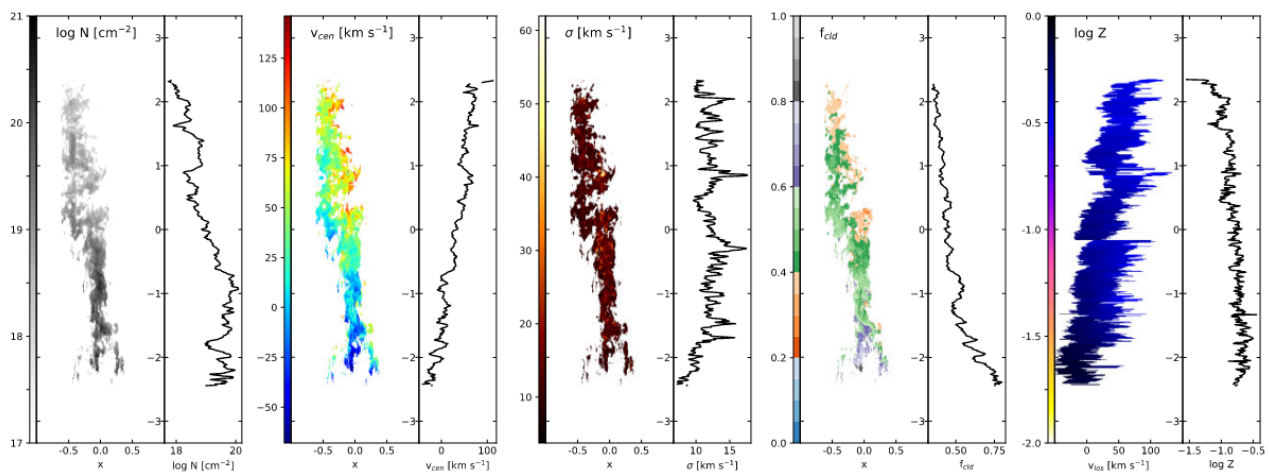


Figure 1.19: Figure 1 (top panel) from [Heitsch et al. \(2022\)](#) which shows the spatial distribution of column density, centroid velocity, velocity dispersion, cold gas fraction and metallicity, in that order left to right, of a cold cloud travelling through an ambient medium at an age of 200 Myr.

(wind-tunnel) or travelling through a hot halo. They found the distinctive head-tail structures in their simulations. They noted that for the wind-tunnel simulations the cloud is fragmented into many, very low mass structures, essentially shredded by the wind. On the other hand, the cloud passing through the hot halo also fragments but the fragmented cores re-cool and are more massive, due to the increasing pressure as they approach the galactic disc.

They found that there was a pronounced velocity difference in their clouds between the head and tail, correlating the velocity gradient with the head-tail structure. [Heitsch et al. \(2022\)](#) investigate the evolution of a HVC passing through a stratified halo. The main aim of this simulation was to investigate the mixing of the cloud material with the ambient medium, particularly in metallicity. They found that the interaction of the HVC with the ambient material could lead to the replacement of over 90% of the material in the cloud, predominantly in the tail. They found that the cloud survives the infall through the cooling and condensation of material in the clouds wake, forming the tail. They also found that the observational signatures, such as gradient in metallicity, velocity, cold gas fraction etc. (shown in [Figure 1.19](#)) are present early in the clouds lifetime but are lost as the cloud evolves.

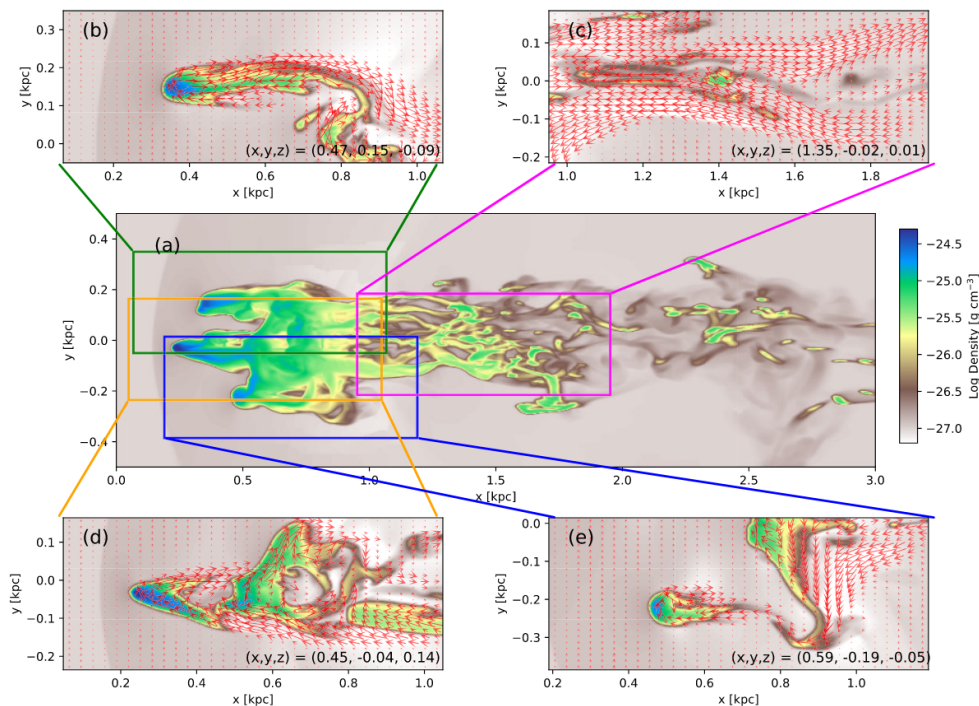


Figure 1.20: Figure 6 from [Jung et al. \(2023\)](#) which shows the magnetic field with red arrows overlaid on the density structures in a magnetised, initially clumpy cloud, with the column density indicated by the colourbar.

[Jung et al. \(2023\)](#) looked at the effect of the magnetic fields on HVCs with initial masses around $10^4 M_{\odot}$ travelling through the galactic halo. They adopted a metallicity of 0.3 solar, which is in the range of metallicities observed for local HVCs, and launch the cloud at 200 km s^{-1} . They found that the magnetic fields drape around the head of the cloud and resist the instabilities that cause the clouds to fragment in the non-magnetised scenario, similar to [Konz et al. \(2002\)](#). They compared the growth of clumpy and uniform clouds and found that both grow in mass, despite the clumpy cloud losing mass initially, indicating that HVCs can survive the infall process. [Figure 1.20](#) shows this draping effect of the magnetic field around the clumps present in the multi-headed cloud produced from an initial cloud with a clumpy density profile.

Chapter 2

Gaussian decomposition and model selection

2.1 Data

This work makes use of the data from the GASKAP-HI survey. The GASKAP-HI survey is a survey selected back in 2009 to be conducted with the Australian Square Kilometre Array Pathfinder (ASKAP) telescope, which was in the construction phase at that time. The ASKAP telescope is one of the Square Kilometre Array (SKA) pathfinders for the high frequency component of the final telescope. The GASKAP-HI survey began pilot observations in 2019, during which the SMC was observed for 20.9 hours. The design of the telescope allows for 36 simultaneous beams to form a 25 deg^2 field of view (FOV). The SMC was observed as a single field centred at RA = 00:58:43.28 and DEC = -72:31:49.03. The data obtained with ASKAP was feathered with previous data obtained with the Parkes radio telescope and has a final restoring beam of $30'' \times 30''$ after smoothing the data. The final data product has a spectral range of 40.0 km s^{-1} to 253.9 km s^{-1} with a channel spacing of 0.98 km s^{-1} . The spatial spacing between pixels is $7''$. The total datacube is $\approx 11 \text{ GB}$ in size and has $3901 \times 3471 \times 220$ voxels. The full details of the GASKAP data reduction are contained in [Pingel et al. \(2022\)](#). The fields explored in this thesis are subcubes taken by simply splicing the data over the ranges of interest in all three axes. The specific fields used in this Chapter are described in Table 2.1 and will be detailed further in Chapter 3 with the choice of velocity range will be explained further in this Chapter.

Three specific clouds were chosen for this work that were first identified as potential outflows from massive star formation in [McClure-Griffiths et al. \(2018\)](#) in the SMC. They are good candidates for analysis with Gaussian decomposition as their emission is very strong in the densest regions of the clouds, with column densities on the order of 10^{20} cm^{-2} . Additionally, while the distribution of CNM in the main body of the SMC can be well-observed in absorption, these clouds have compact structures and were not probed by any background sources in the latest absorption study ([Dempsey et al. 2022](#)), due to the intrinsic limit imposed by the background source density. Thus, this work provides an opportunity to analyse the phase distribution just from the emission of dense HI clouds with high signal to noise ratios.

2.1.1 Noise estimation

To calculate the noise for the GASKAP SMC field, I use a typical empirical method, similar to that used in [Pingel et al. \(2022\)](#). I take the standard deviation along each line of sight of the channels with no HI emission, which we call σ_0 . After inspecting each channel visually, these are defined as where $v < 62 \text{ km s}^{-1}$ and $v > 235 \text{ km s}^{-1}$. Since the velocity range of the cube extends from 40.0 km s^{-1} to 253.9 km s^{-1} , this leaves 43 channels out of 220 (19.5%) that

Field number	RA centre (J2000)	DEC centre (J2000)	Width (')	Height (')
1	00:56:18	-70:50:43	18.9	45.5
2	01:09:21	-71:20:48	14.93	14.93
3	01:07:40	-70:51:21	37.1	45.5

Table 2.1: Field parameters

are emission free, from which to calculate σ_0 . After calculating the stand deviation along each line of sight, I am left with a 2D spatial map of the noise in the cube. As expected from similar results in [Pingel et al. \(2022\)](#), the noise is highest along the edges of the field of view and I get a median value of 1.6 K, compared to their value of 1.1 K. To fairly assess the solutions fits as I will do in this section, I need to have a reasonable understanding of the noise level along each spectrum as well. Noise changes as a function of velocity along each line of sight in the SMC datacube as the brightness of the emission signal increases the temperature of each antenna. I calculate the noise along each line of sight as is done in [Boothroyd et al. \(2011\)](#):

$$\sigma(v) = \sigma_0 \left(\frac{T_{\text{sys}} + T_B(v)}{T_{\text{sys}}} \right). \quad (2.1)$$

I use the quoted value of $T_{\text{sys}} = 55$ K from [Pingel et al. \(2022\)](#). Applying Equation [2.1](#) to the 2D spatial map of the noise adds an extra dimension, so now I have a 3D map of the noise through the whole datacube that can be used going forward in the next Sections.

2.2 ROHSA decomposition testing

Gaussian decomposition techniques have been used extensively for modelling HI data and extracting meaningful information from spectra, as outlined in Section [1.3](#). For optically thin emission, the model is quite simple, with the idea that the HI spectra ($T_B(v)$) can be modelled by sum of Gaussian equations. This model assumes that there are one or more emission signals coming from discrete clumps of HI gas that are each travelling at some bulk velocity. Each instance of the emission line that makes up the total spectrum is broadened primarily by the thermal motions of the gas, meaning broader components correspond to hotter gas. The model can be described as such:

$$T_B(v) = \sum_{n=1}^N a_n \exp \left(\frac{-(v - \mu_n)^2}{2\sigma_n^2} \right) \quad (2.2)$$

where a_n is the amplitude of the Gaussian, μ_n is the central velocity of the Gaussian and σ_n is the dispersion of the Gaussian (or the broadening). This model can be used on any HI spectrum to encode the information within.

There are limitations to this kind of modelling as mentioned in Section [1.3](#). Firstly, it does not take into account self-absorption by the gas, referred to as HI self absorption or HISA. HISA is seen when absorbing CNM sits in front of hotter emitting HI. HISA is normally measured by typical absorption methods, requiring the determination of an ON and OFF spectrum to find the absorption signal in the difference of the two. Secondly, it assumes that the gas motions that cause the line broadening are dominated by thermal motions. The other main cause of line broadening is turbulence that increases the interior motions of the gas. In [Miville-Deschênes et al. \(2003\)](#) it is stated that Gaussian functions well describe the signal if the turbulent motions contribute proportionally to the broadening as the thermal motions do.

Regarding these points in the specific case of the SMC, a recent absorption study that achieved the highest source density of any previous SMC study (Dempsey et al. 2022) showed only one source that had evidence of HISA, out of 229 sources analysed. This study used the same data that I use in this work, so this is a direct indication of how low the prevalence of HISA should be in this work. Also if HISA acts on small scales, I would expect to see large differences in neighbouring spectra at the edge of HISA feature indicating its presence, which I did not come across in this work, thus I do not consider it.

On the point of turbulence, some studies of the SMC (Burkhart et al. 2010, Chepurinov et al. 2015) have shown that the majority of the SMC is in sub or trans sonic regimes when looking at the HI emission data, with enhancements in turbulence typically on small scales and due to star formation activity (Grisdale et al. 2017). So in this work I expect to be primarily in low turbulence regimes, so the Gaussian model is appropriate to use in this case.

In this work I made use of a new Gaussian decomposition package ROHSA (Marchal et al. 2019) which uses an iterative, top-down approach for fitting HI data that enforces regularisation of the solution. The algorithm begins by optimising the solution for the mean spectrum of the entire datacube provided. It then progressively segments the datacube into smaller and smaller grids to which it optimises the solution to each one, until the pixel scale is reached. This already encourages regularisation, by using the model of the larger grid section as the initial guess for the subsequent subdivided grid sections, but to enforce it ROHSA employs an extra term in the cost function that accounts for it. This term is controlled by four hyperparameters set by the user, whose function is to control the amount of regularisation of each parameter the Gaussians fit to the data. These are λ_a , λ_μ , λ_σ , and λ'_σ . These hyperparameters, together with a typical least squares term, make up the cost function of ROHSA ($J(\theta, v)$), shown below from (Marchal et al. 2019):

$$J(\theta, m) = \frac{1}{2} \sum_{r,v} \left(\frac{T_{B,M}(\theta(r), v) - T_B(r, v)}{\Sigma(r)} \right)^2 + \frac{1}{2} \sum_{n=1}^N \lambda_a \|\mathbf{D}\mathbf{a}_n\|_2^2 + \lambda_\mu \|\mathbf{D}\boldsymbol{\mu}_n\|_2^2 + \lambda_\sigma \|\mathbf{D}\boldsymbol{\sigma}_n\|_2^2 + \lambda'_\sigma \|\sigma_n - m_n\|_2^2 \quad (2.3)$$

where θ represents the Gaussian parameter terms i.e. a_n, μ_n, σ_n , m represents the typical value of the velocity dispersion, discussed further in the next paragraph, r is the sky or x and y coordinates, v is the velocity coordinate, and N is the number of Gaussians used, \mathbf{D} is a matrix performing the 2D convolution using the kernel d which is defined in (Marchal et al. 2019).

The first three hyperparameters ($\lambda_a, \lambda_\mu, \lambda_\sigma$) act to regularise the fitted components of each Gaussian in the solution, i.e. the amplitude (a), the central velocity (μ) and the velocity dispersion (σ) between neighbouring pixels. The fourth hyperparameter (λ'_σ) regularises the value of σ across the entire field, as this relates to the thermal properties of the gas and should not vary to a significant degree. As mentioned above, this also relies on the value of m , of which there is one for each Gaussian in the solution. Unlike the hyperparameters, the m term cannot be set by the user, it is rather a fit parameter derived in the optimisation of the cost function.

These four hyperparameters are set by the user. ROHSA requires the user to supply a parameter file when calling the function to run on data. This parameter file is where the values for the hyperparameters, along with a number of initialisation options set by the user. Looking at Equation 2.3 it is clear that increasing the value of any one hyperparameter increases its influence on the value of the cost function. If the values of the hyperparameters are low, we could end up with a result where the least squares term dominates the cost function, so the minimiser brings us to a solution that has focused on fitting the individual lines of sight but has low similarity in neighbouring pixel solutions. On the other hand, if the values of the hyperparameters are high and the regularisation terms dominate the cost function, the minimiser could arrive at a solution that has high similarity in neighbouring pixel solutions, but many individual lines of sight are poorly fit. Striking the right balance between these two scenarios is key, we want a solution that makes physical sense, solutions for neighbouring pixels should be similar, and is also well fit. I discuss this in the next section.

2.2.1 Hyperparameter tuning

When I started working with ROHSA, I tested the effects of varying the hyperparameters on the ROHSA solution extensively. I used a single field of $390 \times 162 \times 220$ to begin this testing phase. The characteristics of this field are detailed further in Chapter 3, but in this section we will consider it just as a testing field for ROHSA modelling.

Initially, I started with an 8 Gaussian solution to test the effects of the hyperparameters on the solution that was output for the field. The 4 hyperparameters were varied over a range of 5 orders of magnitude ($10^1 - 10^5$) which probes a range spanning similar values to that used in Marchal et al. (2019). The range here goes down to 10 for the hyperparameters as I expect to be probing larger physical scales than what were analysed in Marchal et al. (2019), meaning I expect the solution to vary more pixel to pixel. The expected trend in this exploration would be for the solutions with lower regularisation to have smaller least squares errors and thus the best reduced chi squared values. I also expect that the all hyperparameters will have some effect on the least squares errors, though to what degree is unclear. Since I am fitting with the same number of Gaussians each time, the degrees of freedom in the fitted model never changes, so only the residuals of the model subtracted from the model affect the reduced chi squared value. Each of the four hyperparameters were stepped through values: $10^1, 10^2, 10^3, 10^4, 10^5$; creating 625 combination of hyperparameters that were run. After running all of these combinations of parameters, there needs to be some way to differentiate between them to determine the best fit. The obvious metric to use first is the reduced chi squared measure which I define as such:

$$\chi_{\text{red}}^2 = \sum \sum \sum \frac{(T_{B,\text{data}}(x, y, \nu) - T_{B,\text{model}}(x, y, \nu))^2}{\sigma_{\text{noise}}(x, y, \nu) \cdot \text{DOF}} \quad (2.4)$$

where X and Y are the number of pixels in the x and y directions respectively, Z is the number of spectral channels along each pixel position, $\sigma(x, y, \nu)$ is the noise at every voxel, and DOF are the degrees of freedom of the fit, defined

as the number of independent measurements minus the number of parameters fit. In this case the DOF is equal to the total number of voxels in the cube minus the number of parameters fit to the whole cube, as below:

$$DOF = X \cdot Y \cdot Z - (X \cdot Y \cdot 3 \cdot N) \quad (2.5)$$

$$= X \cdot Y(Z - 3N) \quad (2.6)$$

where Z is the number of velocity channels and N is the number of Gaussians fit to the field.

I calculate this for all 625 solutions and find a large spread of values, as is to be expected. The values range from around 1-5. When interpreting an χ_{red}^2 value, a value close to 1 is ideal. A value larger than 1 indicates the data is underfit by the model, a value smaller than 1 indicates the data is overfit by the model. None of the values of reduced chi squared in this exploration fell below 1, so it appears I have not overfit the data. There are two main effects that become clear when you look at the distribution of χ_{red}^2 . The χ_{red}^2 value depends on the λ_a and λ'_{σ} values, as the higher the regularisation parameter the higher the χ_{red}^2 value. However, for the λ_{μ} and λ_{σ} , there is very little change in the χ_{red}^2 value as the level of regularisation changes. This trend is shown in Figure 2.1

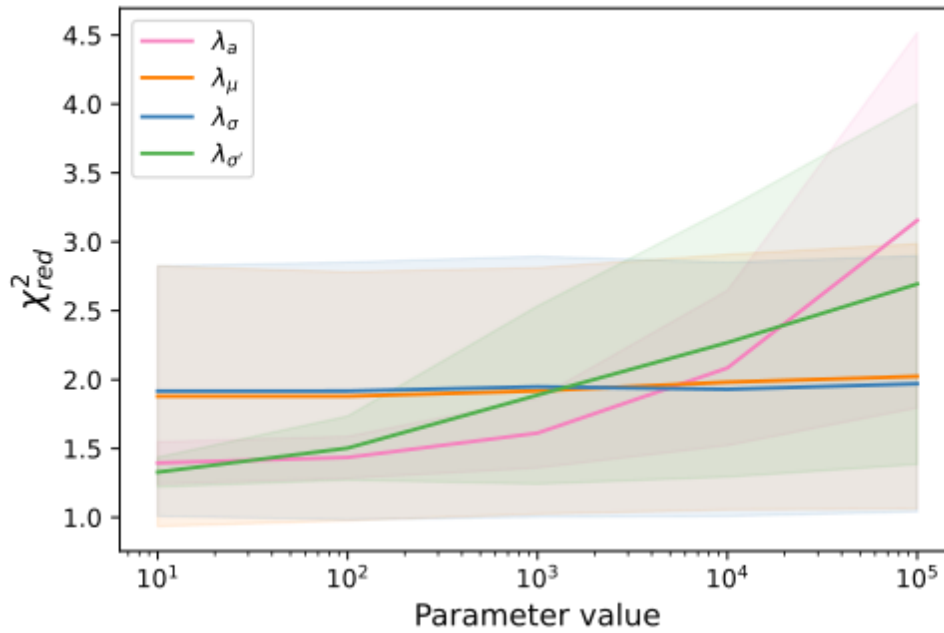


Figure 2.1: The trend of χ_{red}^2 of for each hyperparameter as the values are varied for 8 component solutions. The lines indicate the mean χ_{red}^2 and the windows of the corresponding colour indicate the standard deviation.

This result indicates that regularisation of a and the mean σ value across the field are the most important things affecting the fit. Fitting the most regularised solution without drastically increasing the χ_{red}^2 is what I am aiming for, since the physics would dictate the need for the solution to be consistent pixel-to-pixel. Therefore the values of λ_a and λ'_{σ} , should be 10 and the other two could also be set to 10 as well since there is no strong trend. So now I have

an idea of what the values should be for this field, is there a way to narrow down the best solution even further. There are other metrics that can be analysed to characterise the solutions, I discuss this in the next section.

2.2.2 Differentiating between solutions

Since the CNM distribution is of particular interest in this work, one of the first things that I looked at was the CNM fraction (f_{CNM}) of each solution. If I am able to find a preference for particular f_{CNM} amongst all the solutions, that would indicate there are likely only a handful of preferred solutions, with just small variations on the details. The CNM fraction is defined as so:

$$f_{\text{CNM}} = \sum^R \sum_j^J \frac{N_{H,j}(r)}{N_{H,\text{total}}(r)} \text{ for } j \text{ where } \sigma_j(r) < 3 \text{ km s}^{-1} \quad (2.7)$$

where J is the total number of components, r represents the pixel coordinates, and R is the total area of the field.

The distribution of the f_{CNM} against the χ_{red}^2 is shown in Figure 2.2.

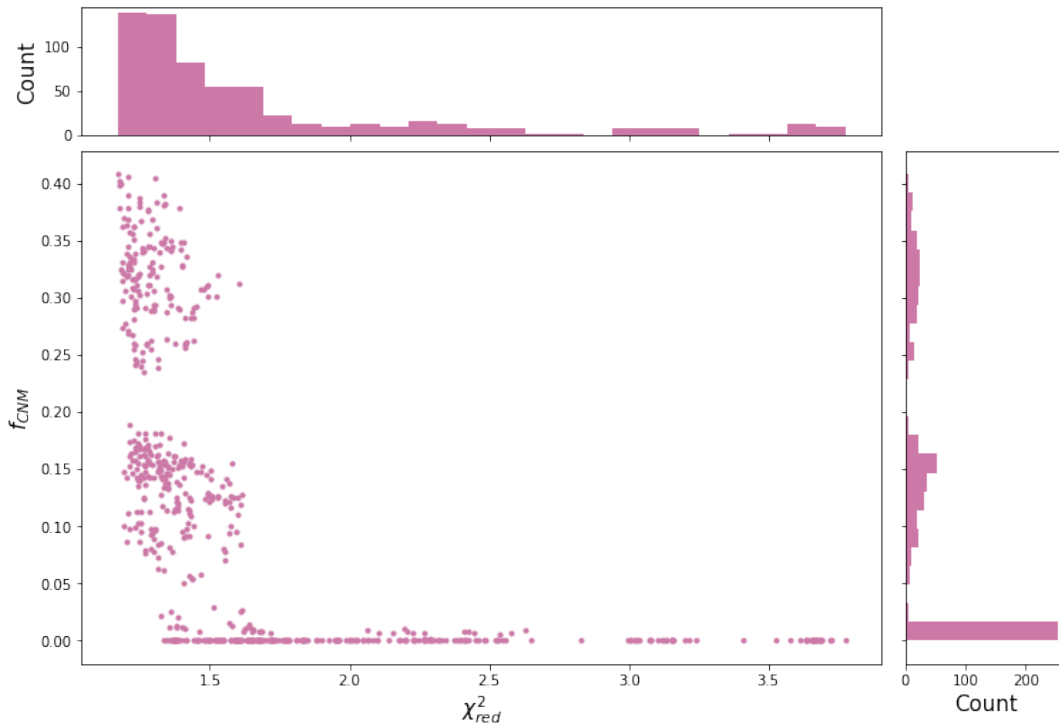


Figure 2.2: The distribution of f_{CNM} and χ_{red}^2 for each 8 component solution, with a histogram of χ_{red}^2 above and f_{CNM} to the right.

When looking at the distribution, it is clear that there is a strong preference for no CNM across the whole field, with the high χ_{red}^2 solutions contributing the most to this clear maximum in the histogram. There are shallow local maxima at an f_{CNM} of around 0.15 and 0.3 that correspond to the better fit values of χ_{red}^2 but there is no strong clumping of f_{CNM} values around any value. So the CNM fraction in this case can not be used as an indicator of

any grouping of solutions as there is a reasonably continuous spread of values from 0-0.45. It would also not be reasonable to judge a solution on its CNM content as that is one of things I am trying to recover in this work. But given that there is such a variation in f_{CNM} values amongst well-fit solutions with $\chi_{\text{red}}^2 < 2$, how can I be sure that I am fitting the right well-fit solution to the data? One thing that can help in reducing the confusion is reducing the number of components fit to the data as this will decrease the degeneracy. Given that all the χ_{red}^2 values are above 1 it does not seem like it would be advisable to reduce the number of parameters fit to the data, however if we can fit the solution just as well with less components then the χ_{red}^2 value will decrease as the DOF increases. To determine what the appropriate amount of components should be, it is possible to make some kind of measurement of the number of components that meaningfully contribute to the solution. I use meaningfully in this context to say that a component represents a non-zero fraction of the solution. To measure this I make a measure of the effective n number of components by ordering the components by column density from largest to smallest, and taking the cumulative function and defining a cut-off point of 80% of the total column density calculated directly from the data. This should indicate how many components are needed to fully capture the signal. Firstly, I calculate the column density assuming the optically thin regime for each component with the following:

$$N_{H,n} = 1.823 \times 10^{18} \int T_{B,n}(x, y, z) dv. \quad (2.8)$$

As T_B is represented by a sum of Gaussians in the ROHSA model, each of the n components has an specific column density. A Gaussian has a definite integral from $-\infty$ to ∞ of:

$$\int_{-\infty}^{\infty} a e^{-(x-b)^2/2c^2} dx = \sqrt{2\pi} a |c|. \quad (2.9)$$

So each component has a column density of

$$N_{H,n} = 1.823 \times 10^{18} \sqrt{2\pi} a_n \sigma_n. \quad (2.10)$$

The way I define the effective n is ordering the column density of each component from largest to smallest, summing the largest K component column densities, for increasing values of K until the following expression is true:

$$\sum_{k=1}^K \frac{N_{H,k}}{N_{H,\text{data}}} \geq 0.8 \quad (2.11)$$

where $N_{H,\text{data}}$ is the total column density along the line of sight calculated from the actual data, as opposed to the model, using Equation [2.9](#). Therefore, the effective n is the smallest K that satisfies Equation [2.11](#).

To calculate a single value for each solution I take the mean value across the data cube of the effective n . I also

calculate the value of the effective n if the threshold is raised from 80% to 95%. For the 625 solutions I can examine the distribution of the effective n value in Figure 2.3

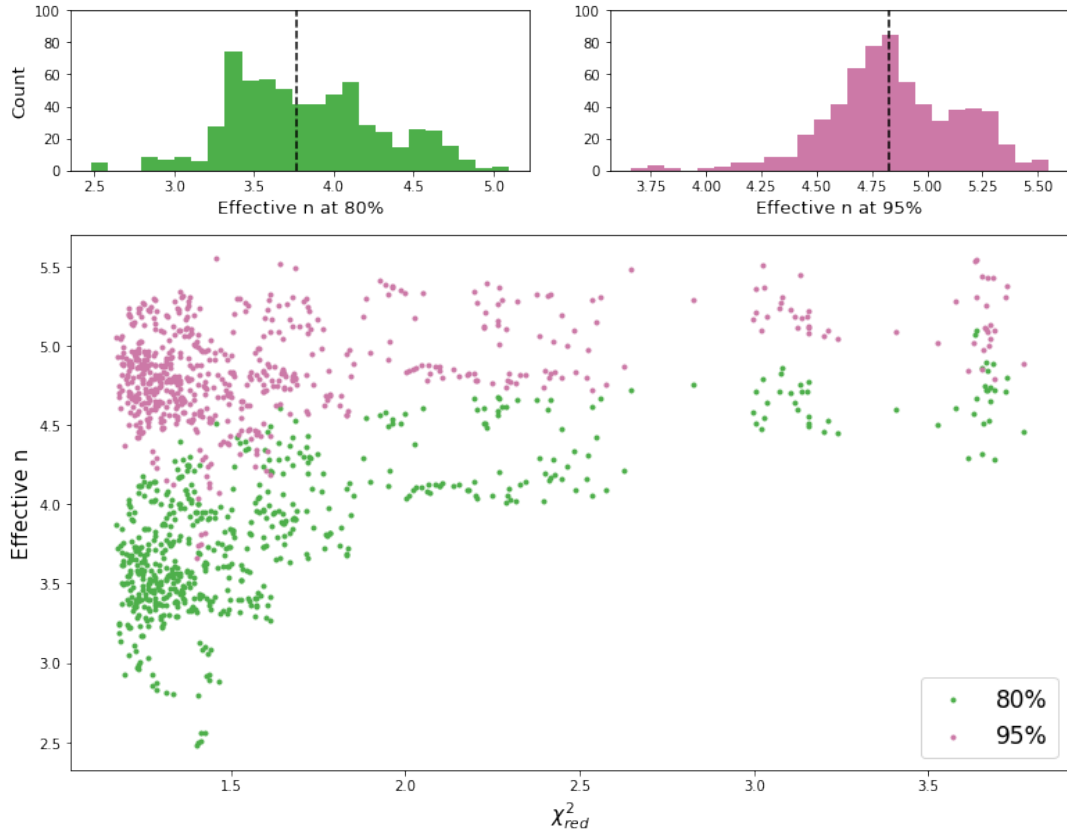


Figure 2.3: The distribution of the effective n and χ_{red}^2 for each 8 component solution. The distribution of the effective n at 80% and effective n at 95%, in green and pink respectively, are shown in the top panels.

For the effective n at 80% there is a weak positive correlation with the χ_{red}^2 that is even weaker with the 95% effective n measure. The medians for the respective 80% and 95% measures are 3.75 and 4.8. Since I can only fit whole numbers of components with ROHSA, I chose to take the value of 5 as the representative effective n that encodes almost all of the signal along any line of sight.

After deciding upon this new value of 5 components, I chose to rerun my ROHSA hyperparameter space exploration with the knowledge gained from the previous one. This not only reduces the amount of computation that has to be done but also the time it takes to do it. I chose to fix the values of λ_{μ} and λ_{σ} to be 10, since they had no effect on the χ_{red}^2 in the previous exploration and then to focus on values of λ_{α} and λ'_{σ} between -2 and 2 in \log_{10} scale. ROHSA also allows the user to set the initial σ that it fits for each component, so I vary this linearly from 2-8 km s^{-1} to see how this affects the solution. In total this exploration produces 175 solutions with 5 components each. In Figure 2.4 I show how the χ_{red}^2 changes with the value of the hyperparameter or the values of the initial σ .

There is no real change as I vary the initial σ except maybe a slight increase in the χ_{red}^2 at a higher initial σ . There is again this correlation between the higher values of χ_{red}^2 with increasing hyperparameter values. It looks very similar

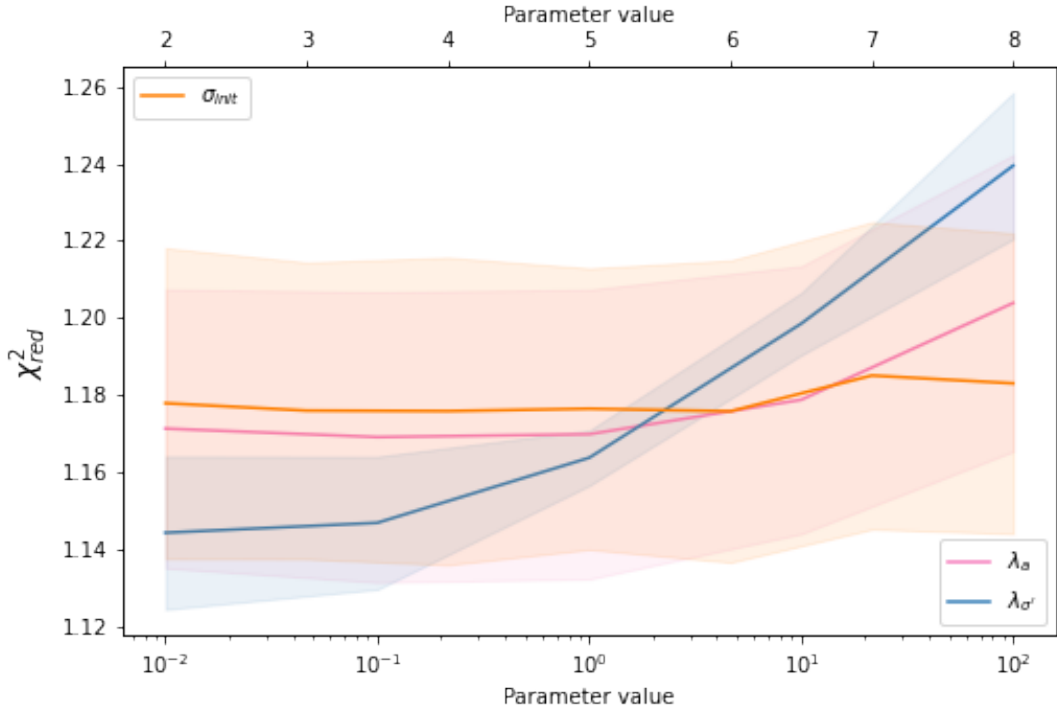


Figure 2.4: The trend of χ_{red}^2 for λ_a , $\lambda_{\sigma'}$, and σ_{init} as the values are varied for 5 component solutions. The lines indicate the mean χ_{red}^2 and the windows of the corresponding colour indicate the standard deviation.

to the trend in the last exploration but the range of values is noticeably different, from around 1.1-1.3 as opposed to 1-5. So although there is an improvement on the goodness of fit, it is fairly inconsequential. One thing I would worry about is that if I select a solution purely based on χ_{red}^2 then I might have sacrificed a lot of the regularisation for a slightly 'better' fit, which defeats the point of using ROHSA in the first place. So to see if this is occurring, I can visualise a high χ_{red}^2 solution and low χ_{red}^2 in this exploration. A good way to visualise a solution is to look at the σ - v diagram, that is a plot of the centroid velocity against the σ at every point in the field. This way we see each fitted component as a clump of points around a mean value for both parameters. In Figure 2.5 I show an example of these σ - v diagrams for the highest χ_{red}^2 solution and lowest χ_{red}^2 in the aforementioned exploration.

These show very different distributions of the central velocity of the solutions vs the σ . The low χ_{red}^2 solution also has $\lambda_a = 0.01$ and $\lambda_{\sigma'} = 0.1$ which are amongst the lowest values of the hyperparameters in this exploration. It is impossible in this solution to distinguish the area in the σ - v space that each fitted component resides in. We expect to see each component tracing a specific phase of the neutral hydrogen, so the value of σ should not vary wildly across the field that is fit. This wild variation is exactly what we see in Figure 2.5 (a) where a lot of the σ - v space is covered by the solution. In contrast the highest χ_{red}^2 solution has more compact groupings that span a wide range of velocities but have much tighter σ distributions. This is what is preferred if we are looking for a spatially regularised physical solution where each component traces a specific phase. Something else to note is the large values of σ that are fit in the low χ_{red}^2 solution. A σ value of 35 km s^{-1} corresponds to a maximum kinetic temperature of 1.5×10^5

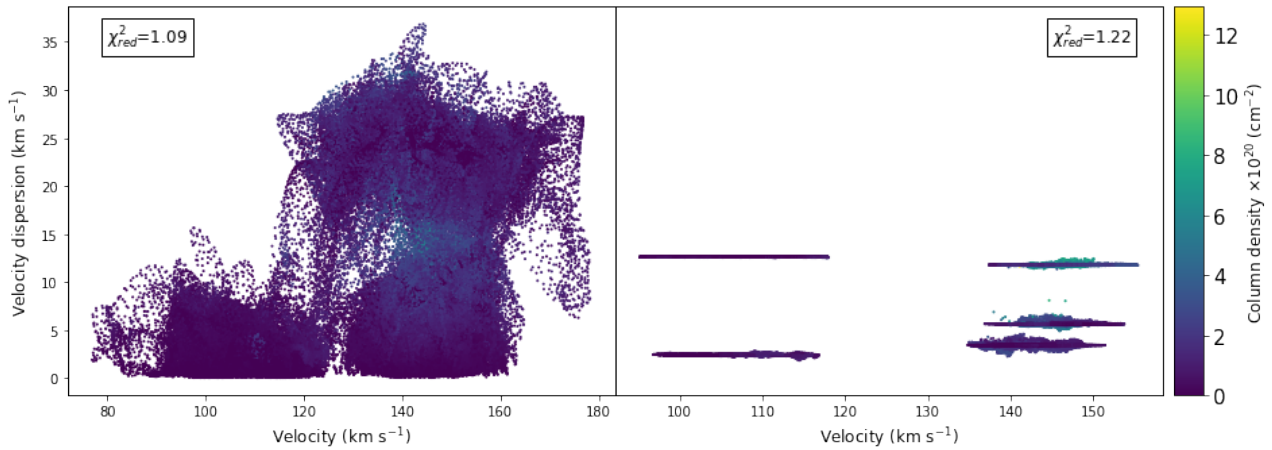


Figure 2.5: Different σ - v diagrams for the solutions with the lowest (left) and highest (right) χ_{red}^2 values of the exploration with 5 components fit to the data.

K, which is well above the standard maximum temperature of the WNM, which reaches around 10^4 K depending on the environmental conditions. This is illustrative of what happens when Gaussian fitting is done without restriction, the solution components could easily exceed physically reasonable values for the velocity dispersion. This is why it is helpful to use a Gaussian fitting algorithm that enforces regularisation in combination with setting reasonable values for the initial velocity dispersion of the solution. This encourages the algorithm to arrive at physically reasonable values itself, without the user having to define a hard limit on the velocity dispersion.

Overall, it seems clear that once we reach λ_a and λ'_σ values on the order of 10 for this field the improvement in the χ_{red}^2 value lessens considerably. We see this shallowing in Figure 2.1 around values of 100 for the 8 component solutions and in Figure 2.4 when we reduced the solution down to 5 components. Additionally, it is clear from the 5 component exploration that a lot of the coherence of the solution is lost, as shown in the σ - v space. So when looking for ideal solutions, we should look for a balance of high regularisation and χ_{red}^2 near 1.

Until now, there has not been much consideration for the specifics of the field that I am testing with. This field was used as a test field, but is also a HI structure that I want to analyse. So I have to consider how the different fits affect that analysis. We can see in Figure 2.5 that there are components fitted over two specific velocity ranges, 95-120 km s⁻¹ and 135-160 km s⁻¹, this division is even seen in the low regularisation solution. The structure that I am interested in analysing is in the 95-120 km s⁻¹ range. From looking at the spectra themselves, it becomes clear that the solution over this velocity range is likely a two component solution, which is what we see in the regularised solution. A two component solution in this case means that there is a very strong narrow peak with a base wider than that expected of the tails of narrow Gaussian. This would correspond physically to a WNM and CNM component. However, there are some areas at which two narrow signals become apparent, offset from each other by a few channels. To fully encode this, we require a 3 component solution in this region of the velocity spectrum. To achieve this, I ran ROHSA again with hyperparameter values all set at 10, with 9 components, as we do not see this needed

component even in the 8 component solution. However, even with 9 components, this needed 3rd component in the lower velocity range just does not materialise. The extra components keep being added in the other end of the velocity range. The problem that I was encountering was that the main body emission in the 135-160 km s⁻¹ range is dominating the mean spectrum of the field and therefore setting the initial guess for the top-level solution that ROHSA begins with. With the version of ROHSA used for this hyperparameter tuning analysis, there was not the option to provide an initial guess to the ROHSA algorithm, so it just uses the mean.

It was clear that running ROHSA with more and more components in an attempt to fit a small number of pixels that in the context of all the data ROHSA was fitting, were fairly insignificant, was pointless. Not only was this increasing the degeneracy issue that I had just tried to reduce, but it was also increasing the value of the χ_{red}^2 by decreasing the *DOF* further and further with extra components. Evidently the next step in this process was to focus on the part of the data that I wanted to fit, rather than attempting the fit the whole spectrum at each pixel. I discuss this in the next section.

2.2.3 Isolating the cloud emission

While the signal from these structures is strong in the densest areas, the emission from the main body of the SMC still dominates the spectra, especially when looking at the mean spectrum over each full field containing the clouds, shown in Figure 2.7. Since the emission from the main body of the SMC is not what I am interested in for this work, it became necessary to isolate the emission from the structures of interest and get rid of the SMC main body emission. The first step to removing the emission of the main body from the total spectrum is to take a subset of the spectrum over the velocity range of interest.

However, there is a big issue that arises when doing this. Whilst the structures we are interested in are offset from the main body emission by around 20-30 km s⁻¹ there are some wide components (WNM) from the main body whose tails bleed into the velocity range I chose to limit the fitting to. This means that the algorithm now has a 'feature' it is trying to fit at one end of the sub spectrum, which is impossible to fit. It is impossible as ROHSA will not fit a central velocity outside of the spectrum velocity range, so instead ROHSA will likely try to fit a narrow low amplitude component to cover the wide component tail. This is problematic for two reasons. The first is that this artificially inflates the CNM fraction, but I can introduce some accepted range for the central velocity of a CNM component for this structure that would exclude these artefacts, so it is not catastrophic. The second problem is that a narrow, low amplitude component will not fit the tail and will leave positive residuals that ROHSA might try and fit again if I add more components.

Clearly this is not ideal, so the solution to this is to properly remove the main body emission. Since I do not care what the true solution is to the main body emission in terms of the different phases and their velocities, I can do an unregulated fit to the data to get the shape of the profile to remove. In Figure 2.5 there is a clear gap in the velocity

distribution that separates the main body from the structure of interest. That σ - v diagram was an example of a very low-regularised solution, so I should see a similar distribution for an unregularised fit. I fit an unregularised solution with 5 components as this is what I found was the ideal number of components for any one line of sight from my investigation of the effective n in Section 2.2.2. This unregularised solution gives four components that represent the bulk emission of the SMC main body. I identify these by their mean central velocity and take everything that has a mean $v > 120 \text{ km s}^{-1}$ to be the main body emission, shown in Figure 2.6. Adding all of these together to create a main body spectrum, it can be removed from the total spectrum leaving just the signal from the structures of interest. After this, the velocity range of the spectrum is reduced to exclude the main body range. This is done to avoid the possibility of any noise spikes left over from the main body subtraction affecting the fit, as well as to reduce the ROHSA computation time. Reducing the spectral range to velocities which have signal also makes the measure of χ_{red}^2 more reliable. To be sure that I did not remove more than was necessary, I show in Figure 2.7 the mean spectrum pre and post subtraction as well as 3 selected points in the field. It is clear that the removal did not affect the structures of interest and only removed the main body emission.

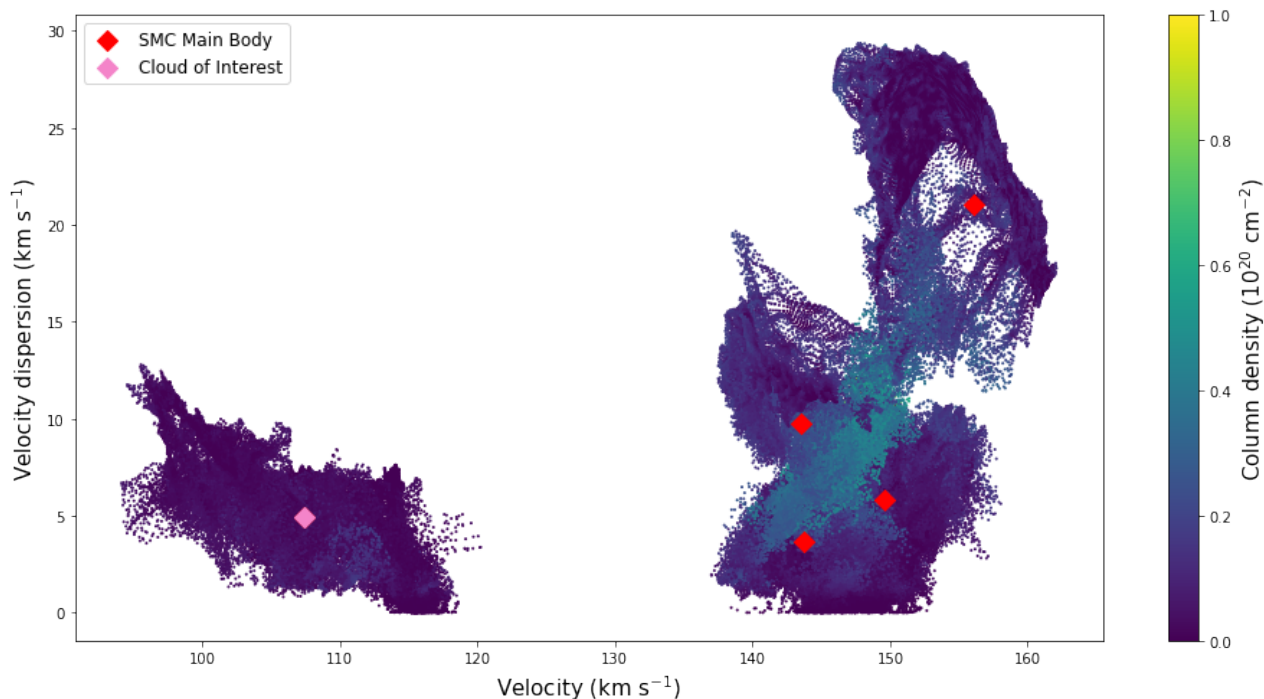


Figure 2.6: σ - v diagram of the unregularised fit to Field 1 to decide the regions for subtraction. The mean velocity and σ of components that were subtracted are marked in red, components kept are marked in pink.

I repeated this main body subtraction process for the other two fields that I am analysing and in Figure 2.8 I show the difference before and after to show that the removal of the main body emission has not affected the remaining signal in the structures.

Another issue that I encountered in Field 1 was the presence of two smaller structures in the velocity range of the larger cloud, seen in Figure 2.9. These smaller structures had the potential to confuse the fit, however they were

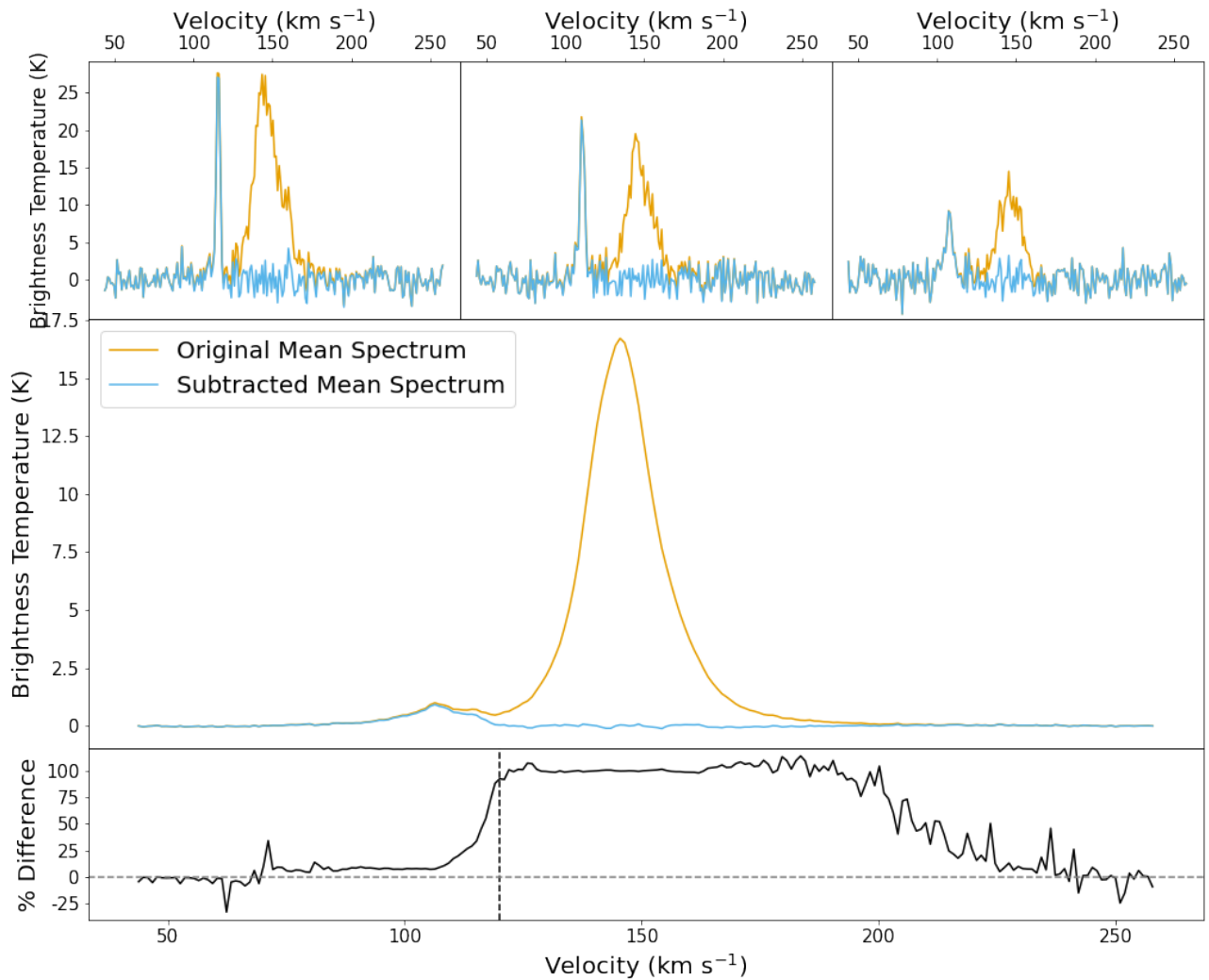


Figure 2.7: Middle: Mean spectrum of Field 1 before (orange) and after (blue) subtraction of main body. Top: example spectra of three individual sightlines before and after subtraction. Bottom: Mean % difference from original spectrum along the spectral axis. Dotted black line indicates where the emission from the feature of interest begins.

fortunately physically separate from the larger cloud of interest. To avoid any potential confusion in the solution fitting, I followed a similar process to the main body subtraction. The main difference was that I just fit the regions around the two small clumps instead of the whole field. They were both easily fit and the relevant components to remove were identified by the central velocity. Figure 2.9 demonstrates the before and after of the subtraction. When subtracting in this way, there is little change to the emission brightness temperature inside and outside the boundaries of these subcubes, which preserves any wide component signal that could be there.

After removing the main body from all fields, I also reduce the spectral range of each field. For Fields 1 and 2 I limit the spectral range to be $v < 150 \text{ km s}^{-1}$, reducing the number of channels from 220 to 109. For Field 3 I limit the spectral range to be $v < 160 \text{ km s}^{-1}$, reducing the number of channels from 220 to 120. I have a slightly larger range for Field 3 because the mean signal from the cloud is around 10 km s^{-1} higher than in the other 2 fields, as is indicated in Figure 2.8.

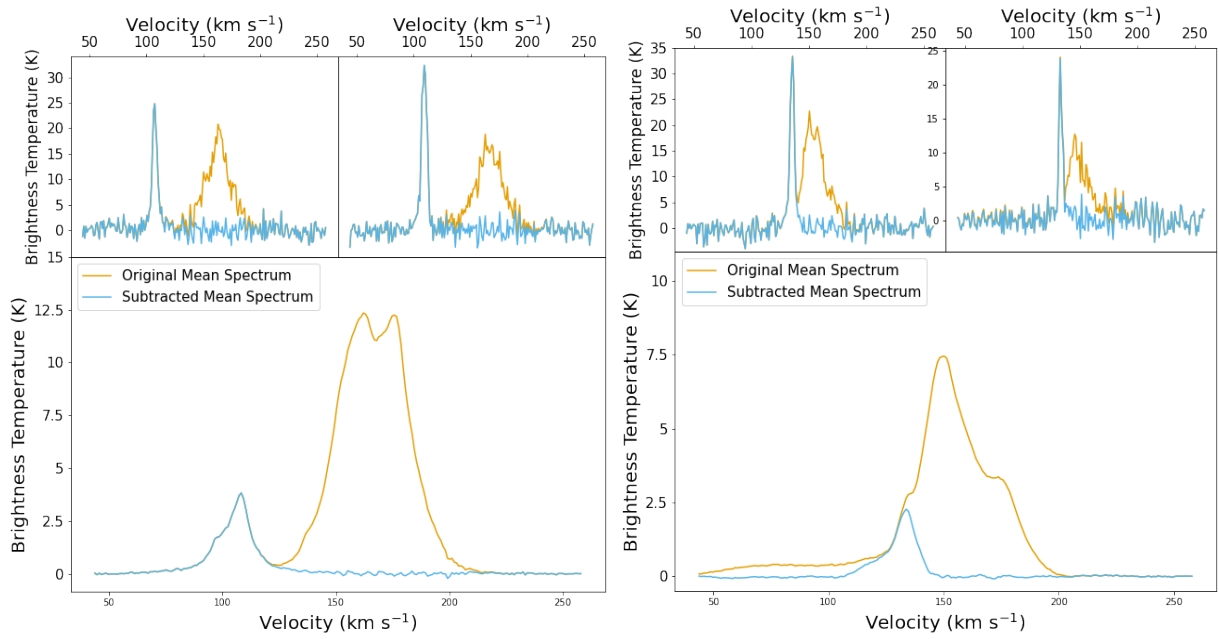


Figure 2.8: Bottom: Mean spectrum of the Field 2 (left) and Field 3 (right) before (orange) and after (blue) subtraction of main body. Top: example spectra of two individual sightlines before and after subtraction for Field 2 (left) and Field 3 (right).

2.2.4 Determining the best fit

After removal of the main body emission, it is now much more straightforward to fit the structures of interest. Now again I want to optimise the hyperparameters used by ROHSA as I did initially, but now I can fit far fewer components. For this section, I will look at Field 2 to show the quicker optimisation method that was used.

Field 2 is a slightly smaller field at $241 \times 236 \times 109$ voxels. From what was learned from the initial exploration with Field 1, I was able to reduce the computation time by exploring a smaller range of values. Since I am analysing the same overall datacube which has structures all at the same spatial scale across the whole field, I would not expect the ideal values for the hyperparameters to change significantly. This field is also nearby to the first field that was analysed ($\Delta RA \approx 13'$ and $\Delta DEC \approx 30'$) and so there should not be a large difference in the amount of regularisation needed. It is more important to obtain a good guess for the number of components required. Now by looking at unregularised fit, it seems that there is a large spread of velocity dispersions in this field. So with that in mind I assume that we have a multi-phase medium meaning at least two components, so I choose to explore the range of 2-4 components (n) and a hyperparameter range of 1-100 in \log_{10} space for all four hyperparameters as well as 3 values of σ_{init} of 3, 5, and 7. This gives me 729 solutions. The trend for all of the hyperparameters and over the 3 values for each the χ_{red}^2 value remains fairly constant. The range of χ_{red}^2 values is also greatly reduced compared to the last exploration. The values range from 1-1.2, which are all well-fitting solutions. The f_{CNM} distribution peaks near 0.3 on average for the field, which suggests a preference for one type of solution. This means that the solutions are fairly similar in what they are showing, the main difference that we would see is in the way each component

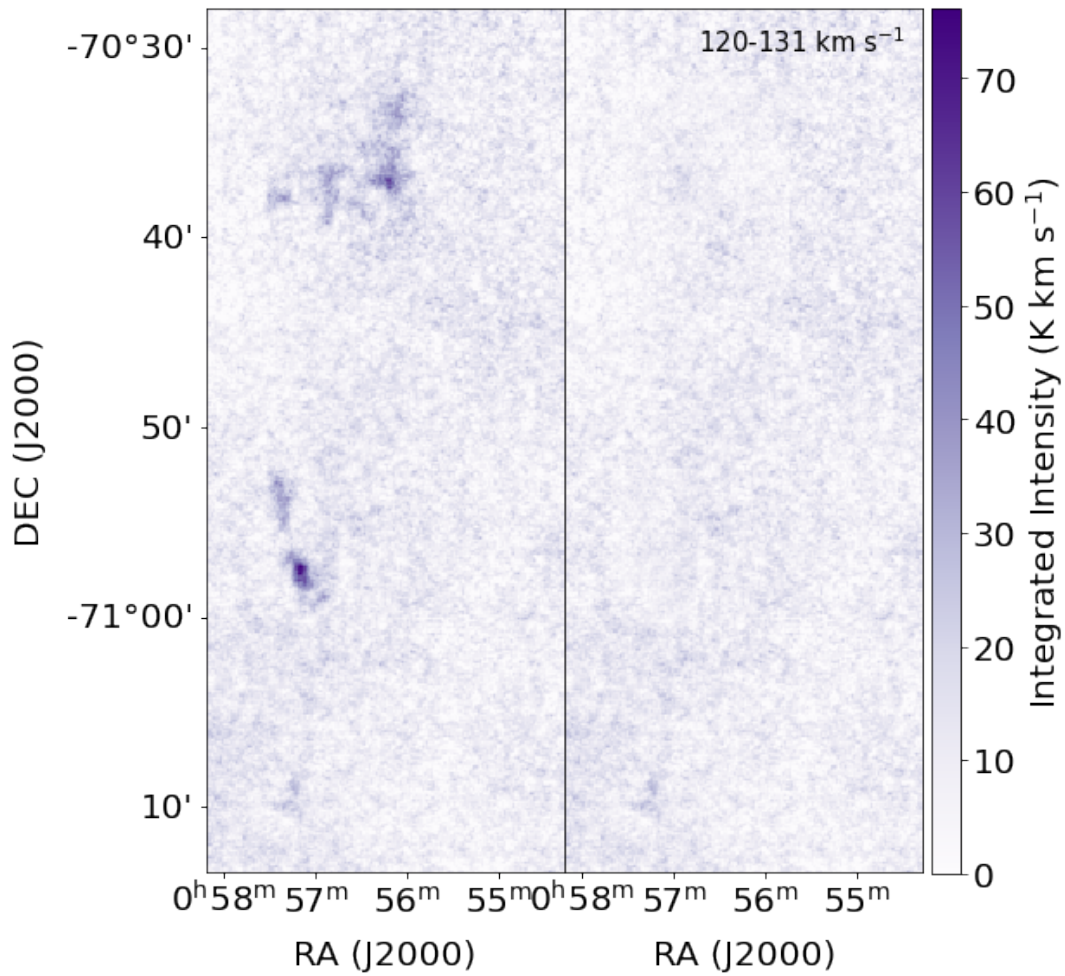


Figure 2.9: Column density integrated over $120\text{-}131 \text{ km s}^{-1}$, before (left) and after (right) removal of clumps that are present in Field 1.

divides the solution, that is, how many components correspond to each phase.

I expect that each component should trace either the CNM, UNM or WNM and so expect that the spread in the σ value for a good solution should be relatively small. If we look at a comparison of the σ - v diagram for a λ'_σ of 1 vs 100 in Figure 2.10 then we can see that the spread of σ for each component in the low λ'_σ case is very large. So in this case I chose the highest value for λ'_σ because this is a more physically motivated solution. The other hyperparameter that has an effect on the fit is λ_a . While the amplitude does not directly have an effect on the width of a component, when fitting multiple components if the amplitude is regulated too strongly it can have an indirect effect. The errors created by an amplitude that is incorrect can lead to the neighbouring component being fit too wide or narrow. In Figure 2.11 I show a solution with a λ_a of 1 vs 100. It is clear that the highly regularised solution pushes some components into a less coherent velocity dispersion structure which we see as this 'drip' down to lower velocity dispersions for the wide component in this fit.

So for this field, it seems that a solution with $\lambda_a = 1$ and $\lambda'_\sigma = 100$ is ideal. Since the other hyperparameters

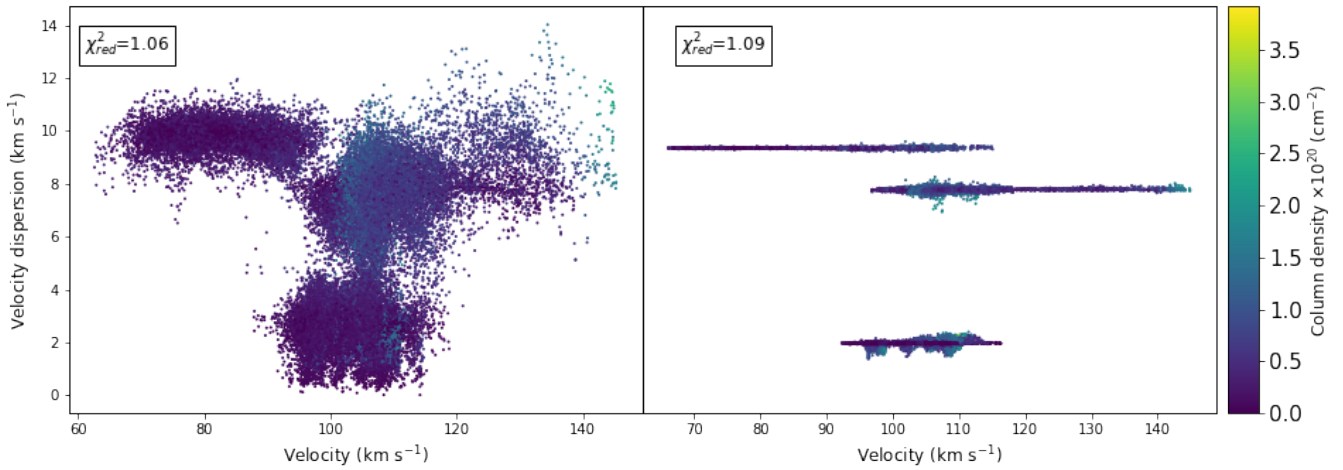


Figure 2.10: σ - v diagram of solution with $\lambda'_\sigma = 1$ (left) and $\lambda'_\sigma = 100$ (right), the other hyperparameters are kept constant.

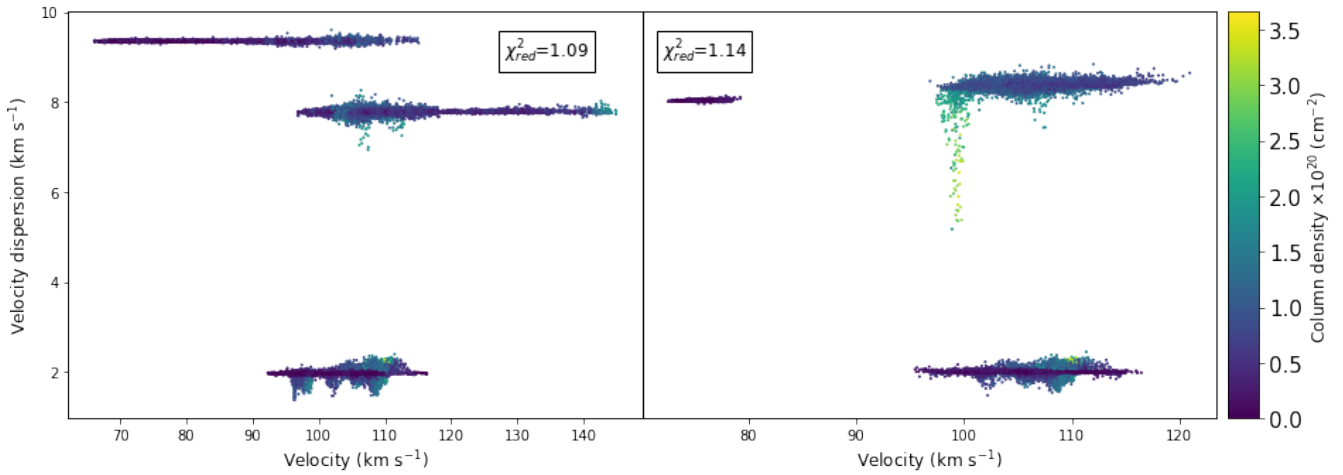


Figure 2.11: σ - v diagram of solution with $\lambda_a = 1$ (left) and $\lambda_a = 100$ (right), the other hyperparameters are kept constant.

have little effect on the goodness of fit, I set them to be in line with λ_a at a value of 1. The last thing to check, is the distribution of the residuals in the solution. In Figure [2.12](#) the brightness temperature residuals follow a normal distribution (with a skewness of 0.02) and in the χ^2_{red} map there is minimal coherent spatial enhancements across the field. It is important to note, that these measures are taken for the velocity range of the structure of interest, $92 < v < 118 \text{ km s}^{-1}$. This means the χ^2_{red} values are higher than what the mean values reported for the fields so far, as the *DOF* have decreased.

These types of checks are important to do, because the mean value of χ^2_{red} for a certain solution can hide large discrepancies over concentrated areas. A spike in the χ^2_{red} value over a sub-range of velocities in a datacube is easily lost when you take mean values. Thus, looking at the χ^2_{red} values and the residuals over these ranges, instils more confidence that the solution fits the structures of interest well. There are two small increases in the χ^2_{red} value in this

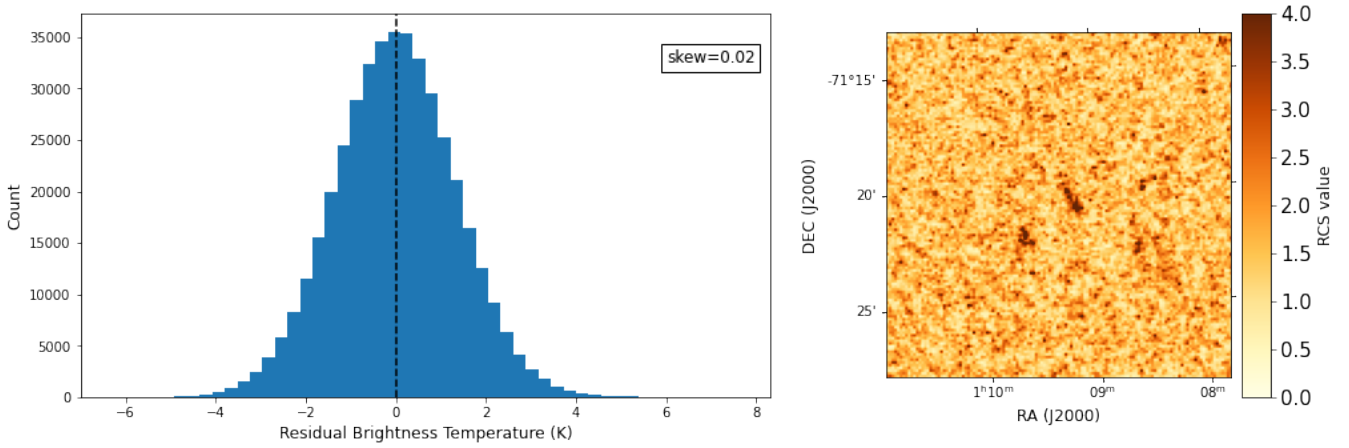


Figure 2.12: Field 2 histogram of the residuals in Brightness Temperature (left), with the skew of the distribution shown in top right corner and the spatial map of χ_{red}^2 values (right).

map for Field 2, however they do not coincide with the highest column density areas over this velocity range, so it is still reasonable to assume that the structure of interest is well fit by this solution.

I conduct the solution explorations after subtraction for Fields 1 and 3 using the same process and show their σ - v diagrams and residual maps in Figures 2.13 and 2.14.

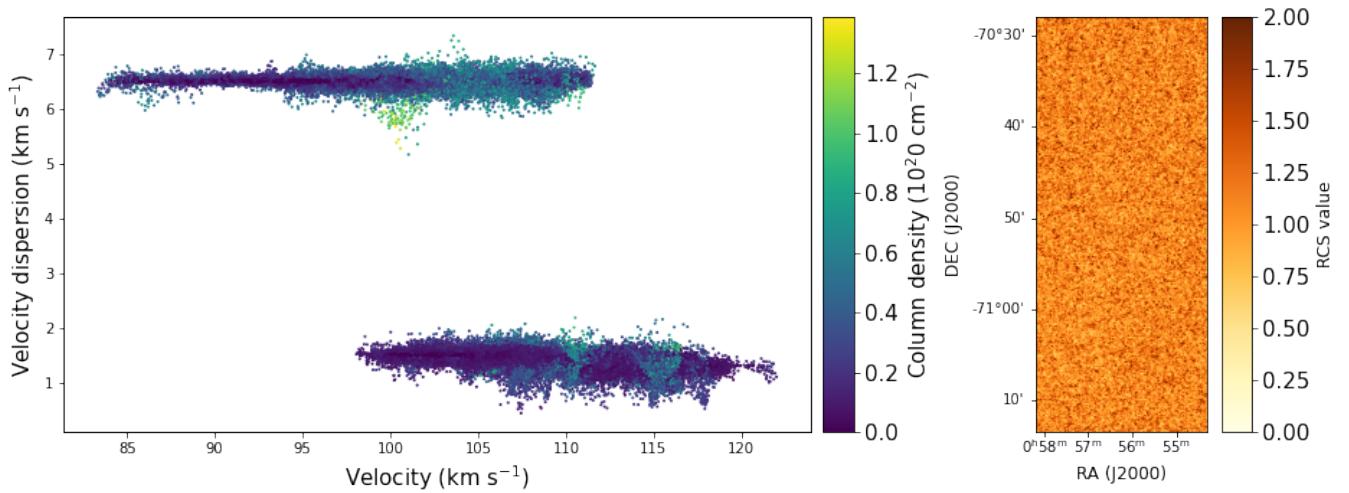


Figure 2.13: Field 1 σ - v diagram (left) and the spatial map of χ_{red}^2 values (right).

In Table 2.2 I summarise the final values of the hyperparameters that were chosen through the exploration process described previously for all fields.

In Figures 2.15, 2.16 and 2.17 I show the full solution maps for Fields 1, 2, and 3 respectively. The maps contain the 3 components (a , v , σ) of each Gaussian fitted to each field.

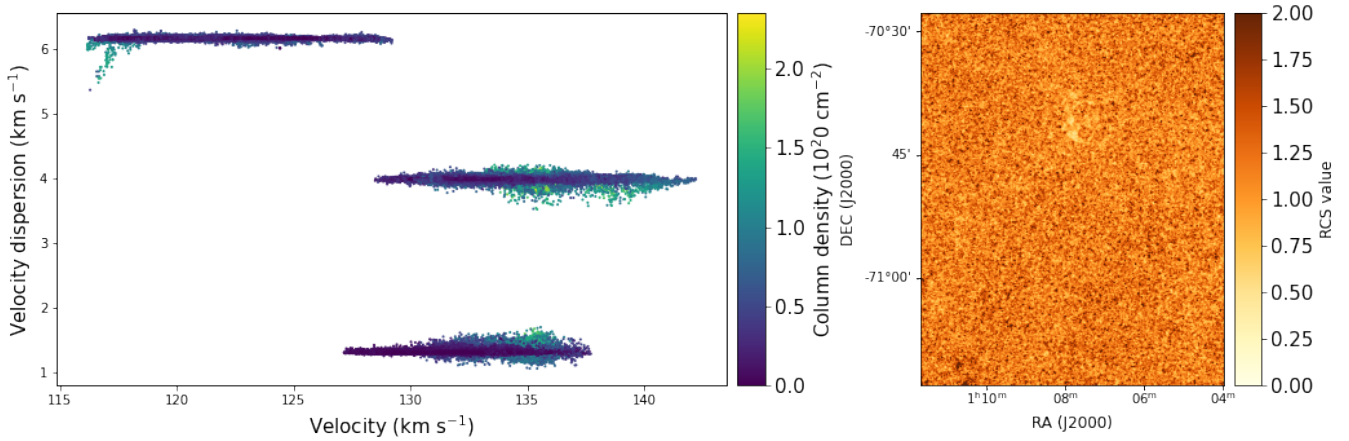


Figure 2.14: Field 3 σ - v diagram (left) and the spatial map of χ_{red}^2 values (right).

Field	n	λ_a	λ_μ	λ_σ	λ'_σ	$\overline{\chi_{\text{red}}^2}$
1	3	10	10	10	10	1.12
2	3	1	1	1	100	1.17
3	3	1	10	10	100	1.19

Table 2.2: Best solution ROHSA parameters for each field with their corresponding χ_{red}^2

2.3 Measuring the solution uncertainty

After using ROHSA to find a best model fitting model, I aimed to find the uncertainties in these models. I followed a similar method to that outlined in [Taank et al. \(2022\)](#). In this work I endeavour to obtain uncertainties on individual model components however, in [Taank et al. \(2022\)](#) they only obtain uncertainties on the column density of each phase, which are integrated quantities. While in that work they use 7 Gaussians to encode their data, increasing the chance of solution degeneracy, I only used 3 for each field. If each solution is stable, I will be able to obtain uncertainties on not just total column density in each phase, but the amplitudes, central velocities and velocity dispersions of each Gaussian fit. The version of the aforementioned method used in this work makes 100 runs each of; the final model injected with random noise, one of these noise realisations with hyperparameters varied by 10%; and one of the noise realisations with initial guesses provided to ROHSA sampled from the FWHM ranges of the original solution distribution.

More specifically for the noise realisations, I use a reconstruction of the PPV datacube from the model solution and the 3D noise map described in Section [2.1.1](#). Using the `numpy.random.normal` module, I supplied the 3D noise cube as the standard deviation for the function, which then produced a 3D noise cube, $N(x, y, v)$, with normally distributed values for each voxel. I then added this noise cube to the model cube ($M(x, y, v)$) to create a synthetic data cube $D_S(x, y, v)$, as below:

$$D_{S,j}(x, y, v) = M(x, y, v) + N_j(x, y, v) \quad (2.12)$$

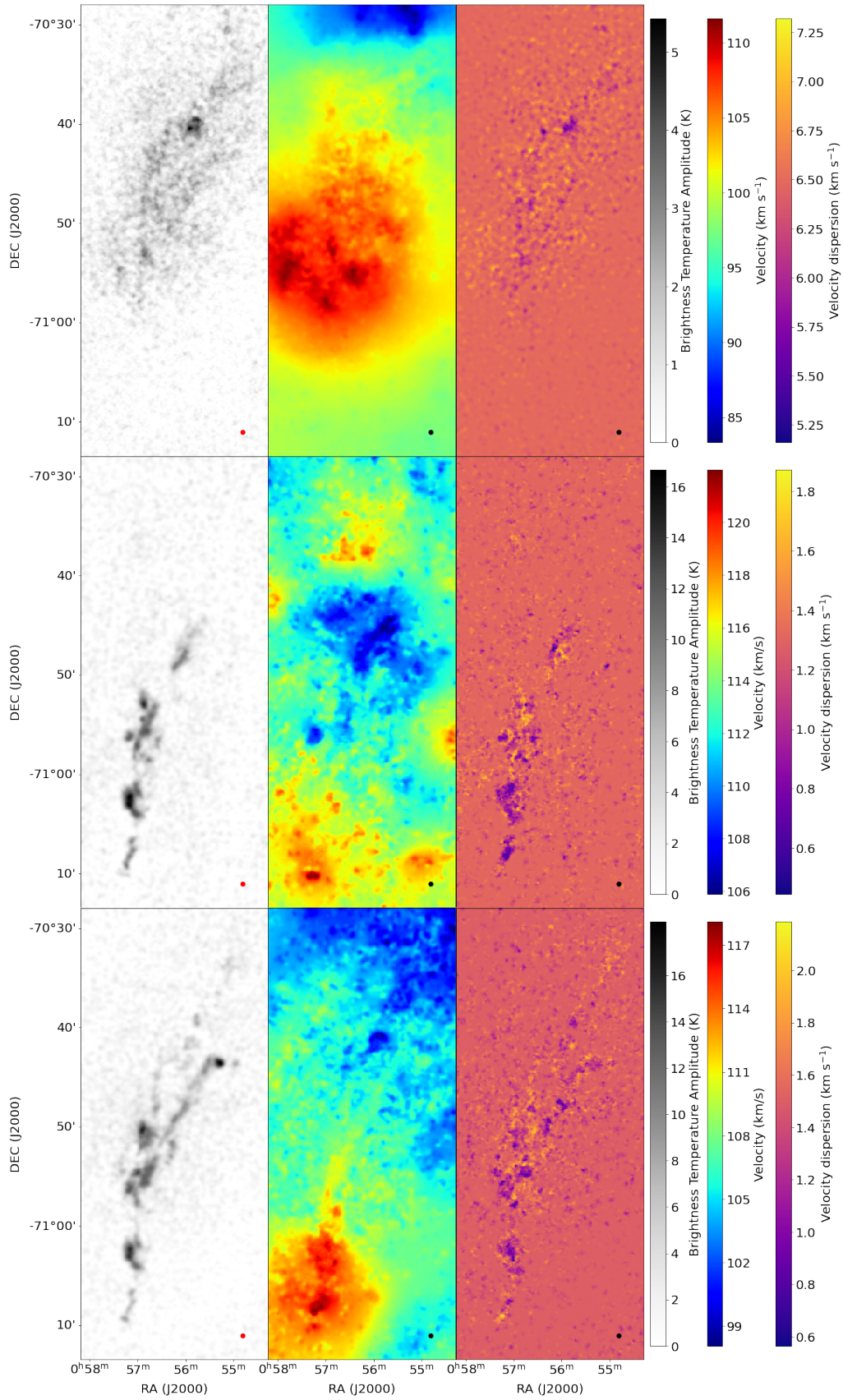


Figure 2.15: Full solution set for Field 1. Top: Component 1, Middle: Component 2, Bottom: Component 3. Left: Brightness Temperature Amplitude (α), Middle: Central Velocity (v), Right: Velocity Dispersion (σ). The red and black circles indicate the beamsize.

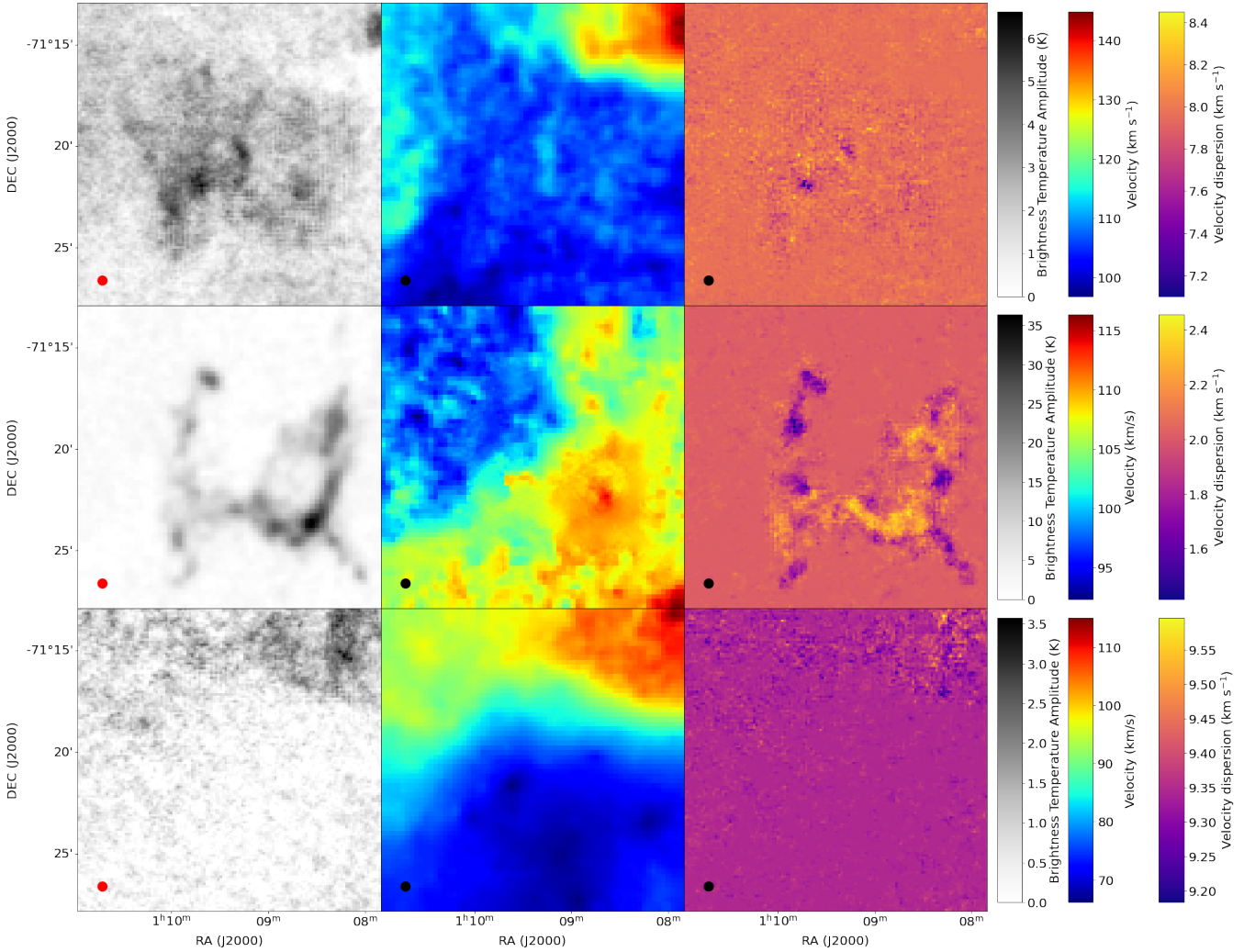


Figure 2.16: Full solution set for Field 2. Panels the same as Figure 2.15

where j represents the realisation number, ranging from 1-100. After creating $J = 100$ realisations I ran ROHSA on these 100 new datacubes, with parameters and hyperparameters identical to those of the best fit solution.

For the following methods, the input data cube does not need to vary, so I chose one noise realisation at random to use for both the following methods.

For the second method of uncertainty quantification, I take the four hyperparameters of the ROHSA algorithm and vary them by up to 10% using the `numpy.random.random` module which draws from a uniform distribution between 0 and 10%. All other parameters of the ROHSA run remain the same. 100 parameter files were generated with differing hyperparameters and ROHSA was run on each of them to obtain 100 solutions.

For the last method, I varied the initial guess provided to ROHSA. As outlined in Section 1.3.2 ROHSA starts decomposing from the largest scale and in doing so looks first at the mean spectrum of the whole field. Providing the algorithm with an initial guess of the parameters changes the starting point of the decomposition at the top level. In Taank et al. (2022) they found this was their largest source of error by a small amount. Changing the initial guess

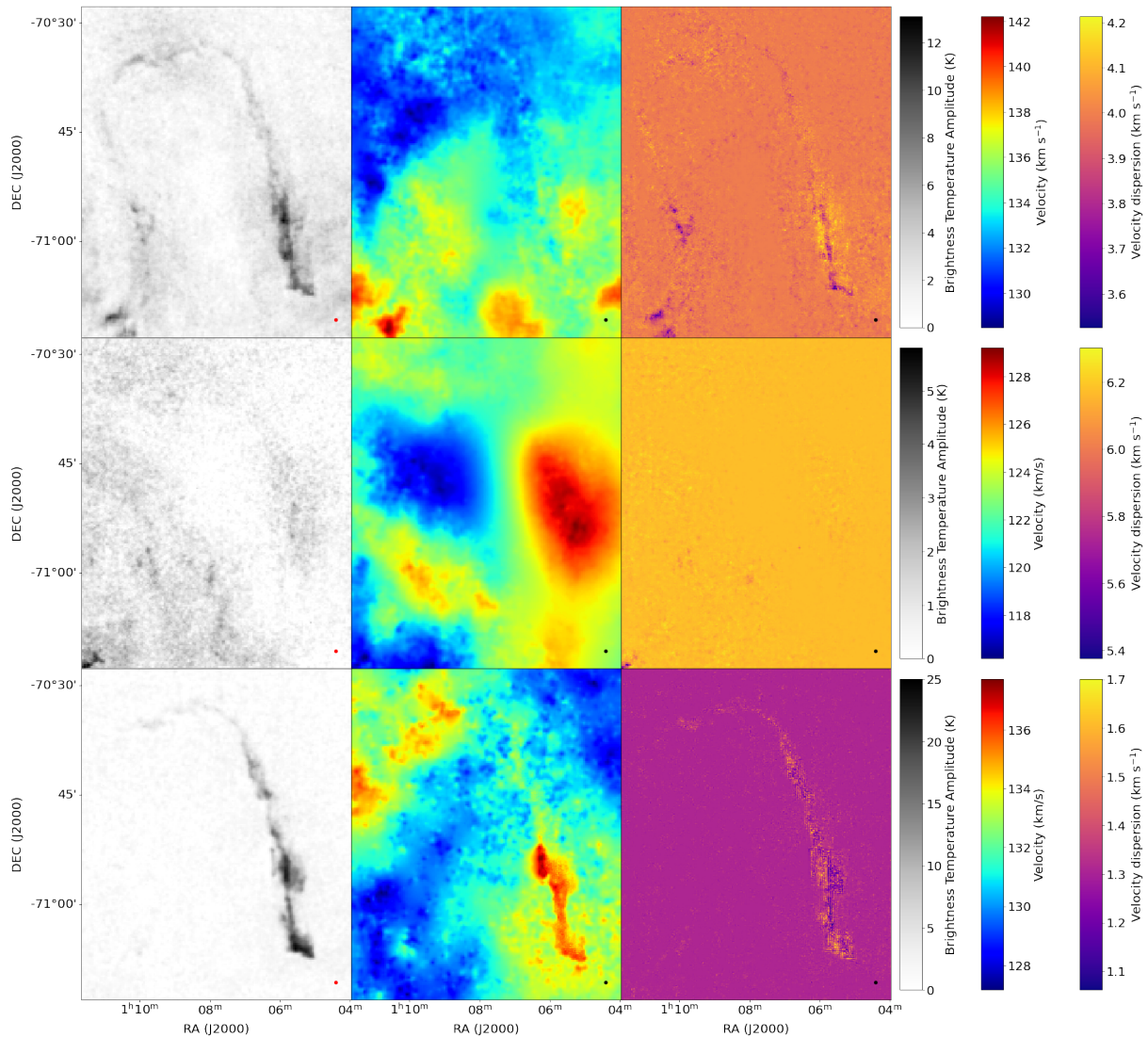


Figure 2.17: Full solution set for Field 3. Panels the same as Figure 2.15

of a decomposition can have large effects on the final solution if there are many degenerate solutions. Typically this is more likely if you are seeking solutions with high numbers of Gaussians, so many algorithms try to minimise the number of Gaussians, e.g. Lindner et al. (2015). I only fit 3 Gaussians to each field in this work, so I do not envision this will be the largest source of error.

The initial guesses were calculated by taking the distributions for each Gaussian component from the best fit model and measuring their spreads in the a , v , and σ axes. I took the FWHM of the distribution of each parameter of each Gaussian from the best fit model. I then randomly selected values from a uniform distribution within the range around the mean up to the value of this FWHM to use as the initial guess. I generated 100 initial guesses which were each supplied to a ROHSA run with the same parameters as the best fit solution.

Once all of these runs were completed, total uncertainties were able to be calculated. First I took the standard deviation of the 100 values from each uncertainty source; random noise (rn), hyperparameter variation (hp) and ini-

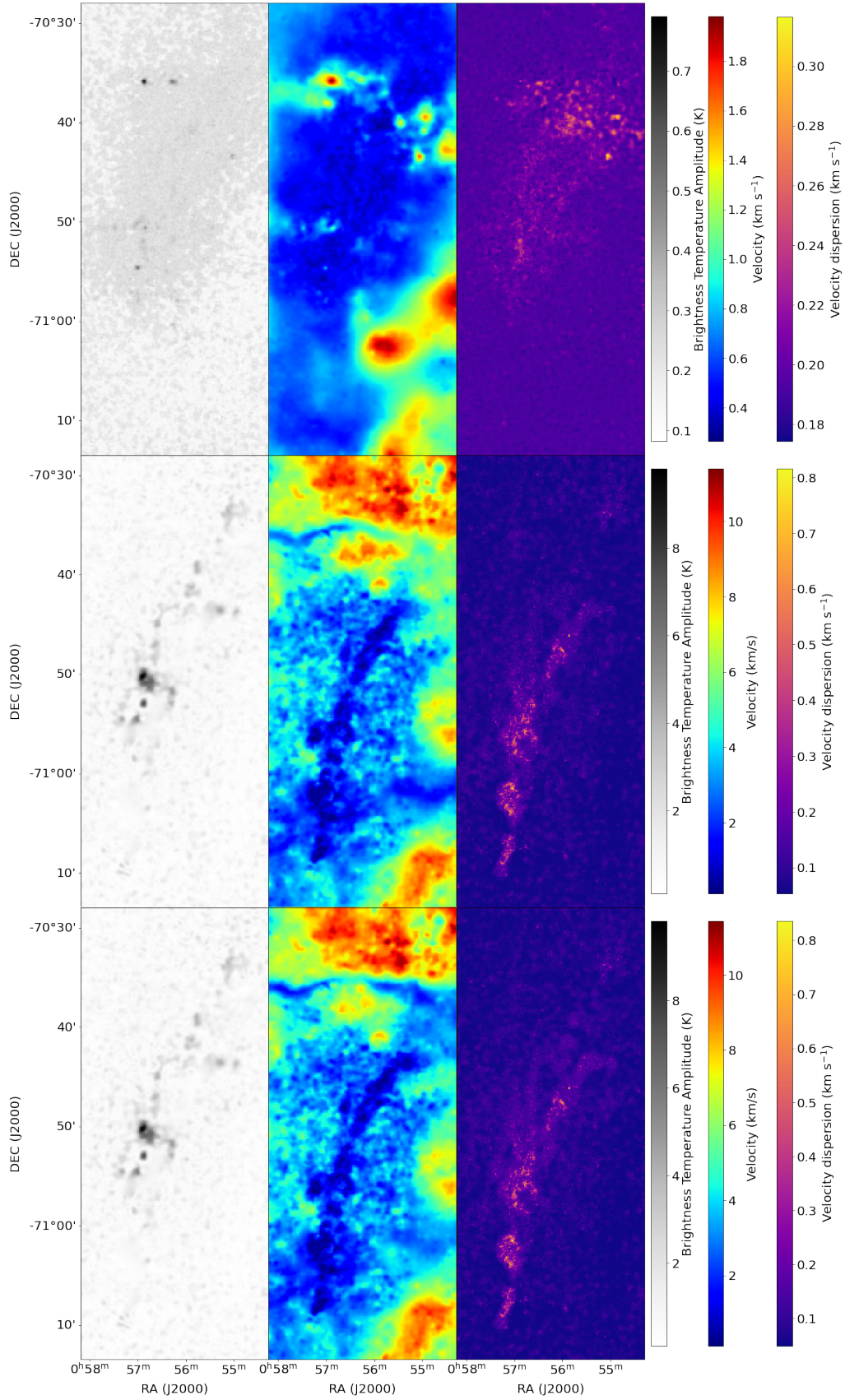


Figure 2.18: Full uncertainty map for Field 1. Top: Component 1, Middle: Component 2, Bottom: Component 3. Left: Brightness Temperature Amplitude (α), Middle: Central Velocity (v), Right: Velocity Dispersion (σ).

tial guess (ig); of each Gaussian component at each pixel. This creates three 3D uncertainty cubes of size $(X_n, Y_n, 3*N)$ where X_n and Y_n are the x and y dimensions of the original datacube and N is the number of Gaussians in the best fit solution. Then I add the 3 cubes together in quadrature as below:

$$\sigma(x, y, 3N)_{\text{tot}} = \sqrt{\sigma(x, y, 3N)_{rn}^2 + \sigma(x, y, 3N)_{hp}^2 + \sigma(x, y, 3N)_{ig}^2} \quad (2.13)$$

This results in one uncertainty cube of size $(X_n, Y_n, 3*N)$ that captures the total uncertainty of each fitted component. The map of the uncertainties for Field 1 is shown in Figure 2.18. There is evidently some confusion between the second and third component in the amplitude and velocity dispersion measures. This is likely because they have a similar velocity and velocity dispersion, evidenced in Figure 2.15. These components are both tracing narrow Gaussians and may change their fitting order between different uncertainty runs. This lead to the higher uncertainties appearing in the same spatial area for these two components.

I investigated the strength from each source of uncertainty compared to the others by taking the mean of each uncertainty divided by the total uncertainty. For example for the strength of the noise uncertainty is calculated as follows:

$$S_n(3N) = \frac{1}{XY} \sum_X \sum_Y \frac{\sigma(x, y, 3N)_{rn}}{\sigma(x, y, 3N)_{\text{tot}}} \quad (2.14)$$

where X and Y are the number of pixels in the x and y axes respectively. This leaves us with an array of $3*N$ numbers representing the relative strength of the noise uncertainty for each Gaussian component. After doing this for the other two uncertainty sources, I normalised the values such that the relative uncertainty strengths for the 3 sources, for each component sum to 1. In Figure 2.19 I show the ratios for each Gaussian component for each cloud. The last column for each field in Figure 2.19 shows the mean ratio across all the components for each field. In hard numbers the mean ratio of $\sigma_{rn} : \sigma_{hp} : \sigma_{ig}$ for Field 1 is 55:32:13, for Field 2 is 48:25:27, and for the Field 3 is

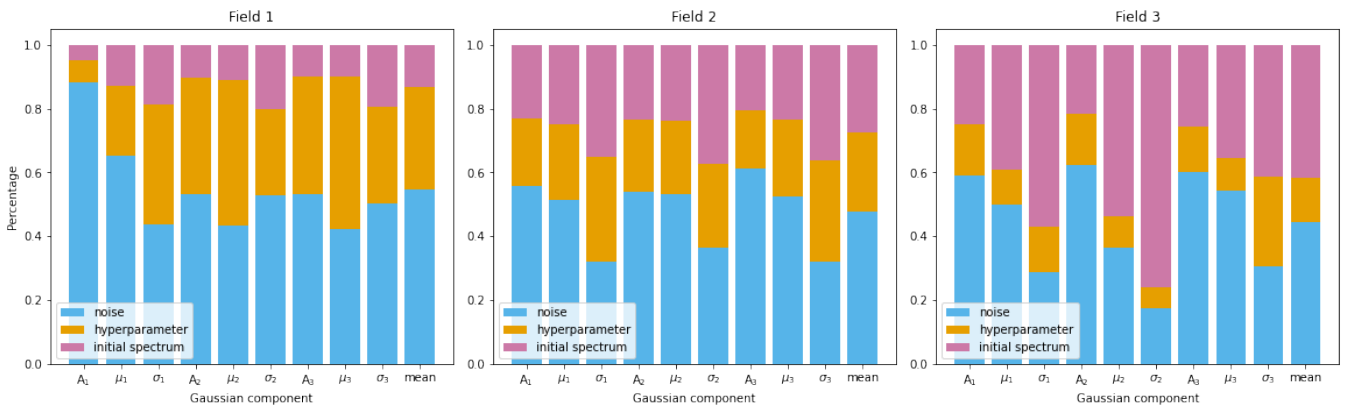


Figure 2.19: The relative strengths from Equation 2.14 of each source of uncertainty in Field 1 (left), Field 2 (middle), Field 3 (right) for each component of each Gaussian as well as the mean of all components.

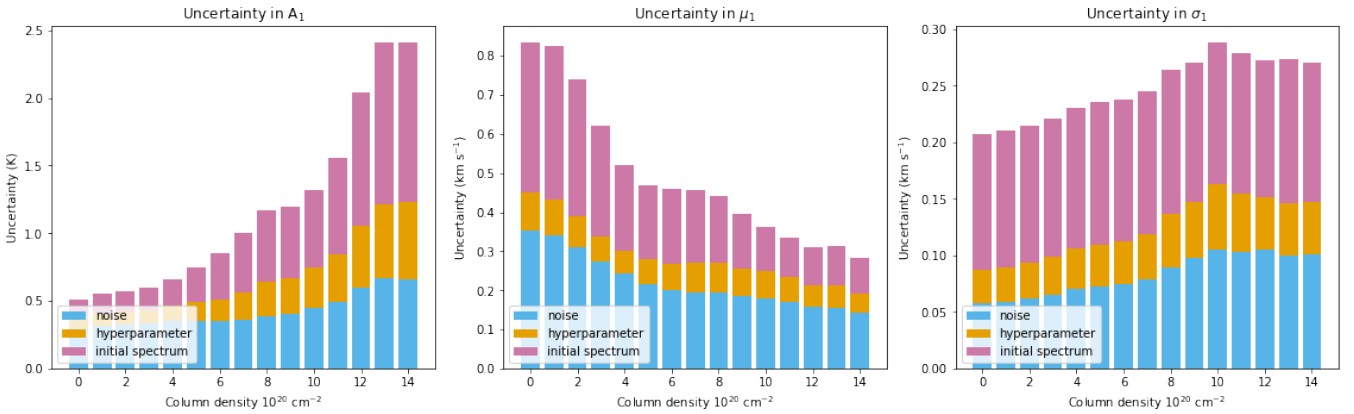


Figure 2.20: The total uncertainty from all three sources across the first component of the Field 3 fit. Left: uncertainty in the amplitude, Middle: uncertainty in the central velocity, Right: uncertainty in the velocity dispersion.

44:14:42. In all 3 cases the uncertainty from the noise error contributes most to the total uncertainty, but the other two sources have differing effects depending on the field.

If we look at the influence of each source on the individual Gaussian components, we see that the influence of the initial guess error is higher relative to the other sources, especially the noise error, for the velocity dispersion components. The velocity dispersion directly influences the column density and gas temperature measurements that we make from decompositions and clearly the initial guess these decompositions work from influences the solution more than small noise or hyperparameter variations.

Additionally if we look at the break down of the uncertainty as a function of column density in Figure 2.20 we find that the total uncertainty in amplitude is positively correlated with column density. Conversely the total uncertainty in velocity is negatively correlated with column density. The latter is easy to understand. When the column density is low, there is little signal, in this case one or even none of the Gaussians are needed to encode the signal. So one or more of the Gaussians will have an amplitude of effectively 0. At this point, the central velocity could have any value and it would still give the same solution, a flat line, for those components. The cause of the relationship between column density and amplitude uncertainty is not clear.

For comparison, I also measure is the uncertainty in the column density of each phase of the HI, just like in Taank et al. (2022). They make this measurement as whilst the individual components that make up these phases may vary significantly in their parameters, the total column density for each phase should vary less for a solution. So to measure this for each cloud, I follow the same method outlined in Taank et al. (2022), calculate the CNM and WNM column density for every iteration from the three different sources of error. I then take the standard deviation at each point of the image across all 300 column density maps as the uncertainty. The mean uncertainty in column density and the relative uncertainty ($\sigma_{\text{NH,CNM}}/N_{\text{H,CNM}}$) for each field of the CNM and WNM column densities is reported in Table 2.3. The means are weighted by column density, to provide relative values that better reflect the relative uncertainty in high column density areas. This is because there are large parts of the fields fit in this Section

Field number	$\sigma_{\text{NH,CNM}} 10^{20} \text{ cm}^{-2}$	$\sigma_{\text{NH,CNM}}/N_{\text{H,CNM}}$	$\sigma_{\text{NH,WNM}} 10^{19} \text{ cm}^{-2}$	$\sigma_{\text{NH,WNM}}/N_{\text{H,WNM}}$
1	0.40	0.215	0.70	0.312
2	1.07	0.214	1.66	0.239
3	0.49	0.328	0.77	0.204

Table 2.3: Measured uncertainties for the CNM and WNM HI phases for each field.

that have low column densities, but similar uncertainties to the high column densities areas, which would make the relative uncertainties high. The mean of these relative uncertainties across each field and phase is 25.2%. These relative uncertainties, are comparable to the ratio of the mean uncertainty to the mean column density reported in [Taank et al. \(2022\)](#). The division of the solutions into WNM and CNM will be defined in the next Section.

Chapter 3

Phase relationship of HI in 3 large clouds of the SMC

This chapter focuses on analysing the results obtained for the 3 fields that contain large HI clouds, outlined in Chapter 2. I look at the phase composition composition of each cloud as well as how that connects to the kinematics of each cloud.

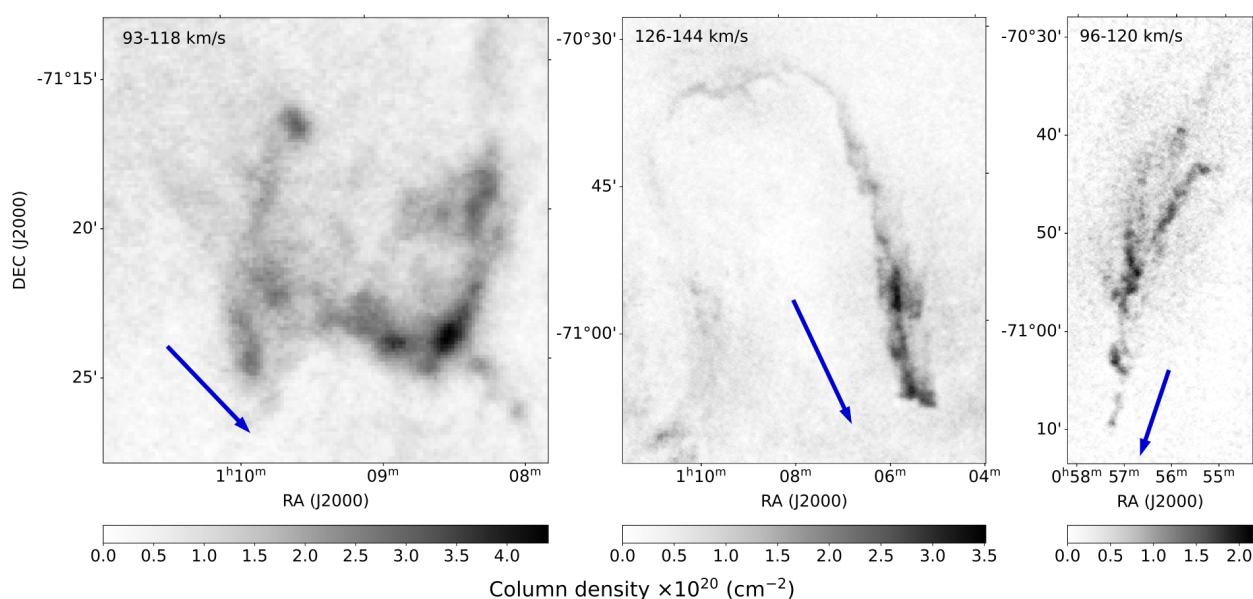


Figure 3.1: Column density of all three clouds (Left: Alpha, Middle: Hook, Right: Gamma) integrated over the velocity range indicated in each panel. The blue arrows indicate the direction of the dynamical centre of the SMC from Di Teodoro et al. (2019b).

In Figure 3.1 I show the total column density for each cloud in Fields 2, 3, and 1 which are referred to from this point on as Alpha, Hook, and Gamma respectively, over the velocity range they span. To orient these clouds in

relation to the SMC, I also indicate the direction of the dynamical centre as reported in [Di Teodoro et al. \(2019b\)](#). Each cloud has a different morphology, with the Alpha cloud being the smallest of the three measuring 240pc across the longest diagonal (north-east to south-west), with an irregular morphology. It lies 4.34 kpc away from the SMC centre, assuming they are at the same distance along the line of sight. The Hook cloud is primarily composed of the strong ridge along the west side forming a long filament that spans 610pc. It also continues looping around the north and down the east of the field, albeit with lower column densities. It lies 4.86 kpc away from the SMC centre, with the same assumptions. The Gamma cloud has a strong core filament, appears to be broken into multiple clumps. It covers a distance of 560pc, similar to the Hook cloud, but with a clumpier appearance. It lies 4.85 kpc away from the SMC centre, again with the same distance assumptions. These distances are calculated with the relation from Equation [3.4](#) with a distance to the SMC which will be detailed in Section [3.5](#).

3.1 Best fit models

Following the fitting process outlined in Chapter 2, the mean values for all the components of each fitted Gaussian are listed in Table [3.1](#). The components are classified as CNM or WNM based on the mean maximum kinetic temperature, $\langle T_k \rangle$, measured from the mean Gaussian dispersion (σ_v) for each component from the the general equation (Equation [1.10](#)):

$$\langle T_k \rangle = 121 \langle \sigma_v \rangle^2 \quad (3.1)$$

Components with $\langle T_k \rangle < 500K$ are classified as CNM, those with $\langle T_k \rangle > 500K$ are classified as WNM, in the stable and unstable regimes, in line with [Heiles and Troland \(2003b\)](#). It is important to note that the measured dispersions can be broadened by non-thermal motions, such as turbulence. As there is no measure of the turbulence along individual lines of sight, I do not consider the effects of turbulence on the values of the measured dispersions in this work. For this reason, I refer to these temperatures as maximums, as they represent the case where all broadening is attributed to thermal motions.

For the Alpha cloud field I have a 3 component model, with 2 of these components clearly corresponding to the cloud in the column density and velocity ranges, shown in Figure [2.16](#). The CNM component has $\langle T_k \rangle = 470K$ and

Cloud	Component 1				Component 2				Component 3			
	A (K)	v (km s ⁻¹)	σ_v (km s ⁻¹)	$\langle T_k \rangle$ (K)	A (K)	v (km s ⁻¹)	σ_v (km s ⁻¹)	$\langle T_k \rangle$ (K)	A (K)	v (km s ⁻¹)	σ_v (km s ⁻¹)	$\langle T_k \rangle$ (K)
Gamma	0.99	102	6.48	5081	1.81	113	1.29	201	2.65	107	1.48	265
Alpha	2.35	108	7.77	7305	6.51	104	1.97	470	0.43	80	9.35	10578
Hook	2.87	135	3.99	1926	0.71	124	6.17	4606	2.84	132	1.32	211

Table 3.1: Fit results, A , v , σ_v , for the components fit to each cloud and the corresponding mean maximum kinetic temperature of each component.

the WNM component has $\langle T_k \rangle = 7305K$. In Figure 3.2 it is clear that CNM component traces the strong filamentary structure of the cloud as it is seen in the integrated data in Figure 3.1 whereas the WNM component accounts for the more diffuse background. When combining the fainter tracing of the WNM with the fact that the two components travel at similar velocities (seen in Figure 3.2) it is clear that the two components are related and both contribute to the whole structure. For this fit the χ_{red}^2 value is 1.17 and the column density weighted mean values for each component parameter are listed in Table 3.1. The means are column density weighted due to the fact that the clouds do not fill the field that was fit, so the unweighted mean can be affected by the values in the low column density areas that are quite uncertain and of little significance. The third unrelated component that is obtained for this fit has a large mean velocity dispersion which puts in the WNM range. It is offset by 25-30 km s^{-1} on average from the cloud and is likely diffuse, low-level WNM emission at the extreme velocity ends of the SMC. The contribution to the overall column density of the cloud from this component is negligible.

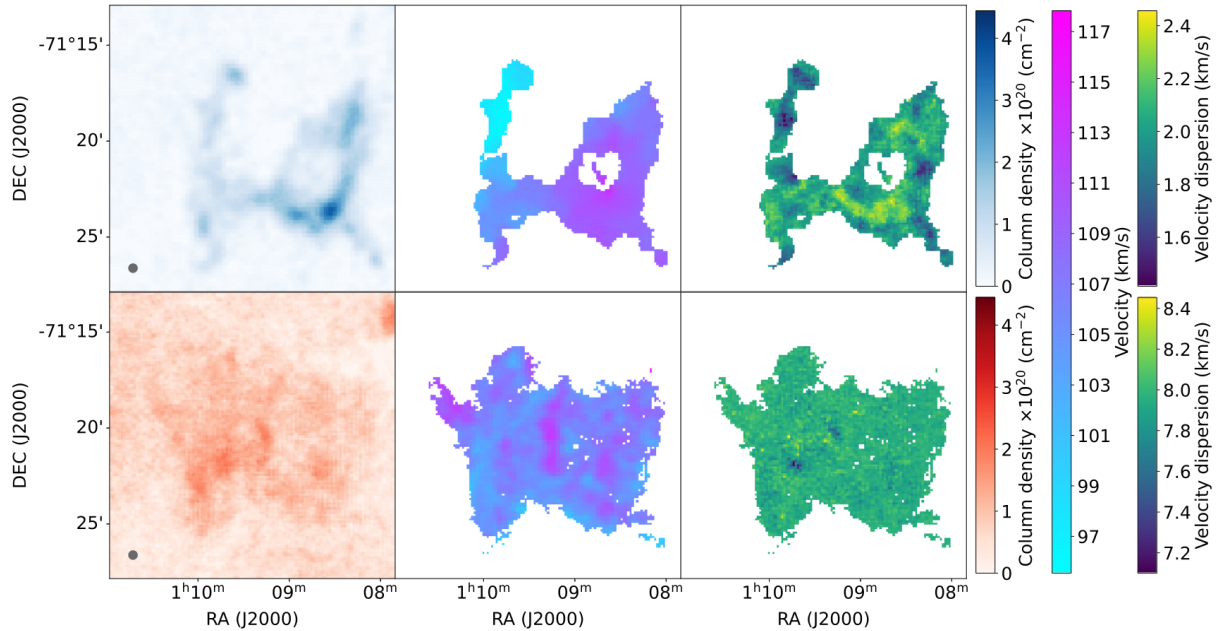


Figure 3.2: Spatial maps of the fitted properties of the Alpha cloud from ROHSA, within a $3 \times 10^{19} \text{ cm}^{-2}$ column density contour for each phase. Top: CNM, Bottom: WNM, Left: Column Density, Centre: Central Velocity, Right: Velocity Dispersion. The grey ellipse in the column density maps indicates the beamsize.

For the Gamma cloud field I have a 3 component model, with all 3 corresponding to the cloud, shown in Figure 2.15. There are two CNM components, with $\langle T_k \rangle = 201K$ and $\langle T_k \rangle = 265K$ and one WNM component with $\langle T_k \rangle = 5081K$. Like in the Alpha cloud, it is evident that the CNM and WNM components overlap in velocity space, shown in Figure 3.3. For this fit the χ_{red}^2 value is 1.12 and the column density weighted mean values for each component parameter are listed in Table 3.1. This fit required 2 components to fully capture the CNM distribution as there are strong, narrow signals at offset velocities in parts of this field. Table 3.1 shows that they are on average separated by 6 km s^{-1} . The advantage of the regularisation conditions that ROHSA imposes means that we can trace how these

components behave individually over the physical axes, which I will discuss further in Section 3.3. For the purpose of mapping the velocity and velocity dispersion of the total CNM, I take the column density weighted mean of the two components in Figure 3.3.

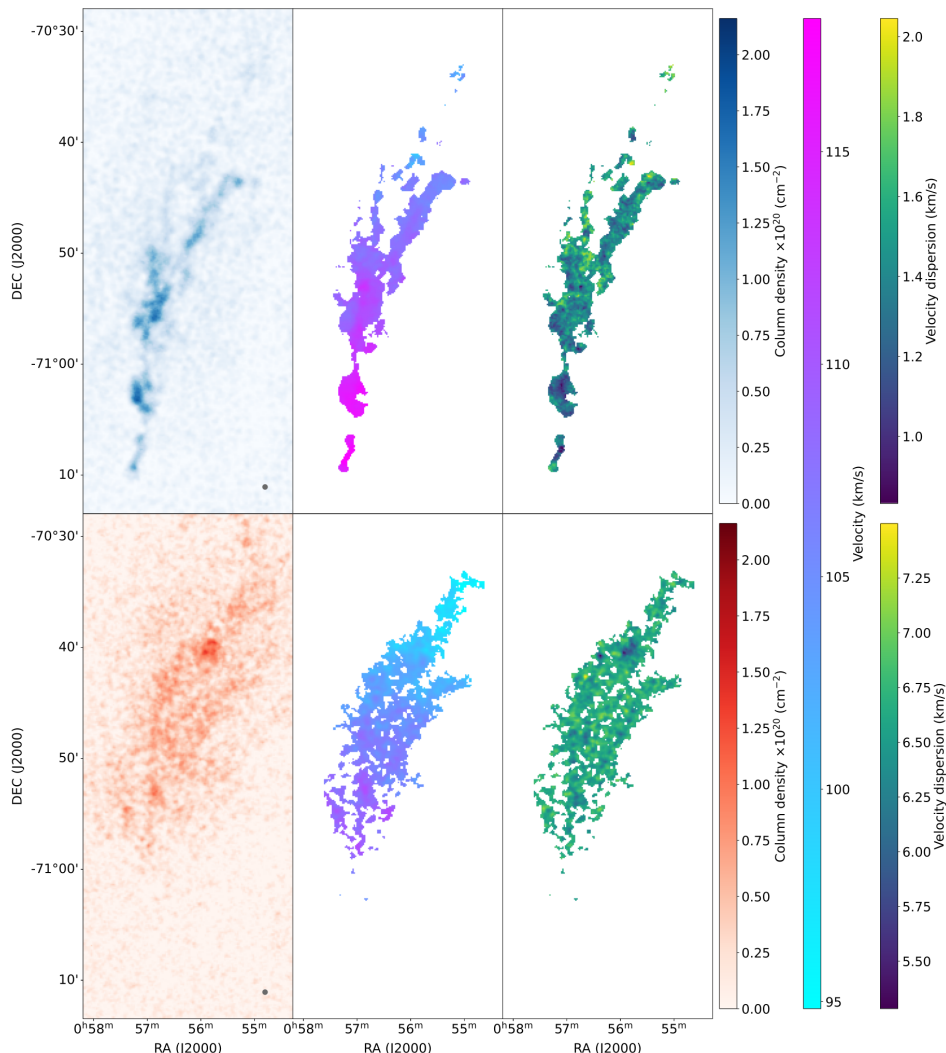


Figure 3.3: Same as Figure 3.2 but for the Gamma cloud.

For the Hook cloud field I again have 3 component model, and like the Alpha cloud just 2 out of these 3 components correspond to the Hook cloud, shown in Figure 2.17. The CNM component has $\langle T_k \rangle = 211K$ and the WNM component has $\langle T_k \rangle = 1926K$. This temperature is on the lower end of the temperatures for WNM, and would more appropriately be considered UNM which is unstable gas, capable of forming either WNM or CNM if perturbed out of equilibrium. The CNM component traces the strongest part of the hook shape along the west and into the north, but does not extend into the eastern part of the field. The UNM component envelops the CNM component and loops all the way around the field from west to east. From Figure 3.4 it is clear that they have very similar central velocities for the areas where they overlap in physical space. For this fit the χ_{red}^2 value is 1.19 and the column density weighted mean values for each component parameter are listed in Table 3.1. The third component for this fit is unrelated

to the Hook cloud. It does not trace any structure of the cloud, and does not follow the same velocity range as the other 2 components, which is evident from Figure 2.17. It is likely a tracing diffuse WNM component, as in the Alpha cloud.

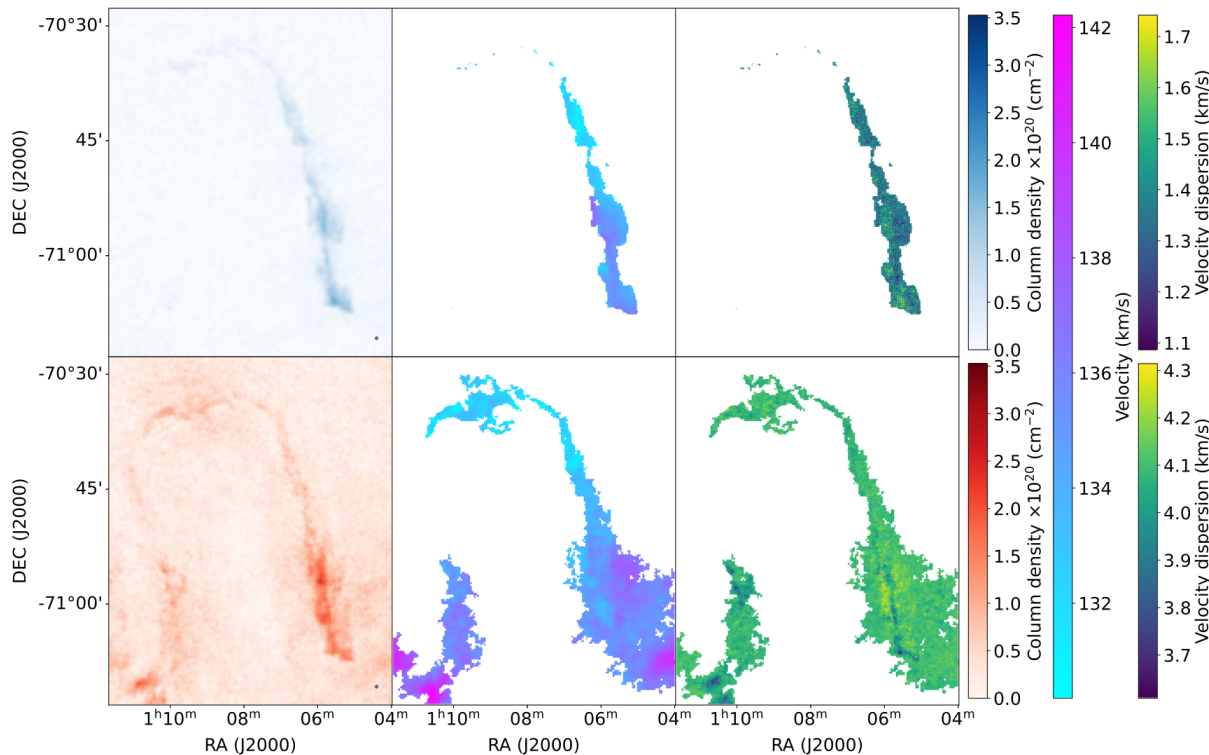


Figure 3.4: Same as Figure 3.2 but for the Hook cloud.

3.2 Phase distribution

Investigating the phase distribution continuously across regions of the SMC through Gaussian decomposition is something that has not been explored in the SMC yet, as measurements of the CNM fraction have only been made towards background sources in absorption studies (Dickey et al. 2000, Jameson et al. 2019, Dempsey et al. 2022). To be able to map the CNM fraction and the dynamics of the HI in the CNM phase continuously across a field is a new step that I take in this work.

The column density for HI is calculated assuming the gas is in the optically-thin regime from Equation 2.8. The optically-thin regime is a reasonable assumption in this case since these clouds have maximum integrated column densities, all below $5 \times 10^{20} \text{ cm}^{-2}$. In Dempsey et al. (2022) they show that the HI column density correction factor for optical depth is less than 1.05 for uncorrected column densities below 10^{21} cm^{-2} . Therefore I am confident that in the relatively low column density environments like these clouds, the optically-thin assumption is reasonable.

As the HI spectra in these fields are modelled as a sum of N Gaussians along the ν axis, the column density of

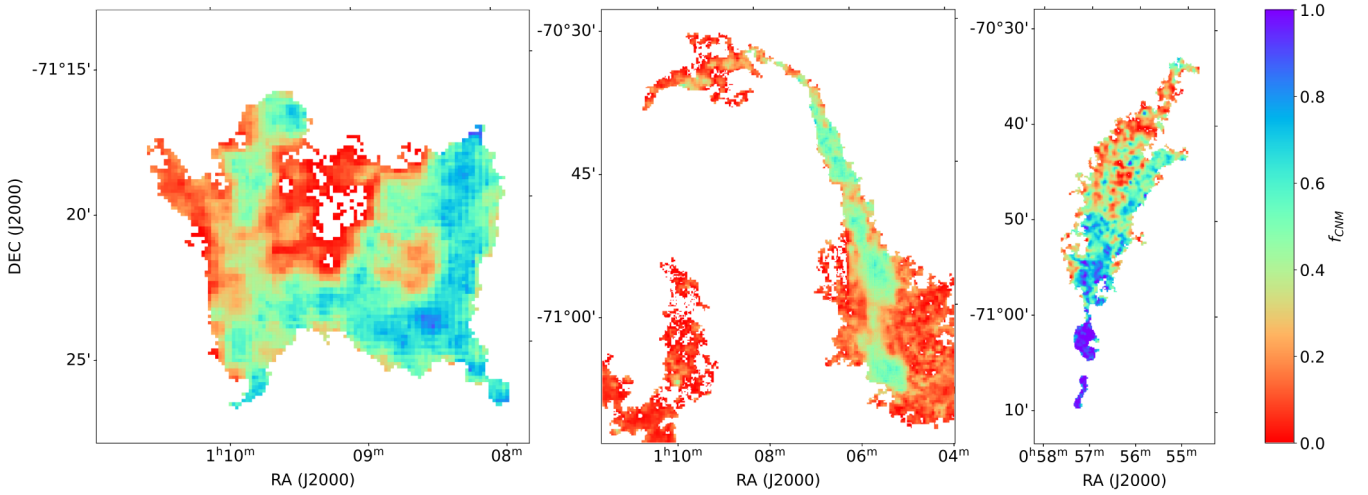


Figure 3.5: CNM fraction for all three clouds (Left: Alpha, Middle: Hook, Right: Gamma) within a $5 \times 10^{19} \text{ cm}^{-2}$ total column density contour.

each component can be derived directly from the fitted parameters as described in Equation 2.10. This gives the column density of each component, which can be summed together to give the total column density of the cloud. The CNM fraction is defined in Equation 3.2 as the CNM column density divided by the total column density (the CNM, UNM, and WNM combined). This fraction can be mapped across the field and I show this for all three clouds in Figure 3.5.

$$f_{CNM} = \frac{N_{(HI,CNM)}}{N_{(HI,CNM)} + N_{(HI,UNM)} + N_{(HI,WNM)}} \quad (3.2)$$

For the Alpha cloud, in Figure 3.5 it is evident that f_{CNM} decreases from south-west to north-east, with the exception of the cold clump at the north-east end of the cloud. At these velocities ($100\text{-}120 \text{ km s}^{-1}$) the main emission from the SMC lies to the south-west of this field, so the closer to the centre of the main emission in the same velocity range, the higher the f_{CNM} . Additionally, at this position, the emission of the main body of the SMC is seen at $v > 135 \text{ km s}^{-1}$. Looking at the velocity maps in Figure 3.2 the south-west region deviates less from the main body than the north-east. This means that the further from the main body emission, in velocity or physical space, the lower the f_{CNM} typically gets. The only area that deviates from this trend is the compact cold clump in the north-east. It is centred at a highly deviant velocity from the SMC and has a $f_{CNM} \approx 0.6$.

The Hook cloud has a stronger contribution from the WNM phase than either of the other two clouds. The maximum value for the CNM fraction is the lowest of all three clouds, at $f_{CNM,max} = 0.67$. So this may lead to the conclusion that the Hook cloud is warmer than the other two clouds. However, the velocity dispersion for the warmer component is almost half what it is for the other two clouds, placing it in the UNM range of the HI. So while the warmer gas dominates in this cloud, it is in the cooler, unstable phase. The CNM fraction also changes as we move around the cloud. The western side of the cloud is where the CNM fraction is highest, especially towards the

centre of the filament. This means, like the Alpha cloud, the CNM is shrouded by an envelope of UNM, but in a more filamentary geometry, a CNM core surrounded by UNM. The main body emission peaks at 150 km s^{-1} in this field and at the velocities of the Hook cloud ($130\text{-}142 \text{ km s}^{-1}$) the main body emission lies off to the south and south-west of this field. There is no strong relation between the velocity structure and the CNM fraction, as the CNM fraction seems to decrease almost to 0 in the northern most part and does not increase again along the eastern edge, whereas, the velocity structure follows a south to north trend, detailed further in Section 3.3. The CNM fraction appears to increase slightly with the thickness of the filament along the western edge.

The Gamma cloud has strong contributions from both phases of neutral hydrogen, with the majority of the CNM located at the southern end of the cloud and the WNM located at the northern end. There is a clear gradient in the CNM fraction as we move northwards, along the cloud. It goes from completely CNM to completely WNM from bottom to top shown in Figure 3.5. In the southern part of this cloud, the CNM fraction approaches 1, which is the highest recorded for any of the three clouds. While the other clouds indicate a scenario where the CNM is shrouded in a more diffuse WNM envelope, this cloud has no WNM envelope at its southern end indicating either all HI was condensed to CNM in this region, or the WNM envelope that previously shielded this cloud from the surrounding environment has been stripped away. So, similarly to the general trend seen in Alpha cloud, the Gamma cloud has a higher CNM fraction in the area that is closer to the main body of the SMC. At this clouds velocity ($100\text{-}112 \text{ km s}^{-1}$) the main body emission is south of the cloud and again at the clouds location, the main emission of the SMC begins at 125 km s^{-1} and peaks at 145 km s^{-1} . Additionally, as the CNM fraction decreases, the velocity increasingly deviates from the peak emission in this field.

3.3 Velocity structure

The velocity structure of the SMC is quite complex, thus to compare and contrast the velocities it is helpful to have a reference point for the bulk velocity of the emission across the SMC. A typical measure of the bulk velocity is the first moment (M_1) of an emission cube, which is the intensity weighted mean value of the velocity of the data, defined in Equation 3.3

$$M_1 = \frac{\sum vI(v)}{\sum I(v)} \quad (3.3)$$

To compare the velocities of the clouds to that of the whole SMC, I removed the emission of the clouds to avoid them being included in the first moment measure. I do this by masking any voxel at which the modelled emission is higher than the noise level at the same position, then calculate the first moment. The map of the first moment is shown in Figure 3.6 with the velocities ranging from $75\text{-}225 \text{ km s}^{-1}$. The velocity gradient across the SMC (due to rotation) is very evident from this figure. Because of this velocity gradient, defining a single value for the bulk velocity

of the entire SMC is unhelpful when trying to identify deviant structures. So now I have a reference from which I can utilise the bulk velocity to define a deviation velocity for each cloud.

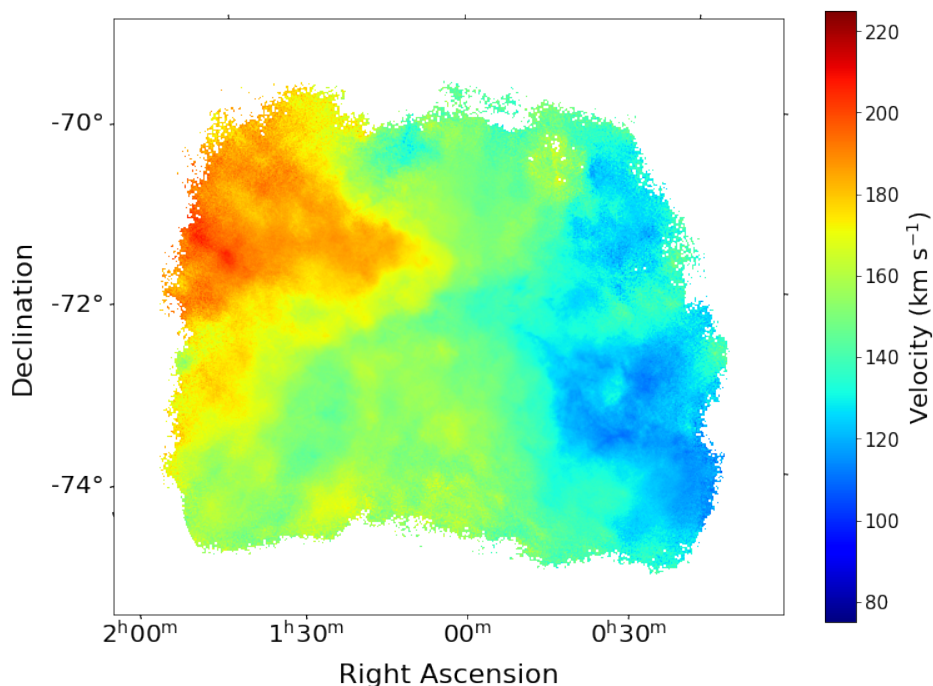


Figure 3.6: First moment map of the SMC, ranging from velocities of 75-225 km s^{-1} .

The mean M_1 velocity in this field for the Alpha cloud is 163 km s^{-1} , thus using the values reported in Table 3.1 the CNM is offset by 59 km s^{-1} and the WNM is offset by 55 km s^{-1} . In this cloud the different components have different velocity structure. The CNM component shows a $\sim -10 \text{ km s}^{-1}$ gradient from west to east whereas the WNM component has no strong gradient from west to east. Both components are centred around the same velocity on the western side, but moving across to the eastern side end up diverging from each other by 12 km s^{-1} . This divergence is driven primarily by the CNM velocity gradient. The CNM velocity decreases as we move to the side of the cloud furthest from the SMC main body at these same velocities.

Figure 3.7 shows the central velocities of each phase along two paths defined through the cloud. The paths traverse the east-west direction and the north to south direction through the cloud towards the densest clump in the south-west of the field. What is evident from this Figure is there is a non-zero difference that is significant above the 3σ level between the WNM and CNM as the path moves towards the densest clump. This offset is between $3\text{-}5 \text{ km s}^{-1}$ depending on the path that is considered, but along each path there is a gradual increase in this offset up to that level. This means that in the outer areas of this cloud the WNM and CNM are well-coupled dynamically, but there is something that occurred at the densest point of this cloud that caused the two phases to develop a dynamic mismatch.

The mean M_1 velocity in the Gamma cloud field is 146 km s^{-1} , thus using the values reported in Table 3.1 the CNM is offset by $33\text{-}39 \text{ km s}^{-1}$ and the WNM is offset by 44 km s^{-1} . For this cloud Figure 3.3 shows a clear velocity

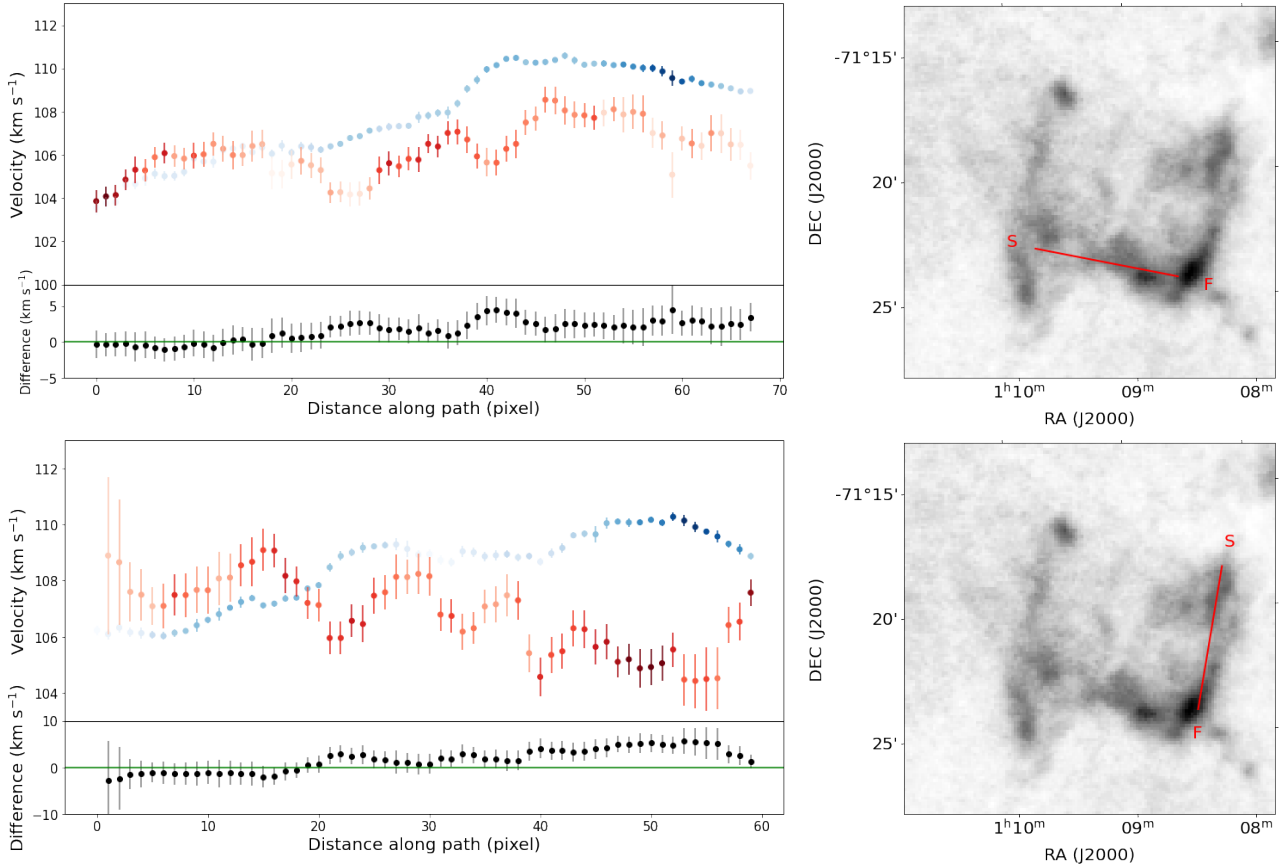


Figure 3.7: Left: Velocity for the different phases for two different paths through the Alpha cloud (top and bottom). The WNM is shown in red, the CNM in blue with the darker colours indicating higher column density. The error bars in the top panels show the 1σ uncertainties from Section 2.3. The error bars in the bottom panels show the same uncertainties at the 3σ level. Right: Path through the cloud for respective velocity relations. S denotes the start of the path, F denotes the end of the path.

gradient from 118 km s^{-1} at the bottom of the cloud to 95 km s^{-1} at the top of the cloud. It is clear that both the components follow each other, with the velocity decreasing as we move from the south end to the north end. It is important to note that the CNM velocities shown in Figure 3.3 are the mean of the two CNM components identified by the ROHSA fitting, which are offset from each other in central velocity by differing amounts across the field.

In Figure 3.8 I show the relative velocities of the mean CNM and the WNM components. It clearly shows the velocity gradient as we move south to north along the cloud length. At the northern end there is an offset that is just above the 3σ significance level at some points that indicates an offset between the CNM and WNM at a similar level seen in the Alpha cloud, $\sim 4 \text{ km s}^{-1}$. The uncertainties on these measurements are dominated by the uncertainty in the CNM components. This is due to the fact that there were two CNM components identified for this field that are separate from each other at some points, but overlap with each other in velocity space, in other parts of the field. This means when the uncertainty is measured, in some areas where the components overlap the amount of the CNM signal attributed to each component depended largely on the neighbouring fits. This confusion is clear in Figure 2.19 where for components 2 and 3 there is a relatively high uncertainty in the brightness temperature amplitude in the

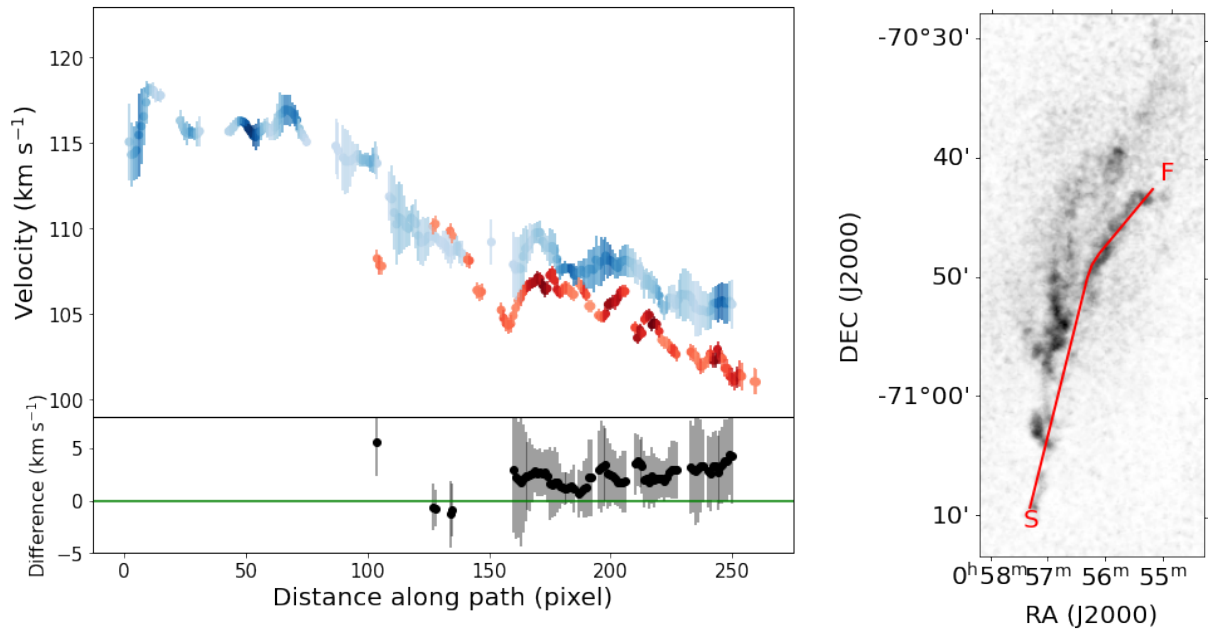


Figure 3.8: Left: Velocity for the different phases for a path through the Gamma cloud. The parameters in this Figure are the same as for Figure 3.7.

area that corresponds to the density enhancement around DEC -70:50:00. The uncertainty on the mean value shown in Figure 3.8 is smaller than the individual components. To see how each component traces the velocity structure, I show the central velocity of all three components in Figure 3.9 and the offsets of the CNM components to the WNM component. There is no significant offset until the density enhancement towards the end of the path, where the uncertainties are small enough to measure an offset of 3-5 km s⁻¹.

It is evident from Figure 3.9 that Component 2 has slightly higher velocity, albeit it within the uncertainties, than that of Component 3. It is also difficult to find an area of the cloud where the mean CNM and the WNM column density are both high enough to trust the fitted values for comparison. The total area where it is meaningful to compare the offset of the two components is smaller for this cloud than for that of the Alpha cloud.

As mentioned in Section 3.2, the CNM fraction decreases along this same path, so overall in this cloud, the warmer the HI, the more deviant the mean velocity of the cloud relative to the main body emission. However, partly due to the inability to compare the velocities of the two phases in most areas of this cloud, I do not observe the same velocity offset that increases as we approach the coldest part of the cloud, measured by the f_{CNM} . In fact, if anything, along this particular path there is an increasing offset between the Component 3 (CNM) and Component 1 (WNM) as the path moves to the warmer part of the cloud in the north.

The mean M_1 velocity in the Hook cloud field is 147 km s⁻¹, thus using the values reported in Table 3.1 the CNM is offset by 15 km s⁻¹ and the WNM is offset by 12 km s⁻¹. This cloud has an interesting velocity structure, it has a gradient that runs south to north with decreasing velocity along both sides. The CNM and WNM follow each other very well, only offset by 0.7 km s⁻¹ on average where the CNM is prominent, along the western edge. The offset

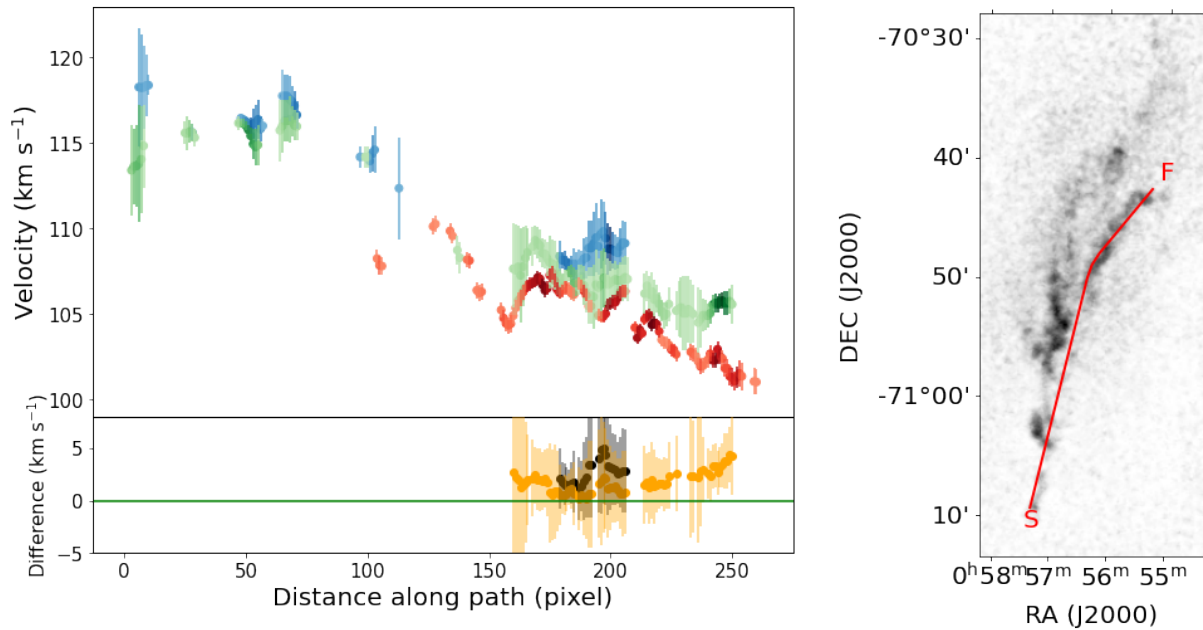


Figure 3.9: Top Left: Velocity for the different Gaussian components in the field. Component 1 (WNM) is shown in red, Component 2 (CNM) is shown in blue, and Component 3 (CNM) is shown in green with the darker colours indicating higher column density. Bottom Left: The offset between Component 2 and 1 is shown in black and the offset between Component 3 and 1 is shown in orange. Right: Path through the cloud, the same as shown in Figure 3.8.

becomes more pronounced at the low column density end of the path, seen in Figure 3.10. The velocity gradient of both phases evident in Figure 3.4 shows that the areas in which the cloud is closest to the main body emission (in the south-west direction) have the least divergent velocities from the main body.

This is a trend seen in all 3 clouds, a velocity decrease as we move away from the SMC centre in opposition to the general velocity trend we see across the SMC. The SMC emission moves east-west across the sky as the velocity decreases from $\sim 200 - 100 \text{ km s}^{-1}$.

The trend of offset between the CNM and WNM in the heads of clouds that is seen in the Alpha and Gamma clouds, is also seen in Brüns et al. (2001), at around the same level ($1-5 \text{ km s}^{-1}$), who looked at an extragalactic CHVC (HVC 125+41-207). They do not comment on what this means, however they do expect that this CHVC is undergoing ram-pressure stripping given its head-tail structure.

3.4 Comparison with CO measurements

The $^{12}\text{CO}(2\rightarrow 1)$ data for these clouds come from targeted APEX observations (Di Teodoro et al. 2019a) and were only available for 2 of the 3 clouds in this work. The observations were taken as a follow up to McClure-Griffiths et al. (2018) where the clouds were first identified in HI. They found spatial correlation, but not perfect overlap between the highest column densities of the existing HI data and their CO clumps. With the new HI data used in this work we

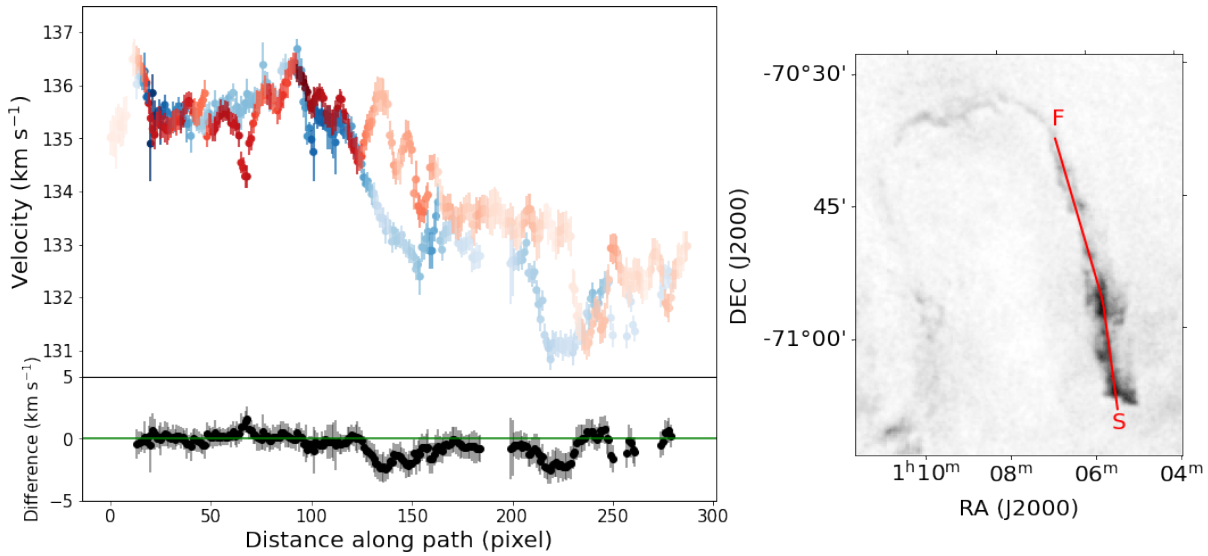


Figure 3.10: Left: Velocity for the different phases for a path through the Hook cloud. The parameters in this Figure are the same as for Figure 3.7

can make a more quantitative comparison of not just the bulk velocity but the velocity of the different phases.

3.4.1 Alpha cloud

This cloud has eight clumps in CO within the field observed. Comparison of the velocities recovered from the HI CNM decomposition solution and the CO data show that four of the clumps agree in velocity within a 3 channel ($\sim 3 \text{ km s}^{-1}$) window, whereas the other four all reside at velocities offset by 15 km s^{-1} from the HI structure, see Figure 3.11. The four clumps that correspond with the velocity structure modelled for the HI are more spatially coincident with the HI structure than the others. This divides these clumps into two distinct populations, one that likely belongs to this HI structure and one that does not. Interestingly, the clumps of CO that do not belong to the Alpha cloud do not correlate with any strong HI emission. There is little HI emission in this region at $v < 93 \text{ km s}^{-1}$ that is at 3σ above the noise. Additionally in the WNM the trend is similar, except on the eastern side, given the divergence of the CNM from the WNM in velocity space which was noted in Section 3.3. This demonstrates that the CO is more dynamically aligned with the CNM than the WNM. The preferable alignment of the CO with the CNM over the WNM is expected, as the CO indicates the areas of cooler molecular gas that form within cool HI regions.

3.4.2 Hook cloud

This cloud has nine clumps in CO within the field observed. The Hook cloud spans 3 times the size of the Alpha cloud in both physical directions, so the entirety of the Hook cloud is not observed in CO. The observations cover the area of strong total HI emission in the base of the western edge of the cloud, show in Figure 3.12. Seven out of nine clumps are in agreement with the CNM within 3 km s^{-1} . These same seven clumps are offset slightly more from the

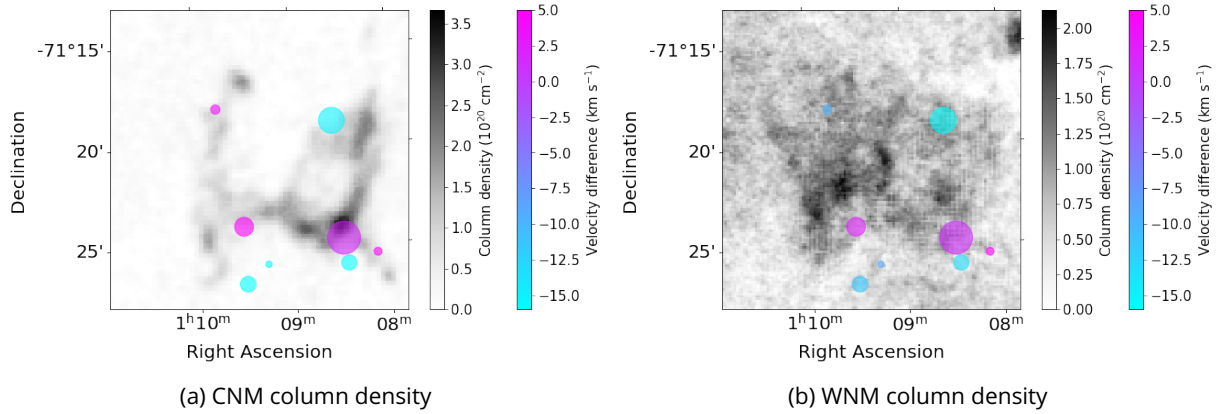


Figure 3.11: Left: Offset of the identified CO clump velocities with the fitted CNM velocities of the Alpha cloud from ROHSA (indicated by colourbar) shown at the position of each clump against the column density of the CNM (in greyscale, darker indicating higher column density). Right: Offset of the identified CO clump velocities with the fitted WNM velocity of the Alpha cloud from ROHSA (indicated by colourbar) shown at the position of each clump against the column density of the WNM. The size of each circle scales with the strength of the CO flux.

WNM than from the CNM velocities. The two clumps that disagree with this trend are slightly spatially offset to the CNM filament, however so is one clump that agrees well, off to the west of the filament. Those two anomalies are also among the weakest clumps in terms of integrated flux density. So if considering just the strongest CO clumps that lie on top of the density structure of the filament, there is not much difference in the agreement between the CNM and WNM contrary to what is seen in the Alpha cloud. This is mainly because while with the Alpha cloud the velocities of the two phases diverged at different points across the field, the velocities of the CNM and WNM in the Hook cloud are near the same. This is especially true at this dense part of the filament, evidenced by looking at the velocities at the beginning of the path through the filament defined in Figure 3.10

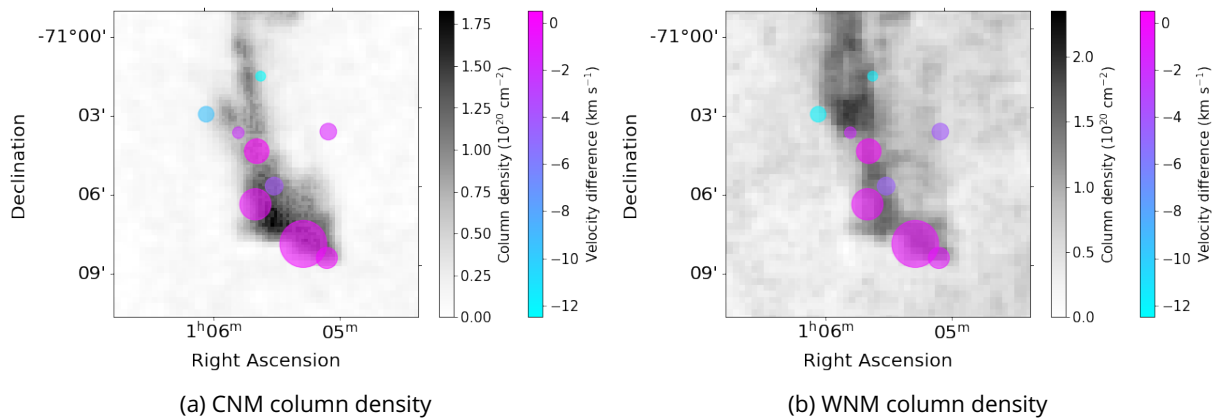


Figure 3.12: Same as Figure 3.11 but for the Hook cloud.

Overall in both clouds, the CO is dynamically aligned with the CNM in areas of high CNM column density. In the Alpha cloud the WNM does not trace the CO clumps as well as it does in the Hook cloud due to the different velocity relationships between the phases in the two clouds.

For comparison, the results from [Noon et al. \(2023\)](#), who looked at three HI clouds (C1, C2, and C3) around the Galactic Centre and their molecular fraction, found good spatial and velocity agreement between the CO clumps and the high HI column density areas. They do not look at the velocity of the CNM and WNM, as the data has insufficient velocity resolution (5.5 km s^{-1}) to resolve the CNM. They do look at the position-velocity relationship of the clouds and show that the CO sits nicely on top of the strong HI emission in one cloud and in another that some of the CO sits closer to the Galactic Centre than the HI. They find that by looking at the chemical states of the clouds, that the amount of molecular hydrogen (obtained by using an X_{CO} factor) is too high to be formed from the amount of HI present. They suggest that the clouds have been affected by a strong wind from the Galactic Centre that boiled off the HI envelope and dissociated the molecular hydrogen into HI.

Their clouds could be comparable to the clouds in this work. It is important to note that their clouds are smaller by a factor of a few in physical size even when projection is taken into account. Additionally, the Galactic Centre of the MW is a completely different environment to the edges of the SMC. However, the presence of multiple CO clumps which are dynamically entrained in the CNM for the Alpha cloud, particularly on the side nearest to the SMC, could suggest that this cloud is undergoing a similar wind force to the clouds in [Noon et al. \(2023\)](#). If the CO was formed in the Alpha cloud CNM structure as it is seen in [Figure 3.11](#) then the clumps would align with the regions of highest density, which is not the case for every clump. This suggests that some force has acted to move the HI away from the CO, or dissociate the HI from the CO.

There are no targeted CO targeted for the Gamma cloud, but it would be of interest to see if it has a similar CO distribution to C1 from [Noon et al. \(2023\)](#). This is because they find that C1 has CO protruding out from the HI cloud in the direction of the Galactic Centre. This unshielded CO, they posit, is in the evolutionary stage where the HI previously surrounding it has been boiled off and the molecular hydrogen has not started to dissociate yet. In the case of the Gamma cloud, the CNM protrudes out of the southern end of the cloud in the direction of the SMC, unshielded by the WNM. This may be the complementary scenario where the WNM envelope has been boiled off or pushed away by a force originating from the SMC main body.

3.5 HI mass

In [Table 3.2](#) I show the calculated masses for each cloud separated by phase. To do this, I sum all pixels that have a column density that exceeds the three times the mean uncertainty of the column density for each respective phase. To convert the summed column density to a total mass, I find the area of each pixel using the standard relation to obtain physical size from angular size, in [Equation 3.4](#) I use the value of $d = 63 \pm 5 \text{ kpc}$ as the distance to the SMC from [Di Teodoro et al. \(2019b\)](#).

$$D = d \tan \theta \quad (3.4)$$

So the conversion of the column density to HI mass (M_H) becomes:

$$M_H = \frac{1.67 \times 10^{-27}}{2.0 \times 10^{30}} (63 \tan(7^\circ) \cdot 3.086 \times 10^{21})^2 \sum N_H(x, y) M_\odot \quad (3.5)$$

$$M_H = 3.64 \times 10^{-20} \sum N_H(x, y) M_\odot \quad (3.6)$$

where this sum is only over pixels for which the column density satisfies the three times the mean uncertainty criterion.

I show these calculated values in Table 3.2. The values that contributed to the the uncertainties are the uncertainties in the column density from Section 2.3 and the uncertainty in the distance to the SMC, quoted above.

Cloud	M_{CNM} ($10^4 M_\odot$)	$e_{M,CNM}$ ($10^4 M_\odot$)	M_{WNM} ($10^4 M_\odot$)	$e_{M,WNM}$ ($10^4 M_\odot$)	M_{tot} ($10^4 M_\odot$)	$e_{M,tot}$ ($10^4 M_\odot$)
Alpha	1.2	0.19	0.21	0.03	1.40	0.19
Hook	1.34	0.21	3.86	0.61	5.21	0.65
Gamma	1.33	0.21	0.68	0.11	2.01	0.24

Table 3.2: The HI phase masses and total masses for each cloud. The uncertainties reported as specified in Section 3.5.

These values of total HI mass, are consistent with the value range reported for structures including these three clouds in McClure-Griffiths et al. (2018) and the mass reported for the Alpha cloud in Pingel et al. (2022).

3.6 Morphologies

The Alpha cloud has the most irregular morphology of the three clouds considered in this work. It has the majority of its cold material at the south-western end of the field, but also a connected dense clump on the north-east side.

This cloud was also analysed in Pingel et al. (2022). They utilised a fit to a mean spectrum obtained by shifting all spectra in the region to be centred at the velocity of the peak brightness temperature and obtained a two component solution with velocity dispersions of 2.93 and 5.52 km s⁻¹ for the CNM and WNM phases respectively. In this work I found a lower velocity dispersion for the CNM of 1.97 km s⁻¹ and a higher velocity dispersion for the WNM of 7.77 km s⁻¹, as shown in Table 3.1. The mismatch between these two methods is likely due to the fact that they took the mean spectra and then fit Gaussians, whereas I fit Gaussians and then extracted mean values. Shifting each spectra to the velocity at which the spectrum peaks to fit the dispersions of the two phases assumes that the CNM traces the WNM at every point, or at least follows it by the same offset everywhere. This is because at the central velocity of the CNM in the spectrum is where the brightness temperature would be highest if there is a comparable amount of CNM to WNM, as the WNM components are shallow and wide. I have shown in this work that the CNM and WNM do not follow each other in a fixed way in the Alpha cloud, so the fitted dispersions in Pingel et al. (2022) do not give

as much physical information about the temperature of the phases across the cloud. In their analysis, they also look at the intensity-weighted velocity and velocity dispersion maps of the cloud, shown in Figure 3.13.

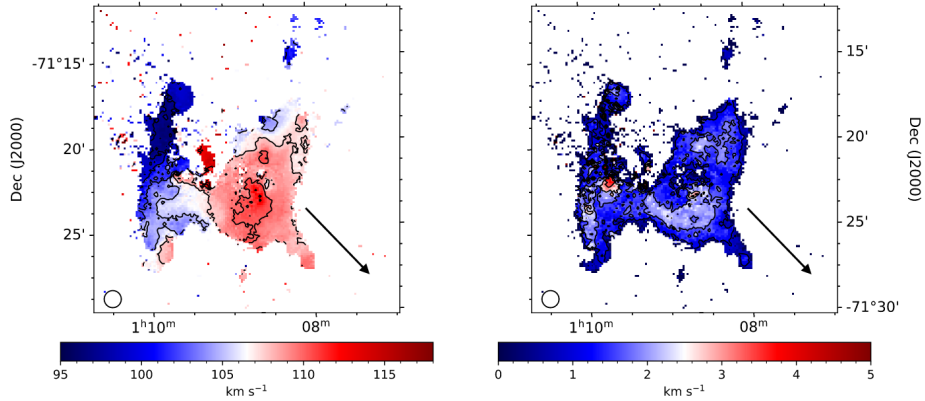


Figure 3.13: Figure 17 from Pingel et al. (2022) that shows the intensity weighted velocity (left) and the intensity weighted velocity dispersion (right) of the Alpha cloud.

Figure 3.13 (b) resembles the CNM velocity dispersion map produced in this work for this cloud in Figure 3.2. There is a similar increase in values just above 2 km s^{-1} in the southern part of the cloud and a decrease of values in the region around the north-east clump. Additionally, the velocity map looks very similar to that of the CNM component in Figure 3.2. The same velocity gradient is found decreasing from the south-west to the north-east.

There is also an intriguing cavity in the western part of the cloud. A shell was previously detected in this field (see shell 369 in Staveley-Smith et al. (1997)) at a radial velocity of 117.7 km s^{-1} . In Staveley-Smith et al. (1997) they have a velocity resolution of 1.6 km s^{-1} . The maximum central velocity measured for the CNM in this field is 113.4 km s^{-1} around this central cavity on the western of the field, which is consistent with the velocity of the shell. In Figure 3.14 I show the position and radius of the shell measured from Staveley-Smith et al. (1997) and the shell fit by Pingel et al. (2022). It is important to note that the resolution of the data in Staveley-Smith et al. (1997) is $1.6'$, so the spatial offset of the shell centre to the cavity in the CNM could be attributed to that inherent positional uncertainty.

The presence of a shell to explain this cavity morphology is consistent with the data, but it does not explain why there is an offset between the CNM and WNM at the south-west point of the shell. There has to be some force coming from the direction of the SMC that has acted upon the gas. It would easily strip the WNM before the CNM since the WNM is less dense, leading to an offset in velocity and an increase in the CNM fraction in the area, both of which I find in this analysis.

Something else to consider is the $H\alpha$ data from the MCELS project, that mapped the whole SMC (Winkler et al. 2015), including most of the fields from these clouds. When looking at the region around the Alpha cloud there is strong $H\alpha$ emission in the base of the cloud (shown in Figure 3.15) which could be a source of energy or ionising photons acting on the HI, creating density enhancements either side of it. There is no record of this emission in the most recent catalogue of HII regions of the SMC (Pellegrini et al. 2012). In Figure 3.7 the top panel shows the path

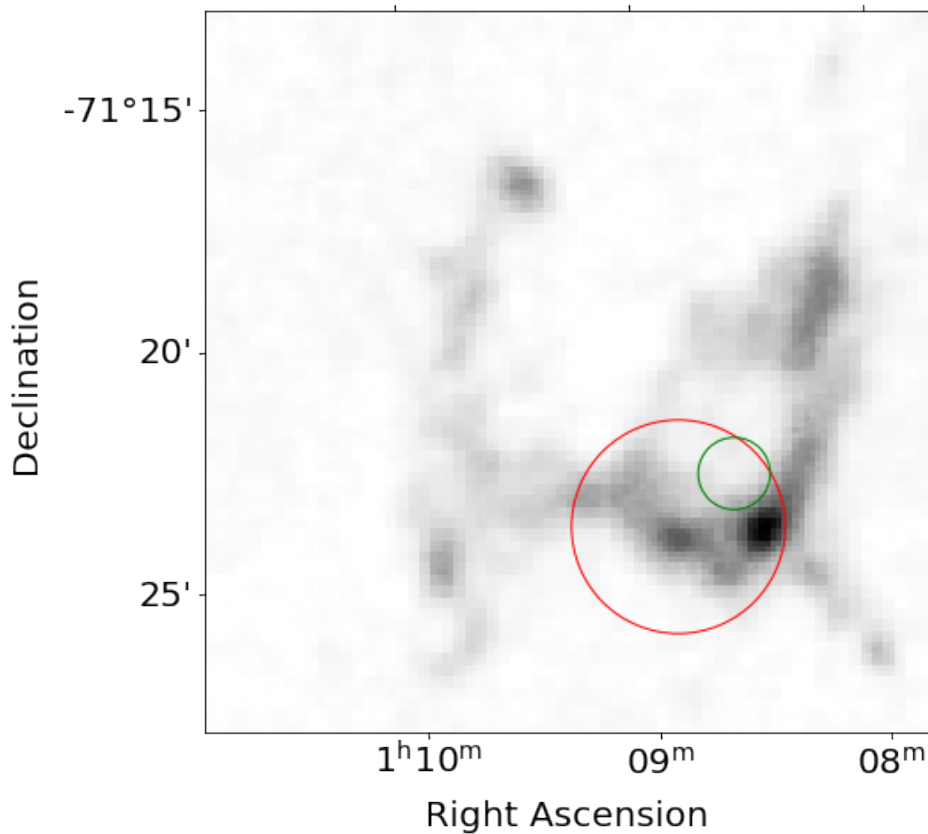


Figure 3.14: CNM column density of the Alpha cloud from ROHSA with the shell identified in Staveley-Smith et al. (1997) shown in red, and the shell from Pingel et al. (2022) shown in green.

along the direction intercepting this region of the cloud. It is around where this $H\alpha$ occurs, that the offset between the CNM and WNM begins. However it may be more reasonable to expect if this $H\alpha$ emission is acting upon the HI to accelerate it away from that point, the offset between the CNM and WNM would peak at that point and fall off either side of it. This is not the trend seen in Figure 3.7. While the spatial correlation of the $H\alpha$ with the HI is compelling, the MCELS data is integrated, so there is no information on the radial velocity of this emission. Investigation of the data used in Smart et al. (2019) taken with WHAM, where there is spectroscopic information for these fields, but at a much worse resolution of 1° , did not yield any results. Since this emission is very compact, $\sim 1'$ wide, any emission from this source is smeared by the beam of WHAM. So this emission cannot be dynamically matched with the Alpha cloud at this time.

Overall, the data paints a picture of a cloud of HI that has undergone more than one process. The expansion of a shell has created a cavity in the western side, ram pressure from the SMC has stripped the head of the cloud of a significant amount of WNM, leading to a higher CNM fraction and offset between the two phases at this end. Additionally, a region of ionised hydrogen has potentially exerted pressure on the areas around it increased the column density either side.

The Hook cloud does not show a large velocity gradient towards the SMC like the other two clouds do and the

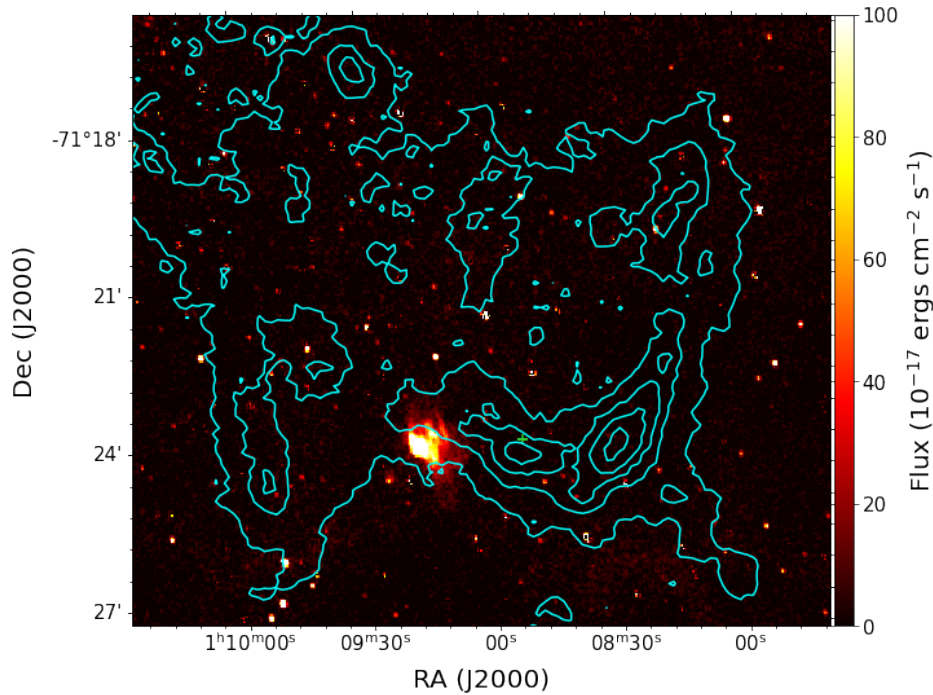


Figure 3.15: $H\alpha$ emission for the field of the Alpha cloud from the MCELS data (Winkler et al. 2015). The total HI column density from ROHSA is shown with the cyan contours.

CNM and WNM components trace each other very well. It is more likely, given the geometry of the feature, that it is formed by an expansive force from the centre of the field. It has a very straight filamentary edge on the west side of the field with a warmer section that loops around and down again on the eastern side, with a drop in column density in between these edges. A possible formation method for straight, long filaments like this is shown in Ntormousi et al. (2011) where the colliding winds of two superbubbles meet to compress material and form a cold filament. However, in their simulations the filament is not as compact as the Hook cloud. In Martínez-Delgado et al. (2019) they identify an arm-like structure of stellar objects in the SMC outskirts that extends into this field with the Hook cloud. Multiple stellar objects in this field could provide the energy to expand a shell out to this size of 100s of pcs.

The Hook cloud is not a perfect spherical shell, this could be due to expansion occurring at multiple points within the structure, leading to uneven expansion in the different axes. Additionally, a southern edge of the shell does not appear to exist. The southern end of the Hook field is closest to the mSMC main body, so any diffuse material at this end could have been swept away by winds from the SMC before the western CNM edge started to form.

The Gamma cloud has a head-tail structure that is typical of material that is undergoing a wind-like force. This head-tail structure is observed in many HVCs around the MW (Brüns et al. 2000, Westmeier et al. 2005, Putman et al. 2011). The head-tail morphology is consistent with a cloud passing through a more diffuse medium, such as a halo, meaning this cloud could be experiencing infall. This infall process would push the warm envelope away, draping it around the head of the cloud leaving the CNM relatively unaffected by the process (Konz et al. 2002). In simulations (Quilis and Moore 2001, Heitsch et al. 2022) they find that the head of cloud has smaller linewidths, thus

cooler temperatures than the tails of the cloud, where most of the material is in the warmer phase. This is consistent with what I find in this cloud. If this infall scenario is true, that has implications for the orientation of the cloud. There is a velocity gradient that shows the velocity becomes less blue-shifted as we move along the cloud towards the SMC. This would imply that cloud is in front of the SMC relative to the observer, with the cold core travelling at a velocity closer to that of the SMC as it accelerates towards it, the tail trailing behind.

It can be hard to distinguish between the scenarios of infall and outflow, because the same velocity gradient would be observed if some force from the SMC main body is acting on this cloud, blue-shifting this cloud towards the observer. This would be a scenario where a ram pressure force acted on a cloud that was already present in the SMC halo, creating this head-tail structure that we see in wind-tunnel simulations. The scenario of ram-pressure stripping is more reasonable in the context of this cloud as it would explain why there is no WNM at the southern end. It is possible we are catching the cloud in a state where the pressure has stripped away the WNM and left the unshielded cold core, that has not had a chance to dissociate yet, similar to the scenario proposed in [Noon et al. \(2023\)](#) for a smaller MW cloud.

As with the other clouds, I compared the HI with the data from the MCELS survey in $H\alpha$ and find an area of compact emission just below the head of this cloud. Again, there are no records of this emission in the most recent catalogue of HII regions of the SMC ([Pellegrini et al. 2012](#)). There is a similar problem that arises with this cloud as with the Alpha cloud, that while the spatial coincidence is compelling, there is no radial velocity information in the MCELS data and the WHAM survey in [Smart et al. \(2019\)](#) does not have the spatial resolution to make a meaningful measurement of this compact emission. An extra piece of information that is available for this field is the presence of a catalogued stellar cluster (HW32, [Maia et al. \(2013\)](#)) which lies inside this $H\alpha$ emission. I show in Figure [3.16](#) the position of this stellar cluster overlaid on the $H\alpha$ image with the HI contours showing the extent of the cloud. There is no information about the distance or radial velocity of this stellar association in the most recent Gaia data release ([Vallenari et al. 2023](#)), so the association is still only spatial, but this provides a potential ionisation source for the hydrogen in this area. Energy ejected from this region, could provide the ram-pressure force that stripped the base of the Gamma cloud. In fact, it is possible that the cold gas that is now being pushed outward, provided the nursery for this stellar cluster.

There are also measures of the line of sight magnetic field from [Livingston et al. \(2021\)](#), which indicate a change from negative line of sight magnetic field to a positive one in the area towards all three clouds looked at in this work. For example at RA: 00:57:53.8 and DEC: -71:18:35.3, which lies a few arcminutes south of the field shown in Figure [3.16](#) a magnetic field strength of $+0.3 \mu G$ was measured. They suggest that these magnetic fields are associated with the outflows and could be due to out-flowing forces from the SMC stretching the magnetic field lines. The results of [Jung et al. \(2023\)](#) show that magnetic fields can elongate an HVC along its streaming direction, reducing the amount of fragmentation. So if this change in magnetic field is due to the Gamma cloud, it could be helping the cloud remain spatially quite linear and particularly filamentary in the CNM. However, the size scales in [Jung et al. \(2023\)](#) are a few

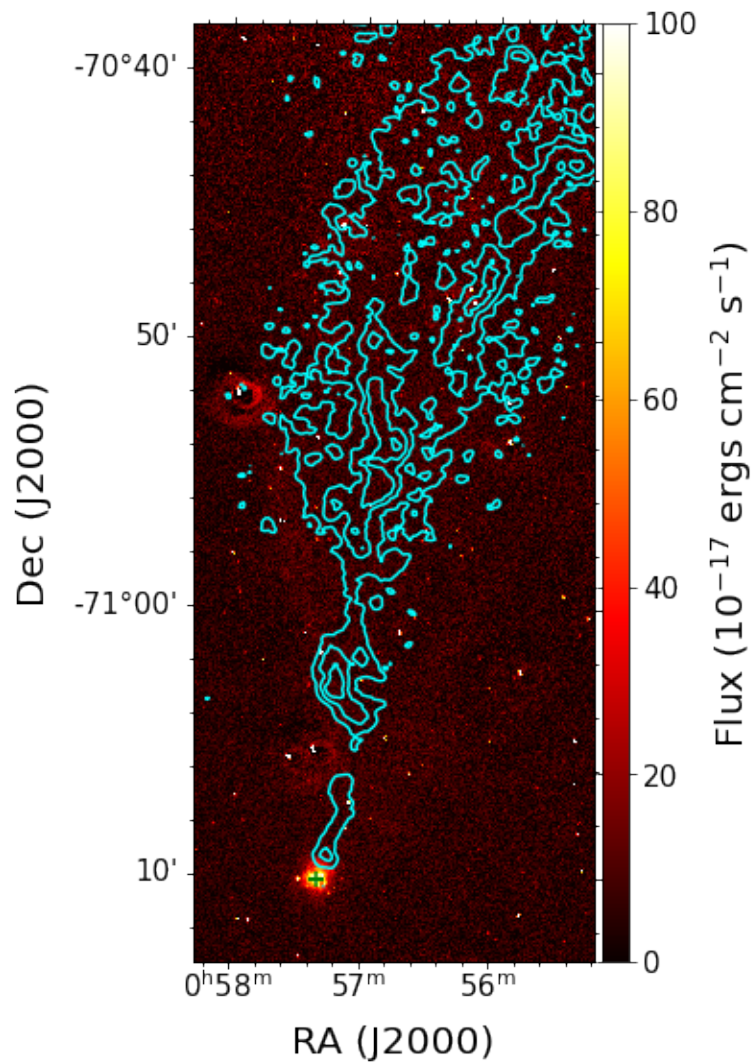


Figure 3.16: $H\alpha$ emission for the field of the Alpha cloud from the MCELS data (Winkler et al. 2015). The total HI column density from ROHSA is shown with the cyan contours. The green marker shows the position of HW 32, a young stellar cluster of the SMC.

times larger than the Gamma cloud appears in projection and they do not consider the separate phases of the ISM, thus it is difficult to make a direct comparison.

If this material has been accelerated by either the stellar cluster identified here, or by star forming regions further inside the SMC main body, such as N66, this could distort the magnetic field to such an extent that the line of sight strength changes from negative to positive in this region.

Chapter 4

HVC/IVC catalogue of the SMC periphery

This chapter focuses on the process used to uncover further anomalous clouds around the SMC. After the analysis of the large structures in the SMC, I decided to search for the presence of smaller structures. It is shown in [McClure-Griffiths et al. \(2018\)](#) that there are smaller clumps at the extreme velocities of the SMC that came up in their analysis. There is a large structure to the north west of the SMC that contains a lot of smaller clumps of high column density gas, as well as small clumps of gas offset by $35\text{-}65 \text{ km s}^{-1}$ from the main body emission. There are obviously many of these clouds just seen by visual examination in the data used in [McClure-Griffiths et al. \(2018\)](#) which was obtained with the reduced ASKAP array of 16 dishes, so in the data used in this work we may expect to find more. To approach this search I knew it would not be feasible to inspect the entire cube by eye. The cube has 220 velocity channels and each channel is 1000s by 1000s of pixels wide. Additionally, by eye analysis can be subjective, so I decided to develop a method to find the small scale structure that lies away from the main body of the SMC to eliminate some of this subjectivity.

4.1 Identifying further anomalous clouds

When looking for smaller anomalous clouds, ideally these would be compact clouds with high brightness temperatures. In this case anomalous would mean separate from the main body emission either in the spatial or velocity axes.

I tackle the main point first, finding compact clumps of gas. A way to probe this is to measure the spatial dispersion in a given velocity channel. The dispersion will give us a quantification of the change of brightness temperature within a given area of the image (defined by some width X and some height Y). For the SMC data I used $X = Y = 10, 20,$ and 30 pixels to cover different physical scales. I also calculate the mean of each box in Equation [4.1](#):

$$\overline{T_B}(v) = \frac{\sum^X \sum^Y T_B(x, y, v)}{X \times Y}. \quad (4.1)$$

So I then define the dispersion σ in Equation [4.2](#):

$$\sigma_{XY}(v) = \sqrt{\frac{\sum^X \sum^Y (T_B(x, y, v) - \overline{T_B}(v))^2}{X \times Y}}. \quad (4.2)$$

A clump that is contained within that area will give a higher dispersion measure than an area that contains a consistent brightness temperature level throughout, whether that be an area of strong emission or area with no emission signal at all. Unfortunately, it is not as straightforward as just declaring that areas with non-zero dispersion are clumps. There will be a base level of dispersion, even in the truly uniform areas of the sky due to noise in the image. This is especially true of the edges of the cube where the noise level is very high. So to avoid flagging these areas as clumps, I have to introduce a noise measurement to discriminate between high noise areas and clumps. I define the noise as:

$$rn_{XY}(v) = \sqrt{\frac{\sum^X \sum^Y \sum^V (T_B(x, y, v) - \overline{T_{B,XY}})^2}{X \times Y \times V}} \quad (4.3)$$

where

$$\overline{T_{B,XY}} = \frac{\sum^X \sum^Y \sum^V T_B(x, y, v)}{X \times Y \times V} \quad (4.4)$$

where V represents the range of channels where $v < 62 \text{ km s}^{-1}$ and $v > 235 \text{ km s}^{-1}$ as these are channels with no emission.

Now I can use the ratio between the two as a proxy signal-to-noise measure to determine if the dispersion is significant enough to classify as a clump. I choose a significance level of 3 in this work. This will avoid all the edges of the datacube from being highlighted in this process. In Figure [4.1](#) it is illustrated how much more obvious the small scale structure becomes in an image with a linear colour scale when looking at this ratio as opposed to the velocity channel map.

Generally I would expect a high dispersion measure if there is a large positive increase in the brightness temperature with respect to the background level within a area. However, it is also possible to get a high dispersion measure if there is a large decrease in brightness temperature relative to the background level. This could be seen if the background level is non-zero due to diffuse emission and there is some self-absorption from an intercepting CNM cloud along the line of sight. This behaviour could also occur if the background level is close to zero and there is an image artefact that produces a large negative signal. However, this is quite unlikely to occur since the data used in this work has been rigorously validated through the processes described in [Pingel et al. \(2022\)](#).

HI self-absorption is a physical process that is important to identify, whereas the image artefacts are non-physical

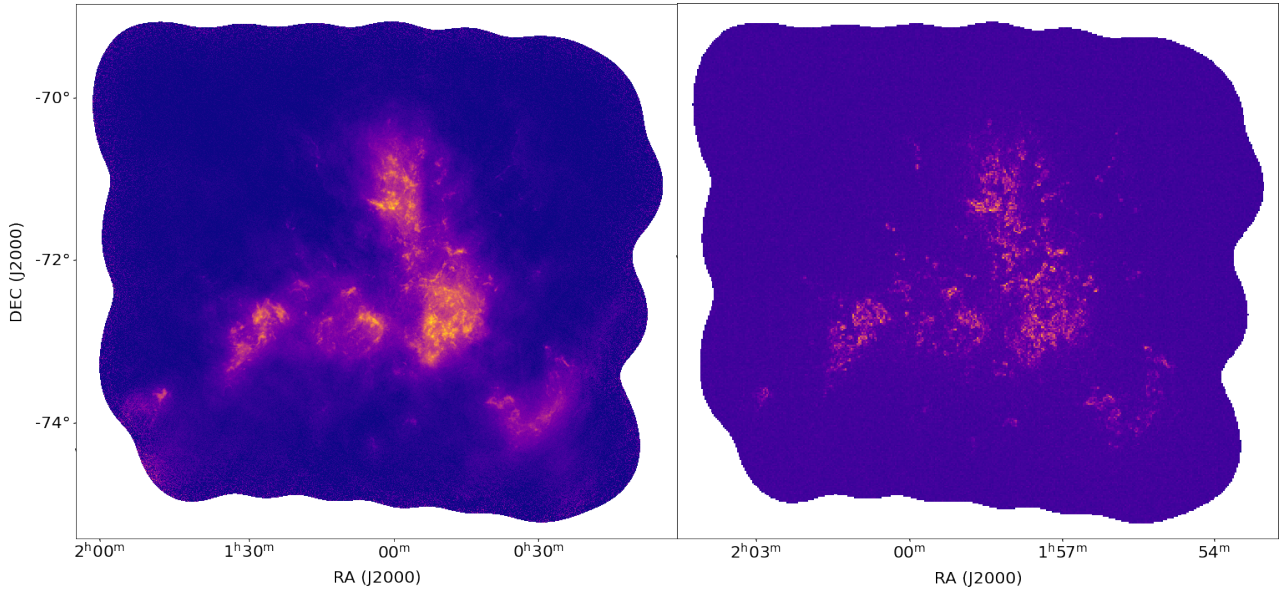


Figure 4.1: Left: Brightness Temperature map of a single channel of the SMC, at 144.5 km s^{-1} , colour scale ranges from 0 to 105 K. Right: Ratio, $\sigma_{XY}(v)/rn_{XY}$ of the same channel, for dispersion measured at the 10 pixel scale, colour scale ranges from 0 to 10.

and should be ignored in this analysis. So to distinguish between the two scenarios, as it is not possible to do so from just the dispersion measure, the mean brightness temperature should be above 0. This will not exclude self-absorption, as physically the self absorption saturates at 0 K, so the condition $\overline{T_B}(v) > 0$ should always be satisfied.

So now I have two conditions to define a clump, $\sigma_{XY}(v)/rn_{XY} > 3$ and $\overline{T_B}(v) > 0$. As I have only defined these clumps in relation to the two spatial axes of the datacube, how do I incorporate the 3rd velocity axis? Both of these measures are a function of the velocity, creating new datacubes with the same velocity axis as the original data, but where each pixel in the spatial axes represents $X \times Y$ pixels from the original data. If I define the size of the original datacube to be (X_0, Y_0, Z_0) , the new datacubes have a size of $(X_0/X, Y_0/Y, Z_0)$. In each velocity channel image I now have a map of potential clumps, but as the data has a spectral resolution of just under 1 km s^{-1} it is very unlikely for the signal from the clump to only be seen in one velocity channel unless the entire clump of gas is completely CNM and very cold. So more often than not the same clump will be seen in multiple adjacent channels, but it would be unhelpful to make any assumptions about the number of adjacent channels a clump occupies. Thus, the approach I take is to use a peak finding algorithm to identify peaks in the $\sigma_{XY}(v)/rn_{XY}$ spectrum. I use the `find_peaks` function from the python `scipy.signal` module. This algorithm finds peaks by going through each channel in a spectrum and looking at the adjacent value on each side. If they are both smaller than the value at that channel, then there is a peak there. However, noise fluctuations could obviously trick this type of algorithm into finding a false peak, so it allows the user to define a threshold above which a peak can be defined. So as per my previous definition, I set this threshold to be $\sigma_{XY}(v)/rn_{XY}$ of 3. I also define another argument the function takes, the distance between peaks in units of the channel number, to be 5. This reduces the likelihood of multiple peaks being identified in neighbouring

channels due to noise fluctuations.

Now I have provided the `find_peaks` function with the spatial dispersion cube that has been masked to exclude areas where $\overline{T_{B,XY}}(v) < 0$. The output of this function is a collection of points with x_{pf} , y_{pf} and v_{pf} coordinates showing peaks identified in the image. In Figure 4.1 I show the brightness temperature of a single velocity channel and the spatial dispersion of the same channel to highlight how much easier it is to see areas of small scale emission. In the Figure it is also evident that there is a lot of small scale structure within the SMC main body as well. This means that the method outlined so far can not isolate only the small scale structure in the periphery of the SMC, it finds everything. So there needs to be some way to determine which of these structures are isolated and which are not.

First, it is necessary to define what 'isolated' is. For comparability with the previous structures analysed, I define it to mean a clump that is separated by at least 250 pc and/or 20 km s^{-1} from the main body. A separation of 250pc was chosen as it is towards the larger end of the HI shell radii distribution, determined in [Staveley-Smith et al. \(1997\)](#) and thus gives an indication of the distance over which HI can be launched within the SMC. In the periphery of the SMC where the gas density is lower, it could be expected that ejected material will travel further than this distance, thus defining the threshold in this work. A separation of 20 km s^{-1} represents the typical separation seen in the large clouds in Chapter 3 between the cloud velocity and the beginning of the main body emission.

For a structure to be isolated at these scales means that either there are few to no identified clumps between an isolated clump and the large population of clumps that belong to the main body within these ranges. It is very possible that truly isolated clumps have more than one identified point if they are more extended than the scale used in the dispersion measure. Thus, there may be clusters of points that represent an isolated clump. For this reason, I use a clumping algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise) from the `sklearn.cluster` python package. This algorithm clusters together points in a distribution based on a user-defined maximum separation and a minimum number of points to define a group. It returns all the points in the distribution with a label that corresponds to their group. Any points that cannot be clustered within a group due to the maximum separation definition, are returned as outliers. Incorporating the aforementioned definition of isolation in all three axes with this algorithm will allow for the identification of outliers in this distribution of small scale clumps.

So, now that I have outlined the process, I will define the exact physical scales that we are working with. I defined the maximum separation as 250pc, but the data is in units of RA (hours:mins:secs) and DEC (degrees:mins:secs) along the x and y spatial axes respectively. As in Section 3.5 I use the simple conversion of angular units to physical units in Equation 3.4. As I am not looking to be extremely precise in this defining this threshold, I do not consider the effect on the declination on measurement of the true angular width in RA. So I take the width of each pixel, which is 7" ([Pingel et al. \(2022\)](#)) and determined the physical scale this represents at the distance of the SMC, which I defined in Section 3.5 as 63 kpc. Also, the uncertainty on the distance to the SMC is larger than the error introduced by not considering the effect of declination. Using Equation 3.4 the physical width of each pixel is then 2.14pc. So then a physical separation of 250pc would correspond to a separation of 117 pixels in both physical axes of the original

datacube. This can then be passed to the clustering algorithm to define the maximum separation.

A small practical problem arose with defining the maximum separation parameter (eps) in the DBSCAN algorithm, that it only takes a single float value. Since the axes have different units, I cannot use the x , y and v coordinates identified by the peak finding algorithm directly with the DBSCAN algorithm. So instead, I normalise all x , y and v values by the isolation criteria for their respective axes so that I can define the isolation radius as eps=1. This looks like

$$x_{DB} = \frac{x_{pf}}{117/X} \quad (4.5)$$

$$y_{DB} = \frac{y_{pf}}{117/Y} \quad (4.6)$$

$$v_{DB} = \frac{v_{pf}}{20/\Delta v} \quad (4.7)$$

where the subscript pf denotes the 3D coordinates obtained in the peak finding step of the process and the subscript DB denotes the 3D coordinates which will be supplied to the DBSCAN algorithm. Δv is the velocity channel width in km s^{-1} . For this data, $\Delta v = 0.98 \text{ km s}^{-1}$.

Now that the coordinates of the peaks are converted, I pass it to the DBSCAN algorithm with maximum separation parameter, eps=1, and a value of 15 for the minimum number of points that define a clump. The result of this clustering is a large cluster of points that correspond to the main body of the SMC and a collection of around 100 points that have been identified as outliers to that large cluster.

This is where I get to the final stage of the identification of isolated clumps. One characteristic of the clouds analysed in the previous section was that they were not connected to the main body by diffuse emission. Since this clustering algorithm only has information about where the small scale structure is, it is impossible for it to distinguish between small scale density enhancements in a large diffuse structure from the main body as opposed to a truly isolated clump of gas. So finally, since 100 is not a prohibitively large sample of clump candidates, I conduct a by-eye inspection of the brightness temperature at the identified peak velocity (v_{pf}) and the spectrum at the point of peak brightness temperature within the $X \times Y$ box the dispersion was measured from. Both of these verification checks are done on the original data. I am looking for the absence of a diffuse envelope connecting to extended strong emission in the brightness temperature map and a obvious separation of the emission signal from the broader main body signal in the spectrum. In Figure 4.2 I show an example of an accepted clump with the absence of spatially diffuse emission around it and rejected clump with significant levels of spatially diffuse emission spanning multiple arcminutes.

It is also important to note that the structures that were analysed in Chapter 3 were also identified while using this method. I rejected these identifications not because they did not meet the criteria I set out, but because they

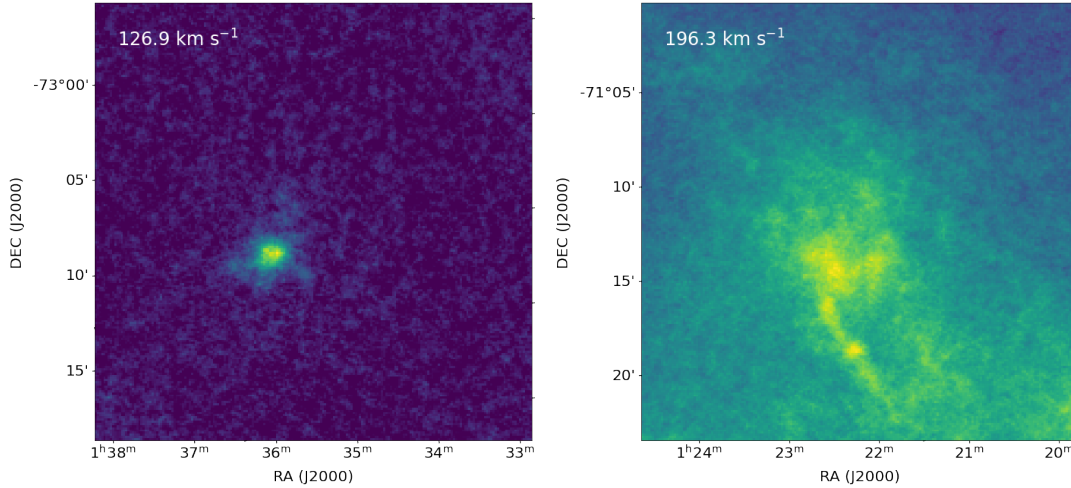


Figure 4.2: Left: A clump candidate that was accepted using by-eye verification. Right: A clump candidate that was rejected using by-eye verification. Colourscale represents brightness temperature at the channel velocity specified in the top left corner of each panel.

were already known.

After the by-eye verification, the sample of clumps was reduced to 31, which are detailed in Table 4.1

4.1.1 SMC

As outlined in Section 4.1, there are 31 clumps around the SMC that were identified through the clustering method. They are distributed as shown in Figure 4.3. The first feature that becomes clear in Figure 4.3 is the cluster of points to the north-west of the SMC. Some of these points can be attributed to a larger structure that was identified in McClure-Griffiths et al. (2018) in the 105-113 km s⁻¹ and 144-152 km s⁻¹ ranges. These correspond to the orange and cyan points in Figure 4.3. The green points also present correspond to structures around 130 km s⁻¹ that do not seem to be connected to the previously identified ones. However, there does seem to be a gradient in the northern-most part of the north-west region, from 105-145 km s⁻¹ north to south, indicating that maybe these clumps represent condensation of HI into the CNM phase along a more diffuse WNM structure. However, there is no way to verify that these structure are physically related. They obviously lie close to each other in the 2D projection we have access to in this data, but since the distance along the line of sight is unknown, there is this inherent uncertainty. The velocity gradient does suggest a relationship between the points that span a range of 30 km s⁻¹ over around 1 degree of declination, but does not confirm it.

Aside from the population in the north-west there are also groupings that can be inferred elsewhere, for example in the north-east, east, and the south. The clumps in the north-east and east are in the direction of the Magellanic Stream and Magellanic Bridge respectfully. The Stream and Bridge are both products of the tidal interaction between the SMC and LMC, a type of interaction that can compress and increase the turbulence of the gas (Renaud et al. 2014). It could be the case that towards these large HI filaments there are clouds that have condensed in their wake. The

Clump number	RA centre (J2000)	DEC centre (J2000)	Peak velocity (km s ⁻¹)
1	00:25:29	-73:08:16	124.9
2	00:35:23	-72:31:58	145.5
3	00:36:47	-71:44:39	142.5
4	00:36:50	-72:00:01	109.3
5	00:39:10	-70:43:27	120.1
6	00:39:16	-71:36:56	145.5
7	00:39:54	-70:51:49	138.6
8	00:40:11	-71:48:27	136.7
9	00:40:29	-70:34:29	132.8
10	00:41:34	-70:37:13	129.8
11	00:41:56	-70:31:09	108.3
12	00:42:36	-71:51:08	142.5
13	00:43:00	-71:33:29	159.2
14	00:44:59	-71:41:25	118.1
15	00:45:10	-70:10:53	109.3
16	00:45:32	-71:50:00	155.2
17	00:47:10	-71:25:09	163.1
18	00:47:55	-71:41:19	180.6
19	00:53:11	-74:26:09	157.2
20	00:56:35	-74:28:21	141.6
21	00:59:57	-75:01:32	153.0
22	01:02:38	-71:41:02	122.0
23	01:03:26	-74:46:32	142.5
24	01:11:14	-71:15:54	143.5
25	01:11:41	-72:07:48	91.7
26	01:15:59	-71:56:28	135.7
27	01:28:51	-71:57:58	126.9
28	01:31:00	-72:34:20	135.7
29	01:31:45	-72:30:28	144.5
30	01:32:40	-72:42:29	128.9
31	01:35:45	-73:10:33	126.9

Table 4.1: Field centres and peak velocities of 31 identified clumps from the SMC

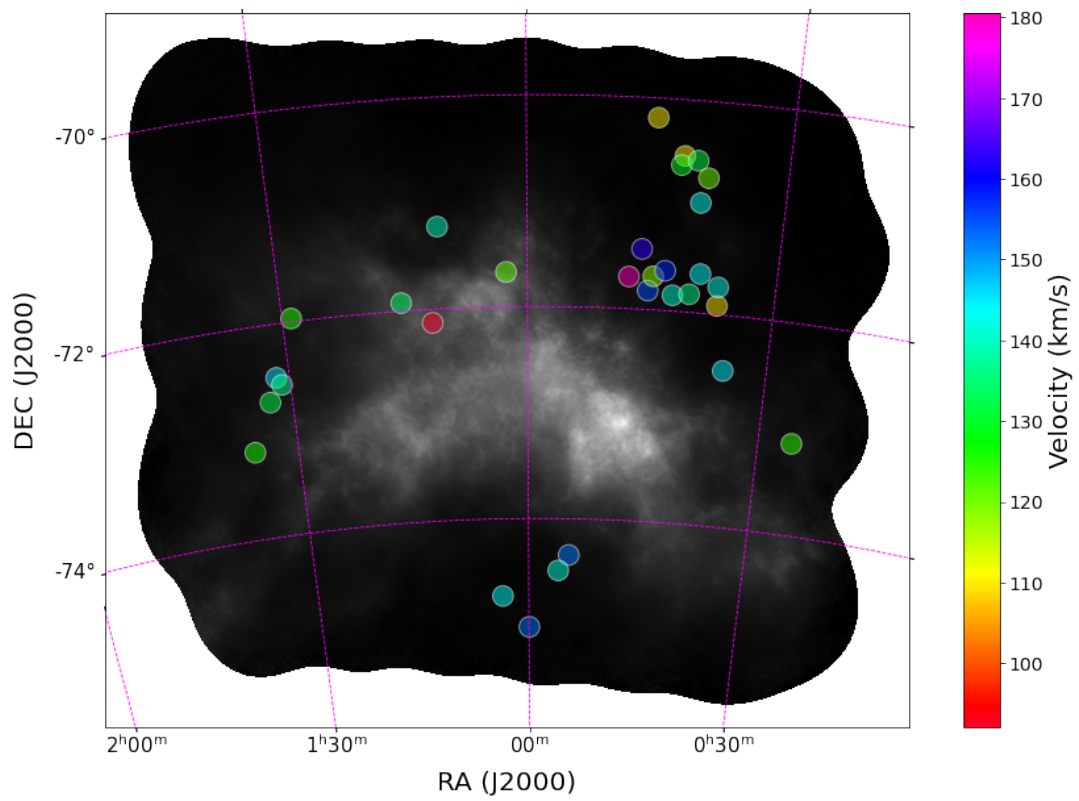


Figure 4.3: The distribution of the clumps around the SMC with the colourbar representing the velocity of the peak brightness temperature of the clump. The total column density of the SMC is shown in greyscale, with white indicating maximum column density.

Bridge and the Stream are both connected to the SMC in all 3 axes, both spatial and velocity. The fact that these clumps are offset from the main emission on the SMC means that while they trend towards the general direction of the Stream and Bridge they have either formed in very diffuse gas trailing those obvious structures or have been moved away from them by some force.

The other population are the four clumps in the south, where there is not any significant tidal structure. They all peak within 10 km s^{-1} of each other, meaning that it is possible that they originated from the same diffuse parent cloud. The one region that could be affecting this area the most is the southern part of the SMC Bar, where there are a number of HII regions that could be responsible for the acceleration of material out of the SMC main body during star formation events.

4.1.2 Other Magellanic fields

Once I had analysed the SMC field, more fields of data had become available from the GASKAP survey. There were datacubes available for a field of the Magellanic Stream adjacent to the SMC, the Magellanic Bridge and a field of the LMC that included 30 Doradus. This led me to ponder if we see the same types of clouds around the main HI emission in those datacubes as well. I followed the same process that I outlined in Section 4.1 for the Bridge and Stream cube. I kept the same physical scales for these data as I had for the SMC as these fields are adjacent to the SMC field, so likely at a similar distance. I encountered a problem with the Stream data as the noise was particularly high around the edges of the datacube. Even with the criterion I set previously, that $\sigma_{XY}(v)/m_{XY} > 3$, outliers were still being detected in these areas. These edge detections look very different from the other and are inherently unreliable, so I could easily reject them in the by-eye verification process, but they should not have been detected as peaks in the first place. To mitigate this, I used the beam response map that was supplied by the GASKAP survey data team, and masked any areas where the beam response is below 85%. Once I had repeated the process I obtained the collection of outliers for the two fields, which are shown in Figure 4.4.

For the LMC, it was necessary to adjust the number of pixels that represent the 250pc scale that I used with the SMC. This is because the LMC is closer than the SMC, at 49.89 kpc (using the recommended canonical distance in de Grijs et al. (2014) obtained from an assessment of 233 measurements). Since Equation 3.4 shows the physical size of an object is proportional to the distance to the object, the ratio of pixels that represent 250pc in the SMC vs LMC is the same as the ratio of the distances between the two. Rather than adjusting the number of pixels that are used in Equations 4.5 and 4.6, I need to change the X and Y values that represent the size of each box for the dispersion measurement. This way, I am measuring the dispersion on the same physical scales between the LMC and SMC to make the comparison reasonable.

$$\frac{X_{LMC}}{X_{SMC}} = \frac{d_{SMC}}{d_{LMC}} \quad (4.8)$$

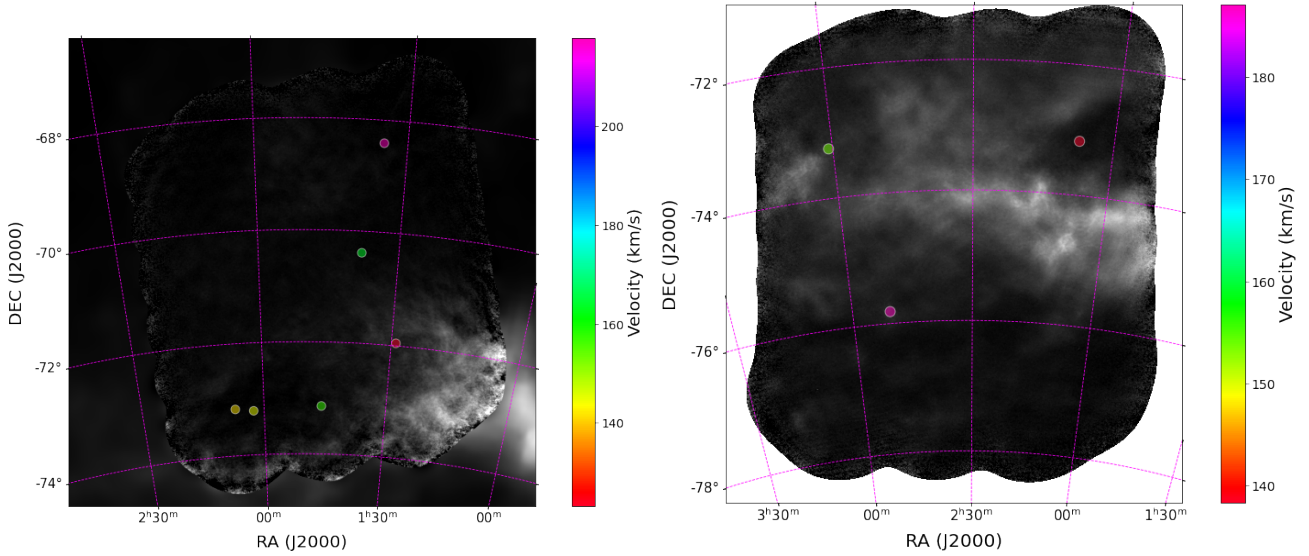


Figure 4.4: Clumps identified for the Magellanic Stream (left) and the Magellanic Bridge (right) with their respective velocity shown with the colourbar. The total column densities of the fields are shown in greyscale, with white indicating maximum column density.

$$X_{LMC} = \frac{63}{49.89} X_{SMC} \quad (4.9)$$

$$X_{LMC} = 1.26 X_{SMC} \quad (4.10)$$

So the sizes of the boxes used for the LMC are changed from the SMC sizes of 10, 20, and 30 to 13, 26, and 39 pixels respectively (rounded to the nearest integer as pixels can not be non-integer values). After making this adjustment, the outliers of the LMC field are shown in Figure 4.5

One trend becomes very clear when looking at the results for fields other than the SMC, the relative absence of small scale structure. There are 5, 3, and 5 clumps for the Stream, Bridge and LMC field respectively. I count 5 clumps for the Stream field, even though 6 were identified as the clump at $\sim 130 \text{ km s}^{-1}$ in the south-west of the field is a duplicate of SMC clump 27. In this region there is overlap between the two fields, so identifying the same cloud in separate images indicates that this method can identify small clumps consistently between images.

Compared to these three fields, the SMC has 6-10 times more clumps. The measy amount of small scale structure in the LMC field can be reasonably explained by the fact that the field does not cover what would be considered the periphery of the LMC. It is focused on the area around the 30 Doradus region where there is a lot of strong emission. From Figure 4.5 it is clear that there is an area of little emission on the east side of the of field, however, the noise levels are too high to find any meaningful structure in this region.

The absence of clumps in the Bridge and Stream fields are more difficult to understand compared to the SMC. It is clear that in these fields there is diffuse large scale emission, so possibly there are less clumps that are removed

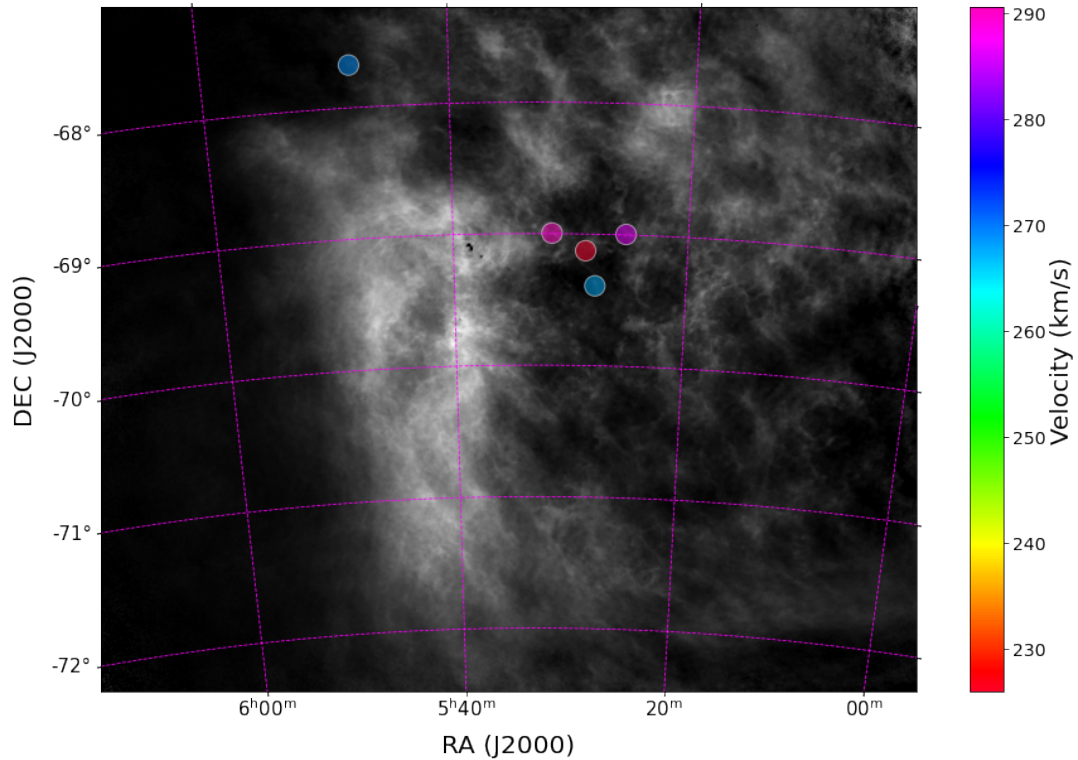


Figure 4.5: Clumps identified for the LMC with the velocity shown with the colourbar. The total column density of the LMC is shown in greyscale, with white indicating maximum column density.

from the diffuse emission. This would suggest that the SMC has an environment in which small scale structures are more likely to survive than than in the Bridge or Stream. Alternatively, are we seeing these fields in different stages of clump formation, where the Bridge and Stream have only just formed clumps and have not yet had events that have removed them from the main filaments or stripped them of their large diffuse envelope. Another point to consider is that the clumps that I am selecting for are on the periphery of the main HI emission in all fields. These fields have also been observed for only 10 hours, about half the integration time of the SMC field. This would mean the noise level is higher, precluding the detection of low column density clouds or clouds in already high noise areas like the periphery. This means that they are harder to detect using the criteria outlined in Section 4.1

The fields used in this section are preliminary data from the GASKAP survey, reduced with the same method as outlined in Pingel et al. (2022). They are from the pilot stage of survey. With the full survey, these fields will be re-observed and have a longer integration time, so the noise level will be lower in the full survey data. Repeating the same process I have outline in this Section on the full survey data could uncover more substructure in these areas.

4.2 Gaussian decomposition of anomalous clouds

Now that I had identified all of these clouds around the SMC periphery, the next step was to use ROHSA on each field to decompose them into the different HI phases. To do this, I created subcubes for each cloud by defining a

box with some width (Δx) and height (Δy) on the physical plane of the sky such that the full extent of the cloud was captured with the smallest i and j possible, where $i = \log_2 \Delta x$ and $j = \log_2 \Delta y$ and are both whole numbers. This is to reduce the number of levels of computation that ROHSA goes through in its multi-gridding process that scales by a factor of 2 each time.

I followed a similar process to what is detailed in Section 2.2 reducing the spectral range by excluding the part of the spectrum that includes the main body emission and passing this through to be fitted. I then began to fit these fields, estimating from the mean spectrum of each field how many components were required. I used a small parameter space exploration of 1, 5, 10, 50, and 100 for both λ_A and λ'_σ , as these affect the fit the most and set $\lambda_v = \lambda_\sigma = 10$. If the solution could have a small enough spread in the dispersions fitted to each component that there was a clear separation of the different phases, as well as no visible cloud structure in the χ_{red}^2 spatial map, then the solution was accepted. If a solution was accepted, I would reduce the number of components fitted until the solution is unacceptable to define the smallest amount of components required. Conversely, if from the initial estimate of component number, the solution is unacceptable, I would increase the number of components until an acceptable solution is achieved. Overall, I am looking to find a solution where the components fitted clearly belong to a specific phase of the HI, marked by the spread of their dispersion with the smallest number of components needed.

The results of this fitting process are shown in Table 4.2 with the number of components fitted, the λ_A and λ'_σ hyperparameter values, the χ_{red}^2 and skew measurement of the residual distribution for each clump region. I report the smallest mean σ of all the components to pull out the CNM structure, but notably many clumps have a $\overline{\sigma_{\text{min}}}$ above 2 km s^{-1} , especially clump 29, with a $\overline{\sigma_{\text{min}}} = 5.2$. This means that many of these clumps have components that fall into the temperature range of the UNM rather than CNM when measured by maximum kinetic temperature. So the population of clouds show a wide range of kinetic temperatures.

Now that I have a measure of the coldest components of each clump, I investigate the possibility of comparing these to a recent absorption survey of the SMC in the next section.

4.3 Comparison with absorption measurements

A comprehensive, unbiased absorption survey is detailed in Dempsey et al. (2022), which was conducted on the same ASKAP data used in this work. There were 229 continuum sources identified in this survey and so it is possible that some of these source lie within the regions defined for each of these clumps that have been identified. In the absorption survey they have a few notable detections at low column density, so even though the clumps looked at in this work are in typically low column density regions it may be possible to see absorption corresponding to their emission. I overlaid the absorption source catalogue on the map of all 31 identified clumps and found 7 sources that laid within the regions fit with ROHSA. I then cross-referenced these with the detections reported in Dempsey

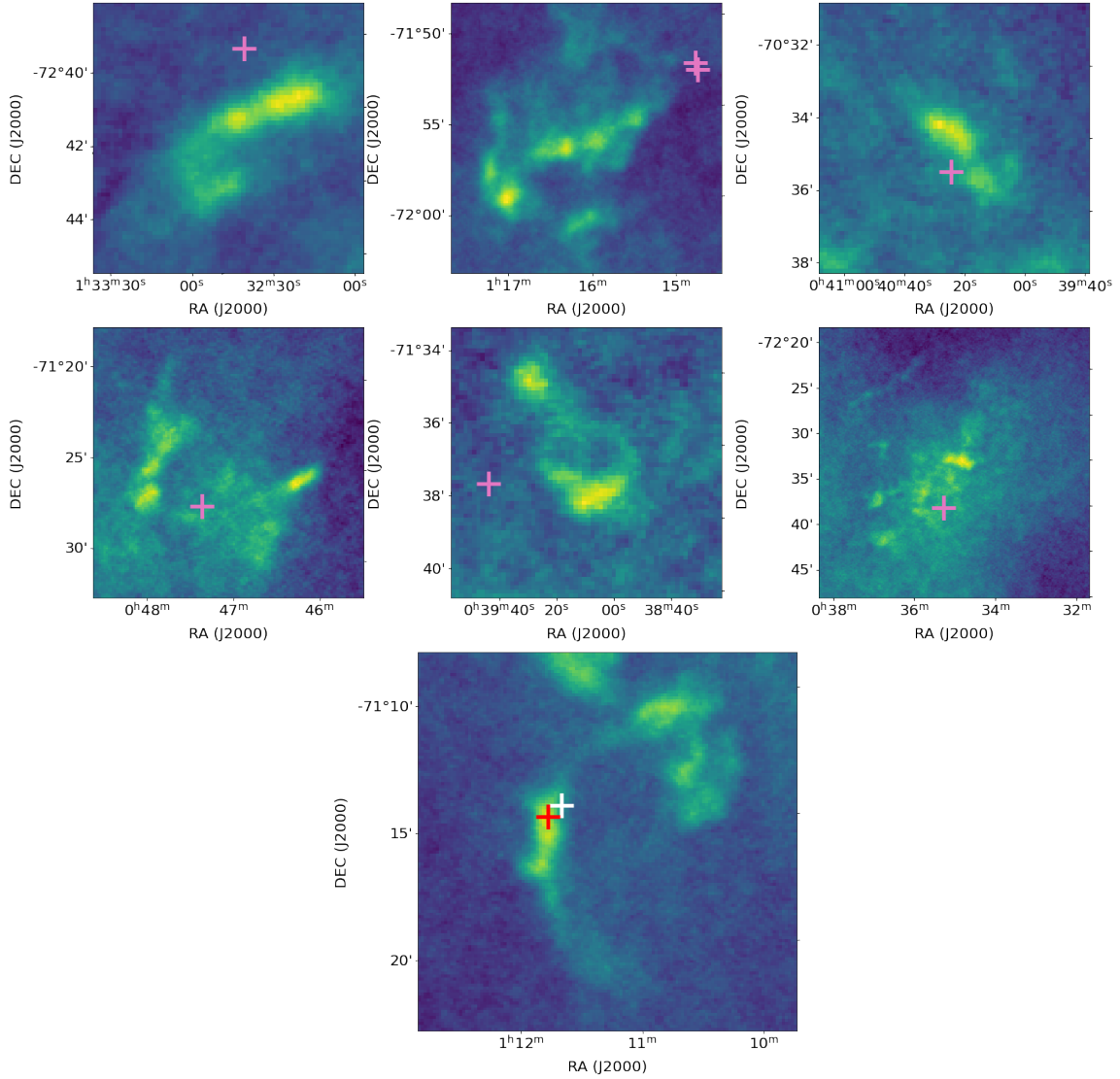


Figure 4.6: The absorption source locations for clumps 30, 26, 9 (top row), 17, 6, and 2 (bottom row) with non-detections in pink, overlaid on clump column density. Bottom: location of absorption source detection in white for clump 24. Red point indicates point of high column density for clump 24.

et al. (2022) and found that only one of those source has a significant detection. Clumps 30, 26, 9, 17, 6 and 2 have non-detections within their field whereas clump 24 has a detection within its field. These fields and the locations of the absorption sources, both detections and non-detections, are shown in Figure 4.6

What is clear from the positions of the absorption sources in the non-detections is that most of them lie off the dense areas of the clumps. For clumps 30, 26, and 6 their respective absorption sources do not lie on the clump at all. For clumps 9, 17, and 2 their respective absorption sources lie in the more diffuse regions where we do not detect much CNM at all. In fact, the CNM components fitted to these clumps all contribute less than 4.5 K of brightness temperature, at most, to the spectra at the locations of the absorption sources.

There is a detection of absorption in the field of Clump 24, for source J011134-711414. At this position the central velocity fitted to the CNM component is 139.4 km s^{-1} . The absorption detection is at a central velocity of ≈ 115

km s⁻¹. From inspection of the emission cube by-eye, there is clearly other emission that is associated with the absorption feature at that velocity. The brightness temperature of the data at that point at 139.6 km s⁻¹ (the closet channel to the CNM fitted velocity) is only 1.8 K. It seems that the absorption source just misses the CNM detected by ROHSA. This is suggested in Figure 4.6 where the position of the absorption source in white lies just to the side of the high column density region of the clump. In Figure 4.7 I show the emission spectrum from the emission cube use in this work and the absorption spectrum from Dempsey et al. (2022) for absorption source J011134-711414. The blue line shows where the absorption in Dempsey et al. (2022) was detected and the green line shows where the CNM in emission is centred at that location. It is clear that there is no absorption above the noise level at these velocity of the clump and very little emission at that point. In contrast, when shifting onto the densest part of the clump, the emission spectra changes drastically (shown in red), with the emission from the cloud peaking at the purple line. Clearly absorption source J011134-711414 just misses the CNM of the clump.

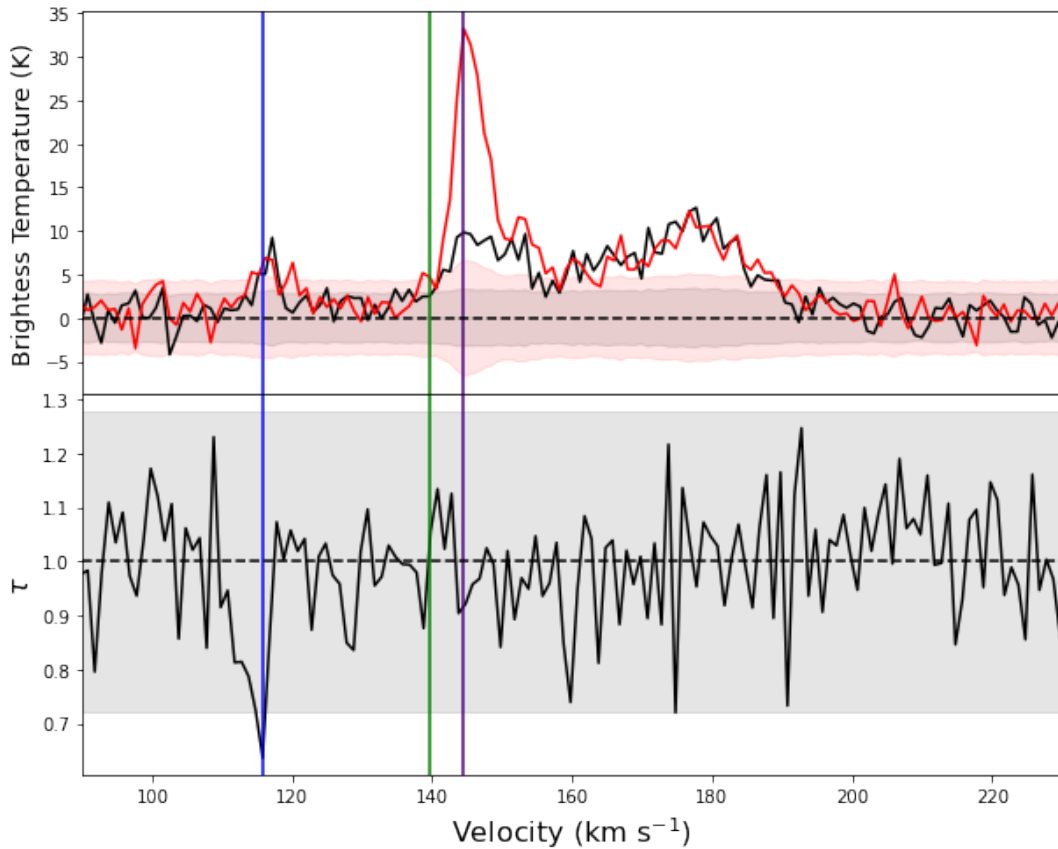


Figure 4.7: Top: Emission spectra for absorption source J011134-711414 (black), emission spectra for nearby point, positions shown with white and red crosses in Figure 4.6 respectively. The grey and red envelopes indicate the 3σ uncertainty level. Bottom: Absorption spectra for the same source with the 3σ uncertainty level in grey. The blue line is the peak of the absorption (115 km s⁻¹), the green line is the CNM velocity fitted with ROHSA at absorption source (139.6 km s⁻¹), and the purple line is the CNM velocity fitted with ROHSA at the red cross (144.5 km s⁻¹).

A non-detection in the survey, like any other absorption survey, does not definitively prove that there is no CNM in absorption towards a source, it simply means that there was no absorption detected above the required signal-

to-noise level, which was 5 in [Dempsey et al. \(2022\)](#). More sensitive surveys in the future may uncover shallow absorption that was previously hidden in the noise for known sources, or detect fainter sources which will give better coverage of the field. The distribution of background absorption sources is a variable out of an observers control, so it is purely up to chance whether an absorption source will intersect a HI structure of interest and absorption can be detected from that structure. Unfortunately in this work, this did not occur, so I am unable to make any comprehensive comparison to the absorption survey undertaken by [Dempsey et al. \(2022\)](#).

4.4 Measuring cloud HI density

To accurately characterise the properties of the HI phases in the newly identified clouds, there needs to be a way to measure density. Unfortunately, in this data there is no measure of the 3D structure of the SMC, even in other tracers it is an area of ongoing research. There are stellar studies that suggest the SMC is elongated over multiple kpc along the line of sight ([Tatton et al. 2021](#), [Jacyszyn-Dobrzyniecka et al. 2016](#), [Ripepi et al. 2017](#)). These studies are based on stellar objects and do not provide a lot of information about the periphery of the SMC. With sparse knowledge of the 3D structure, it is difficult to definitively assert the depth of the SMC in HI along the line of sight, much less the 3D structure of any one HI cloud or filament within it. Integrating HI data along the line of sight gives the column density after multiplying by the conversion factor, so to obtain a number density (3D density) all that is needed is a way to estimate the depth along the line of sight. Previous studies of HVCs and filaments ([For et al. 2016](#), [McClure-Griffiths et al. 2006](#)) have made the assumption that HI structures are as deep as they are wide. I work under this same assumption, that in the case of a spherical cloud, the width in both axes is similar to the depth and in the case of a filament, the width is similar to the depth. Of course, some clouds may be in orientations that would make this assumption invalid, for example a filament may be orientated in such a way that its longest axis is lying parallel to our line of sight. However, typically a filament is distinct from a cloud when the aspect ratio (length/width) is above a certain value (e.g. 7 in [Ma et al. \(2023\)](#) or 10 in [Clark et al. \(2014\)](#)), so treating this case as we see it, a spherical cloud, we would overestimate the density by that same factor. Overestimations on the order of 10 should be obvious when comparing with the other cloud measurements.

The shapes of some of these clumps are fairly irregular, so there needs to be enough flexibility in the model of density to capture this irregularity. Some clumps can be well-described by an elliptical shape in the plane of the sky, while some are filamentary with varying degrees of linearity or curvature in their shape.

It is very straightforward to fit a model to an elliptical clump, as the centre is easily defined by taking the point with the highest column density. It is not so straightforward to define the extent of a filament. So to tackle this problem, I employed the use of the python package `radfil` ([Zucker and Chen 2018](#)) which is designed to fit the width of a filament in a column density image. The package first defines a spine for the filament using the package `FilFinder` ([Koch and Rosolowsky 2015](#)) and then takes profiles of the column density of the image perpendicular to the spine

along its length. Once all these profiles are compiled, the user can chose to fit to the mean profile a Gaussian or Plummer profile. One of the problems that can arise when defining the filament spine is that the spine can take an unexpected path if the data is not masked to exclude the low column density areas. So to avoid this issue, the data is masked to the largest contour within certain contour levels. I do this by defining 7 levels of contour for each field, where the maximum level is defined by the maximum column density in the field. The largest contour from second highest level is then used as the mask, to capture the area around a density enhancement, to define the filament spine. So the threshold column density that defines this contour is not a consistent number across all clumps, but a consistent percentage of the total column density, regions above 67%. This spine is use to define a path through the filament in the image plane such that I can access the column density profiles along the filament length. For every point along the filament I take the column density profile to fit different density models to. To measure the width of a structure, a few methods were tried. I detail them in the following sections.

4.4.1 Model selection

The first method that was tried was assuming the structure had a constant density and a depth equal to the width of the structure. An ellipsoid with constant density (n_H) that has a height, width and depth defined as $2a$, $2b$, and $2c$, when projected onto the 2D plane has a column density profile as such:

$$N_H(x, y) = n_H \times 2c \sqrt{1 - \left(\frac{(x - x_c) \cos \theta + (y - y_c) \sin \theta}{a} \right)^2 + \left(\frac{(x - x_c) \sin \theta - (y - y_c) \cos \theta}{b} \right)^2} \quad (4.11)$$

where x_c and y_c are the x and y coordinates of the centre of ellipsoid and θ is the position angle in the anti-clockwise direction. I also assume in this model that c is the mean of a and b, as I am working on the assumption that the depth is equal to the width. This is because unlike in a filament where there is a clear large length and then a much smaller width, a and b are both measurements of the width of the structure, just in different axes. They both give us an estimate of the depth of feature, so I take c to be their mean.

If I treat these clumps as having a 2D Gaussian density profile then the results will differ, but it is instructive to compare for the ellipsoids how each model fits the data.

To define the 2D profile I use the general equation:

$$N_H(x, y) = A \exp \left(- \left(d(x - x_0)^2 + 2e(x - x_0)(y - y_0) + f(y - y_0)^2 \right) \right) \quad (4.12)$$

where:

$$d = \frac{\cos^2 \theta}{2a^2} + \frac{\sin^2 \theta}{2b^2} \quad (4.13)$$

$$e = -\frac{\sin 2\theta}{4a^2} + \frac{\sin 2\theta}{4b^2} \quad (4.14)$$

$$f = \frac{\sin^2 \theta}{2a^2} + \frac{\cos^2 \theta}{2b^2} \quad (4.15)$$

where θ is the position angle of the 2D profile in the anti-clockwise direction and a and b represent the Gaussian dispersion measures in the major and minor axes.

In this model c is not explicitly defined as a variable, but I calculate it in the same way as the first model, as the mean of a and b . Additionally, the number density can not be found directly from the fit, I have to make further calculations. This fit just provides the major and minor axes as well as the position angle of the ellipsoid that we calculate the density from. I take the widths a , b , and c to calculate the volume (V) of the ellipsoid as such:

$$V = \frac{4}{3}\pi abc. \quad (4.16)$$

Now there is a volume, but to get a density I need to know the total number of hydrogen atoms ($M_{H,tot}$), so I integrate over the area of the ellipse on the sky defined as every point (x,y) for where:

$$\frac{((x - x_0) \cos \theta + (y - y_0) \sin \theta)^2}{a^2} + \frac{(-(x - x_0) \sin \theta + (y - y_0) \cos \theta)^2}{b^2} \leq 1. \quad (4.17)$$

So with the $M_{H,tot}$ and V it is quite straightforward to get the mean number density in the clump:

$$n_H = \frac{3M_{H,tot}}{4\pi abc} \quad (4.18)$$

These are the two models I consider for modelling the column density distribution in the plane of the sky for ellipsoidal clumps.

I divided my sample of clumps into ellipsoids and filaments by eye, identifying compact clumps and structures with spherical density enhancements as ellipsoids. This included clumps 5, 12, 14, 15, 16, 18, 20, 25, 26, 28, 29, and 31. The rest were treated as filaments.

To fit the ellipsoids, I identified the x and y coordinates of the peak brightness temperature and used initial guesses of 10 pixels as values for a , b , and c . I passed these to the scipy minimiser function (`scipy.optimize.minimize`) to minimise the cost function (J) for both $N_H(x, y)_{\text{model}}$ models, a least squares measure of the residuals:

$$J = \sum_x \sum_y (N_H(x, y)_{\text{data}} - N_H(x, y)_{\text{model}})^2. \quad (4.19)$$

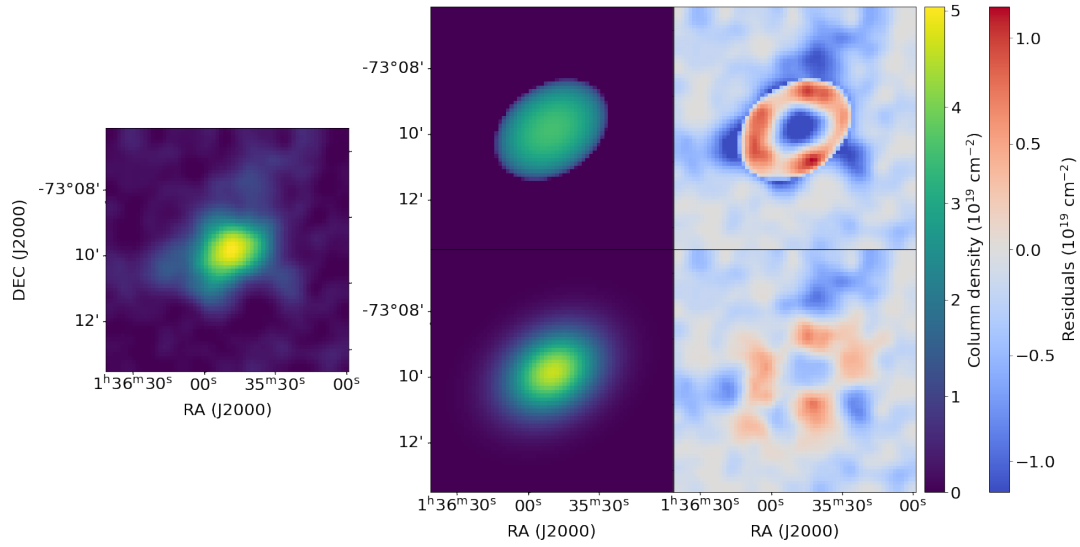


Figure 4.8: Left: Column density of clump 31 CNM component fitted with ROHSA. Middle: Column density modelled as a sphere of constant density (top), column density modelled as a 2D Gaussian profile (bottom). Right: residuals (model-data) for the respective models.

This minimisation gives the best fit parameters for both models so I can now look at how the models perform in recreating the 2D column density distribution. I use clump 31 to demonstrate the appropriateness of each model for a spheroidal clump. In Figure 4.8 it is clear from the residual map for the Gaussian model that it fits the column density distribution of the data better than the constant density model. This was consistent in all spheroidal clumps that were fit. So going from this point on I used the Gaussian profile method which has been used in other studies for filaments (Zucker et al. 2021, Juvela et al. 2012, Syed et al. 2022). Now that there is a model for the ellipsoidal clumps, I need to define some model for the filaments. Since the ellipsoids will be fit with this model and Gaussian profiles are used in the aforementioned studies for filaments, that is how the column density profile will be defined.

For the filaments I treat them as cylinders with length a and radius b (represented by the Gaussian dispersion measure) so that along the filament each slice of thickness Δa has a profile of :

$$N_H(x) = A \exp\left(\frac{-(x - x_c)^2}{2b^2}\right). \quad (4.20)$$

Fitting a function to a filament is not as simple as it was to do with an ellipsoid. There is no definite centre without defining a length, which can be dependent on what threshold of column density is used to define the whole filament. So I follow the bounds outlined in the beginning of this section, of identifying the filamentary structure by using contours. This is passed to the radfil package to define the filament spine and produce the 1D column density profiles along the spine. The spine is determined by producing a one pixel-wide representation of the mask and finding the shortest path through this, as explained in Koch and Rosolowsky (2015). The 1D profiles of the column density perpendicular to the filament spine are then compiled from the output. To each of these profiles I fit Equation 4.20 to obtain the dispersion b along the filament spine. Now there is a measure of the width along the

length of the filament, I utilise a similar method of calculating density from this as was done with the 2D Gaussian model for the ellipsoidal clumps. I take the volume of the filament to be:

$$V = \sum_{a'=0}^a \pi b(a')^2 \Delta a \quad (4.21)$$

where a is the length of the filament and Δa is the size between each point along the filament spine.

To calculate the area over which to sum the column density I take rectangles with the width of the step between each point of the filament spine, centred at the point and length that is the width defined from the fit. I create a mask that includes the pixels which fall within these rectangles and this is the 2D area of the filament in column density. I apply this mask to the column density image to return only the column density within the filament area. Summing up all pixels in this mask gives the total hydrogen mass $M_{H,tot}$.

There is one extra step to obtain the true total hydrogen mass $M_{H,tot}$, that is to convert the area of each pixel to a physical area. I use Equation 3.4 to convert the pixel scale which is 7" to a physical scale in cm, using the SMC distance of 63kpc derived in Di Teodoro et al. (2019b). This gives a physical scale of 6.60×10^{17} cm, so the area of each pixel is 4.36×10^{35} cm². Applying this conversion factor $A_{pix} = 4.36 \times 10^{35}$ to the column density image converts it to an image of total hydrogen atoms within each pixel. Then when summing over the mask area, a total number of hydrogen atoms within the filament area is obtained.

This conversion is also applied to the ellipsoidal method to calculate the total hydrogen density before converting it to number density in Equation 4.18. More generally, $M_{H,tot}$ is defined as:

$$M_{H,tot} = \sum^R 4.36 \times 10^{35} N_H(r) \quad (4.22)$$

where r represents the x and y position of the pixel and R is the total number of pixels with the defined structure, either defined by Equation 4.17 for the ellipsoid case or defined by the aforementioned mask for the filament case.

This method worked well for fairly straight filamentary structure, but did not manage to describe the area of curved structures well. My definition of the surface area is used to allow for variation in the thickness of the filament along its length using a measurement of the width of the filament which is determined from a fit of the profile, not from the width of the initial contour level used to define the spine. When the spine is curved, it leads to the rectangles used to define the area at each point overlapping in the inner part of the curve while simultaneously not fully covering the outer part of the curve, the comparison to this scenario and a straighter filament scenario is shown in Figure 4.9.

This highlights that while this method was used in an attempt to allow a flexible solution to the width of the filamentary structure, it does a poor job at handling strongly curved structures. It also has some very large widths fitted since when the structure curves around about 180 degrees the profiles along the spine then includes the density enhancement from either side of the curve. This can lead to cases where the 1D Gaussian profile is being



Figure 4.9: Areas derived for filaments on clump 22 (left) and clump 30 (right) when the filament area is modelled as a sum of rectangles along the filament spine.

fitted to a bimodal distribution, leading to strange results. It is difficult to automate a way to identify and correct the bimodal distributions in this process. Going through each clump individually is not particularly tedious for this sample size, but choosing where to segment a bimodal distribution will likely affect the widths fitted to the central peak in that area. Doing this kind of segmentation for each point along the spine could lead to inconsistent results due to human subjectivity.

Evidently, model fitting is not straightforward for irregular structures and the filamentary model described in this section is not robust enough to be unaffected by some rogue data points. Additionally, when I treated the more spherical clumps as filaments, to test how closely these models matched, the densities derived from the ellipsoidal model were consistently 1-2 times greater than those derived from the filamentary model. This inconsistency makes it difficult to treat these two groups of clumps the same in the sample. To combat these issues, in the next Section I describe a process to fit the clumps using their mean properties.

4.4.2 Measuring cloud depths

From the previous model attempts, it was clear that measuring properties along the length of a filament to build an area profile does not describe curvy filaments well. Additionally there are discrepancies between the densities calculated from the same clumps when treated as a filament or an ellipsoid in column density maps. The subjectivity of the distinction between ellipsoidal and filamentary is difficult to justify as some clumps look like a mix of the two. So, to be consistent, in this method I treated every clump as a filament.

I again employed the `radfil` package with the same contour level criteria for each clump. I then made use of the fitting methods that `radfil` provides. It first collates all the profiles along the spine, centring them at the position of the peak column density at each point in the spine. It then allows the user to define the window size that it uses to fit

the data and another window which it uses to fit a background profile. These windows can be defined symmetrically or asymmetrically by the user specifying the distance from the centre of the profile as a single value or a value for each direction away from the centre of the profile. I define this window by computing the distance from the peak value to 50% of the peak in each direction for the profile with the highest column density in the clump. I calculate this in both directions away from the centre and take the mean value (which I call x_W) to define a symmetric window for the radfil fit. I then specify the background fit window to be from x_W to $3x_W$.

The background fit can either be a horizontal line, assuming some constant value along the profile, or a sloped line, assuming a linear variation in the background from one end of the profile to the other. I chose to fit a horizontal line since should be no linear variation in column density as I am fitting the column density derived from the ROHSA model. Fitting the model CNM column density means there is no diffuse WNM emission in the column density map, so a constant background level fit should be able to adjust for any signal that may have been fit to narrow noise spikes in the original data. I show an example for clump 24 of the fitting results in Figure 4.10

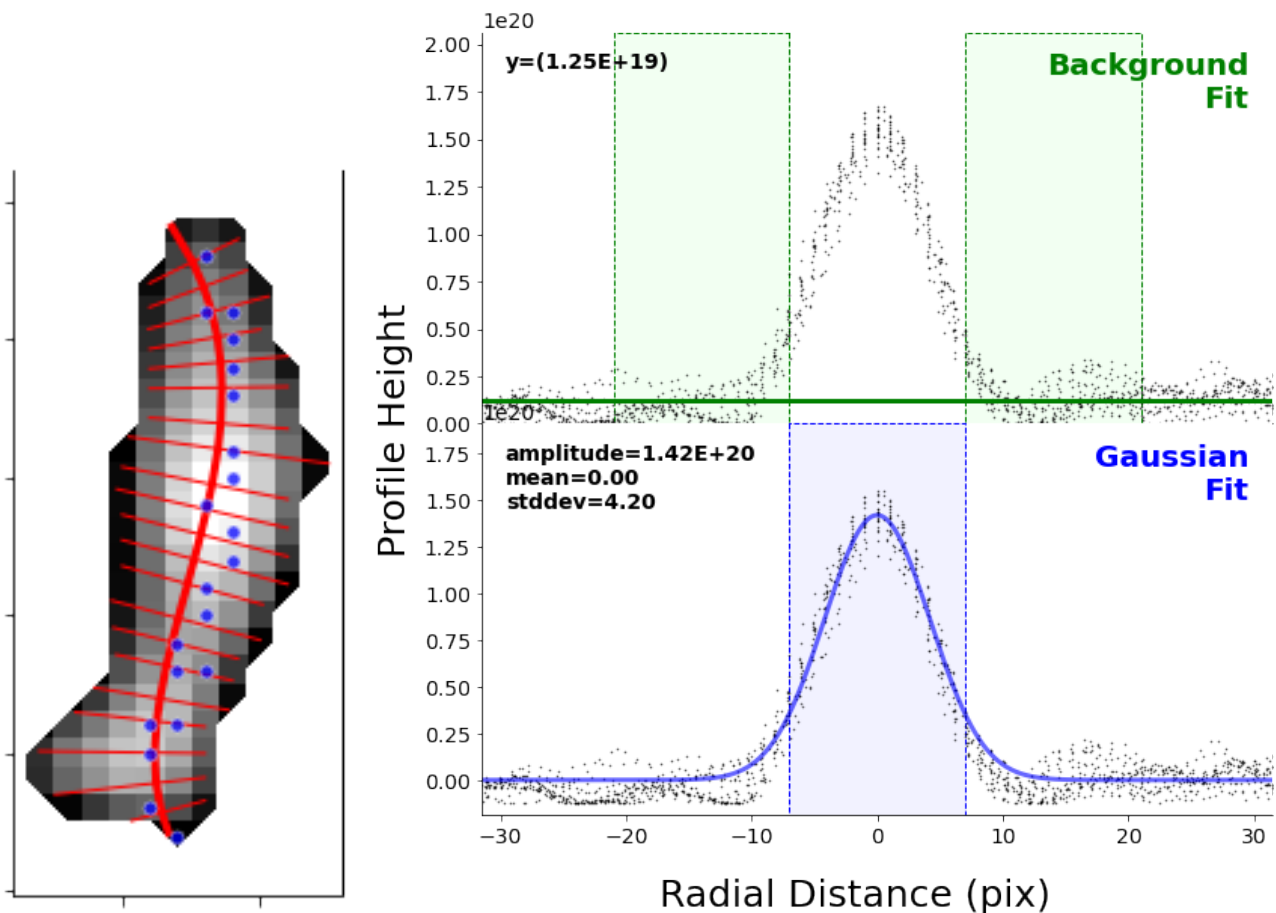


Figure 4.10: Left: The spine constructed along the densest part of clump 24. Right: The fit produced by radfil (Zucker and Chen 2018) showing the background level fit area in green (top) and Gaussian fit to are highlighted in blue (bottom).

Then I pass these to the radfil.fit_profile function selecting to fit a Gaussian profile, as the FWHM will be used to

ultimately measure the density. This is consistent with other studies that have looked at the number density of HI in coherent structures as they use the FWHM as the depth estimate for filaments or spherical clouds (McClure-Griffiths et al. 2006 and For et al. 2016). The fit provides a dispersion (σ) which is converted to a FWHM as $\text{FWHM} = 2.355\sigma$. The fitting function also provides a statistical error, but as stated in Zucker and Chen (2018), the variation introduced by the choice of fitting window is often larger than the statistical error. I also expect that the choice of contour levels that define the mask used for the spine definition will affect the fit. To estimate the uncertainty in the fitted width, which will be a combination of these two effects, I vary both conditions. I use the function in radfil which calculates the fit for all combinations of a set of fit and background windows supplied to it. I previously stated I would chose these windows by computing the distance from the peak value to 50% of the peak in each direction. I do this for calculation for 40% and 60% as well to obtain a larger and smaller fit window respectively. This give 3 different fit windows and 3 different background windows, so 9 different permutations. I then also change the contour levels used to calculate the spine. In Section 4.4 I stated that I used 7 levels for the contours, so I vary this to also include 6 and 8 contour levels, using the second highest contour level in both, defining the contour level at 60% and 75% of the peak column density respectively. This gives another 3 different measures, so all up 27 fits to the width of the filament. From these 27 values I take the median (\widetilde{W}) and standard deviation (σ_W), using the standard deviation formula for a small sample size:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (4.23)$$

One caveat to the contour level definition is that on occasion the contour level provides a contour area that is too small for a spine to be constructed for the filament. In this case, I move to lower and lower contour levels until a spine can be constructed. This only affects clumps that are fairly compact and spherical.

After this variation bootstrapping is completed, I combine the uncertainties on the fitted widths. To convert the pixel widths to physical widths, I use the pixel scale of 7" that corresponds to a physical scale of 2.14pc. There is an uncertainty on this conversion factor, since the distance to the SMC has a uncertainty of 5 kpc or 8% (Di Teodoro et al. 2019b). Combining the uncertainty from both sources follows Equation 4.24:

$$\sigma_W = W \text{ (pc)} \sqrt{\left(\frac{5 \text{ (kpc)}}{63 \text{ (kpc)}}\right)^2 + \left(\frac{\sigma_{W,p} \text{ (pix)}}{\widetilde{W} \text{ (pix)}}\right)^2}. \quad (4.24)$$

Then taking the peak column density within the contour for which the width has been measured, I obtain the number density n with Equation 4.25:

$$n = \frac{N_{\text{H, max}} \text{ (cm}^{-2}\text{)}}{\widetilde{W} \text{ (cm)}}. \quad (4.25)$$

There is an uncertainty in the column density too, however since these clumps were fitted with more components

Clump number	n	λ_A	λ'_σ	χ^2_{red}	skew	$\overline{\sigma_{\text{min}}}$ (km s ⁻¹)	\widetilde{W} (pc)	σ_W (pc)	$N_{\text{H, max}}$ 10 ²⁰ cm ⁻²	n (cm ⁻³)
1	4	10	80	1.24	0.003	1.5	66.4	13.9	2.92	1.09
2	4	10	10	1.23	-0.0001	2.1	21.1	41.4	2.24	3.18
3	2	10	10	1.20	0.007	1.8	21.3	6.2	1.53	2.33
4	4	10	1	1.23	-0.003	2.2	25.5	3.1	2.69	3.42
5	3	10	10	1.20	-0.018	2.2	21.3	2.1	1.39	2.11
6	3	10	50	1.22	-0.004	1.9	14.0	2.0	1.27	2.93
7	3	10	50	1.23	-0.001	3.0	18.5	1.9	1.86	3.25
8	2	10	10	1.25	0.007	1.4	18.1	2.1	1.98	3.54
9	3	50	10	1.20	0.004	2.3	15.4	1.7	1.61	3.39
10	2	10	10	1.17	0.005	1.5	22.7	3.6	2.29	3.26
11	3	10	10	1.17	-0.013	2.0	23.2	4.3	2.03	2.84
12	2	10	50	1.25	-0.002	3.7	24.0	3.4	2.93	3.96
13	4	10	50	1.24	-0.009	1.6	26.4	12.8	2.88	2.69
14	4	10	50	1.31	-0.015	1.9	39.0	12.9	4.61	3.83
15	2	10	10	1.18	0.008	2.3	17.7	2.7	1.52	2.77
16	2	10	10	1.23	0.006	1.4	15.3	2.6	1.56	3.29
17	3	10	50	1.21	-0.001	1.7	15.7	1.4	2.34	4.83
18	2	10	10	1.17	0.014	2.1	12.2	1.9	1.05	2.79
19	3	10	50	1.36	-0.004	2.5	13.9	4.6	3.09	3.39
20	3	10	100	1.35	-0.034	1.9	26.4	12.2	2.92	1.69
21	2	10	10	1.26	-0.004	1.3	23.0	6.1	1.35	1.90
22	3	10	10	1.25	-0.016	1.9	19.0	1.9	1.87	3.18
23	4	10	50	1.31	-0.005	1.9	29.5	6.0	1.72	1.89
24	4	3	3	1.88	0.0005	1.9	20.1	2.6	1.68	2.72
25	2	10	10	1.14	0.002	1.9	18.5	2.4	1.18	2.06
26	6	10	5	1.43	-0.007	1.6	26.3	4.4	2.72	3.34
27	2	10	10	1.17	-0.012	2.9	13.6	1.6	2.30	5.49
28	3	100	10	1.30	-0.075	2.0	20.2	2.0	0.87	1.40
29	4	70	10	1.24	-0.021	5.2	24.0	2.7	2.68	3.63
30	3	10	10	1.19	-0.012	3.5	24.6	8.3	2.28	3.00
31	3	50	10	1.20	-0.016	1.3	28.4	3.7	1.26	1.44

Table 4.2: Clump parameters and measured properties, with $\lambda_v = \lambda_\sigma = 10$. For each cloud is listed: Clump number, number of components fitted, λ_A , λ'_σ , χ^2_{red} value for the field, skew measure of the residuals after model subtraction, mean velocity dispersion of coldest component fit with ROHSA, width of cloud, uncertainty in the width, maximum column density of coldest component fit with ROHSA, number density.

than the clouds from Chapter 2 it is computationally prohibitive to run the full 300 ROHSA permutation bootstrapping method as described in 2.3 for all 31 clumps. Instead, I use the mean relative uncertainty of 0.252 from the CNM and WNM column density maps of the aforementioned clouds, calculated in Section 2.3. This means the total uncertainty in the density is then:

$$\sigma_n = n \sqrt{(0.252)^2 + \left(\frac{\sigma_W}{W}\right)^2}. \quad (4.26)$$

All these values are reported in Table 4.2. Note that for clump 2 there is a very large uncertainty. It is skewed by 3 out of the 27 permutations of the width fitting process using the fitting window determined by the 40% level measurement. These three widths were 6 times higher than the rest of the distribution and consequently greatly increased the uncertainty. It is for this reason I use the median widths in Table 4.2 and Equations 4.24 and 4.25.

Chapter 5

The conditions of the Interstellar Medium around the SMC

In this chapter I detail the process of calculating the phase diagram of HI following the equations outlined in [Wolfire et al. \(1995\)](#) and [Wolfire et al. \(2003\)](#) that are discussed in Section [1.2.1](#). Then I discuss how to relate the properties of the clouds analysed in the previous Chapters to the phase diagram. I detail the process used to fit the number densities and kinetic temperatures of each cloud to phase diagrams with varying ISM conditions. I then discuss the best fit ISM conditions distribution for all the clouds. Finally, I compile a full table of cloud properties from all the work in this thesis, including a measure of deviation velocity.

5.1 Modelling the HI phase diagram

The phase diagram of HI represents how the density and pressure should relate to each other if the gas is ideal and in equilibrium. Equilibrium is obtained when the rate of cooling equals the rate of heating. More practically, it can be calculated by taking all the important cooling and heating rates for the densities of the HI and numerically solving for the temperature where the total heating rate equals the total cooling rate. I follow the equations outlined in [Wolfire et al. \(1995\)](#) and [Wolfire et al. \(2003\)](#) for heating and cooling. First the heating. I take into account photoelectric heating (PE), heating by X-rays (XR), and by cosmic rays (CR), which I list in Equations [5.1](#) - [5.7](#) below:

$$\Gamma_{pe} = 1.3 \times 10^{-24} \frac{n_e}{n} \epsilon G'_0 \quad (5.1)$$

from Equation 20 in [Wolfire et al. \(2003\)](#) where n_e is the electron density (defined in Equation [5.13](#)), n is number density, G'_0 is the incident FUV radiation field normalised to the value in [Habing \(1968\)](#). ϵ is the heating efficiency, defined as such:

$$\epsilon = \frac{4.2 \times 10^{-2}}{1 + 4.0 \times 10^{-3} (G'_0 T^{1/2} / n_e \phi_{\text{PAH}})^{0.73}} + \frac{3.7 \times 10^{-2} (T/10^4)^{0.7}}{1 + 2.0 \times 10^{-4} (G'_0 T^{1/2} / n_e \phi_{\text{PAH}})} \quad (5.2)$$

where T is the temperature in Kelvin and ϕ_{PAH} is the PAH collision rate factor. The heating rate from X-rays is defined as:

$$\log_{10}(\Gamma_{XR}) = f_6(p_2)(-26.5 - 0.920p_1 + 5.89 \times 10^{-2}p_1^2)0.96 \exp \left[- \left(\frac{p_1 - 0.38}{0.87} \right)^2 \right] \quad (5.3)$$

from Equation A9 in [Wolfire et al. \(1995\)](#) where

$$p_1 = \log_{10}(N_W/10^{18}) \quad (5.4)$$

$$p_2 = \log_{10}(n_e/n) \quad (5.5)$$

$$f_6(p_2) = 0.990 - 2.74 \times 10^{-3}p_2 + 1.13 \times 10^{-3}p_2^2. \quad (5.6)$$

N_W is the absorbing column density of WNM. It is important to note that Equations [5.3](#)[5.6](#) are empirical relations derived for the conditions of the local ISM. I discuss the implications of this in Section [5.1.1](#).

The heating from cosmic rays is defined as:

$$\Gamma_{CR} = \zeta_{CR} E_h(E, n_e/n) \quad (5.7)$$

from Equation 4 in [Wolfire et al. \(1995\)](#). Where ζ_{CR} is the cosmic ray ionisation rate, E is the primary electron energy, defined from [Spitzer \(1978\)](#) as 35eV, and E_h is defined in [Wolfire et al. \(1995\)](#).

Those are the 3 heating terms included in this work. The other side of this equilibrium necessitates the cooling terms including cooling by CII, OI, Ly α and recombination, which I list in Equations [5.8](#)-[5.12](#) below:

$$\Lambda_{\text{CII}}^{\text{H}} = 3.15 \times 10^{-27} e^{-0.92/T_2} Z'_g \quad (5.8)$$

$$\Lambda_{\text{CII}}^{\text{e}} = 1.4 \times 10^{-24} T_2^{-0.5} e^{-0.92/T_2} Z'_g \quad (5.9)$$

$$\Lambda_{\text{OI}}^{\text{H}} = 2.5 \times 10^{-27} T_2^{0.4} e^{-2.28/T_2} Z'_g \quad (5.10)$$

where $T_2 = T/100$ and Z'_g is the metallicity normalised to the solar value of 1. from Equations C1, C2 and C3 in

Wolfire et al. (2003). The cooling from Lyman α is defined as:

$$\Lambda_{\text{Ly}\alpha} = 7.3 \times 10^{-19} n_e n \times e^{-118400/T} \quad (5.11)$$

from Equation 6-12 in Spitzer (1978). The cooling from recombination is defined as:

$$\Lambda_{\text{rec}} = 4.65 \times 10^{-30} T^{0.94} (G'_0 T^{0.5} / n_e)^\beta \frac{n_e}{n} \quad (5.12)$$

where $\beta = 0.74/T^{0.068}$ from Equation 9 in Wolfire et al. (1995).

The other important function that we need is the electron density which is calculated as so:

$$n_e = 2.4 \times 10^{-3} \zeta'_t{}^{0.5} T_2^{0.25} G'_0{}^{0.5} Z'_d{}^{0.5} \phi_{\text{PAH}} \quad (5.13)$$

where ζ'_t is total ionisation rate of hydrogen normalised to the local value and Z'_d is the dust to gas ratio normalised to the local value, from Equation C15 in Wolfire et al. (2003).

So to determine the temperature T at which the system is in equilibrium at a given density n we just add all the heating rates together and all the cooling rates together, set them as equal and solve for T , like so:

$$n (\Gamma_{pe} + \Gamma_{XR} + \Gamma_{CR}) = n (n \Lambda_{\text{CII}}^{\text{H}} + n_e \Lambda_{\text{CII}}^{\text{e}} + n \Lambda_{\text{OI}}^{\text{H}} + n \Lambda_{\text{Ly}\alpha} + n_e \Lambda_{\text{rec}}). \quad (5.14)$$

To go from the relation between temperature and density to the relation between the thermal pressure (P/k_B) and density which is most often shown, we make one assumption, that the gas is ideal. This way through employing the ideal gas law:

$$P = nk_B T \quad (5.15)$$

$$P/k_B = nT \quad (5.16)$$

So to get the relation between thermal pressure and density I actually calculate the relation between nT and density (n).

In the Wolfire papers they typically look at galactic conditions, which are not necessarily applicable to SMC, however the obvious difference between the SMC and the Milky Way is the metallicity (Z_g), which is a variable present in the cooling rate equations. The dust to gas ratio is also affected by the metallicity and is expected to scale with metallicity so I adopt that $Z_g = Z_d = 0.2$, which is consistent with the range of the gas to dust ratio found in Welty et al. (2016) and the mean metallicity of the SMC (Russell and Dopita 1992). In Wolfire et al. (1995) and Wolfire et al. (2003) they investigate the effect of metallicity and also vary the dust to gas ratio and metallicity together. The trend

that is evident from their results, shown in Figure 5.1 is that the decreasing metallicity pushes the pressure window for the two phase equilibrium to higher and higher values.

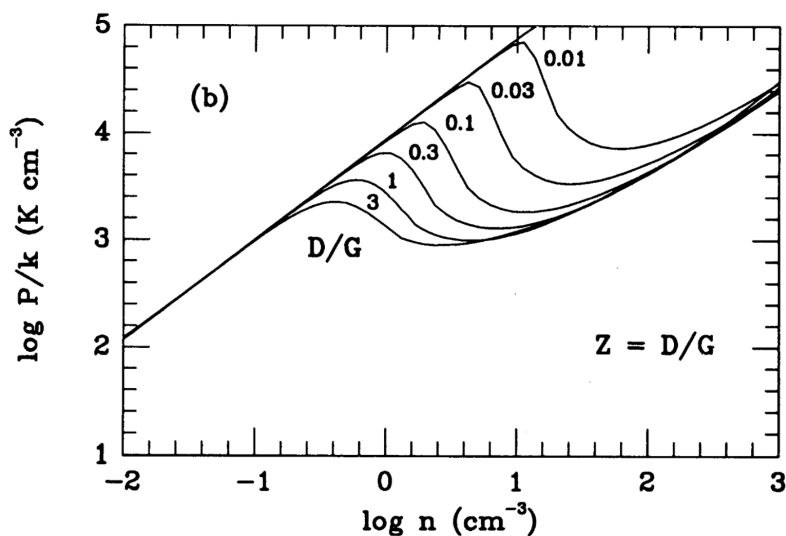


Figure 5.1: Figure 6(b) from Wolfire et al. (1995) showing how the phase diagram changes as Z_g and Z_d change.

In this work I use `scipy.optimize.minimize` to numerically solve this using the Powell minimiser. I solve Equation 5.14 for 300 values of density (n) equally spaced over of log range of -2 to 3 , which captures the lower range of the WNM density and the higher range of the CNM density and is consistent with the ranges used in Wolfire et al. (2003). I do not take into account any of the fine structure lines and using the analytical expressions does not account for the full chemical modelling conducted in Wolfire et al. (1995) and Wolfire et al. (2003), so there will be some variation in the numerical values derived for the curves compared to those in the Wolfire series of papers. To compare the limits of the modelling, I explore some of the dependency trends that are demonstrated in the Wolfire series of papers. I show in Figure 5.2 the same trend with metallicity that is demonstrated in Figure 5.1

Evidently the low density end of the phase diagram does not behave in the same way as expected from Figure 5.1 so this is likely a limit of the modelling and values in this region should be treated with caution.

In Figure 5.3 I show the trend as G'_0 is varied by specified factors from its standard value of 1.7.

Again, there are values at low density for some curves that are not as linear as in the Wolfire curves, so this region should again be treated with caution. However since the focus in this work is on the colder parts of clouds, I do not expect to find any clouds residing in this region of the density range.

5.1.1 Measuring the thermal properties of clouds

In Chapter 4 I measured the densities of a number of clouds around the SMC periphery, but if I want to compare them to the phase diagrams I outlined in the previous section, I need to know the temperature as well, as shown by Equation 5.16. The temperature of the coldest part of the clouds is something that can be measured from the

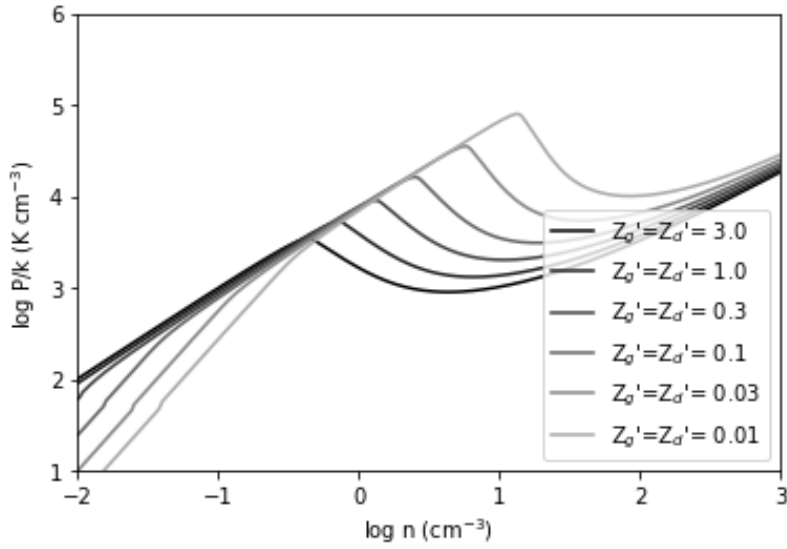


Figure 5.2: Reconstructed phase diagram in this work as the values of Z_g and Z_d are varied together over the same range as in Figure 5.1.

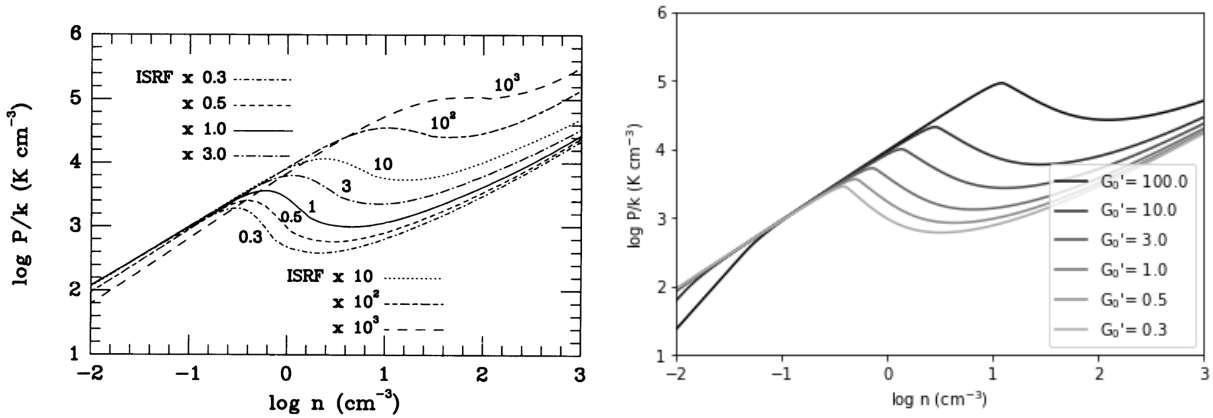


Figure 5.3: Phase diagram as G'_0 is varied from Wolfire et al. (1995) (left) and reconstructed in this work (right).

dispersion (which I have previously defined as σ , but to avoid confusion with the uncertainties defined, I will define as S) of the Gaussian component. From Equation 3.1 the maximum kinetic temperature can be derived directly from a measurement of the dispersion of the Gaussian component. I take the mean value of the fitted dispersion measure within the contours defined in Section 4.4.2 to use for the calculation and the standard deviation in this contour as the uncertainty (σ_S) in that value. I combine this with the measure of density from Table 4.2 and derive the thermal pressures of each cloud as in Equation 5.16. To get the uncertainty (σ_T) in $T_{k,max}$ I propagate the uncertainties in Equation 5.17.

$$\sigma_T = T_{k,max} \times 2 \frac{\sigma_S}{S} \quad (5.17)$$

Clump number	n (cm^{-3})	σ_n (cm^{-3})	$T_{k,max}$ (K)	σ_T (K)	P/k (K cm^{-3})	$\sigma_{P/k}$ (K cm^{-3})
1	1.09	0.32	298	39	326	105
2	3.18	6.04	569	185	1811	3489
3	2.33	0.79	292	53	679	261
4	3.42	0.91	492	130	1682	633
5	2.11	0.56	565	44	1193	329
6	2.93	0.80	390	70	1143	372
7	3.25	0.86	837	210	2717	992
8	3.54	0.94	201	23	713	207
9	3.39	0.90	835	202	2831	1018
10	3.26	0.90	327	42	1066	325
11	2.84	0.81	470	71	1336	430
12	3.96	1.07	1351	304	5349	1885
13	2.69	0.75	297	75	799	300
14	3.83	1.40	467	72	1790	710
15	2.77	0.76	570	85	1580	492
16	3.29	0.92	216	33	712	227
17	4.83	1.28	468	54	2261	652
18	2.79	0.77	365	80	1018	359
19	3.39	0.93	711	103	2411	750
20	1.69	0.50	502	71	850	280
21	1.90	0.61	361	140	686	347
22	3.18	0.84	443	23	1410	380
23	1.89	0.55	541	51	1024	315
24	2.72	0.73	328	36	892	259
25	2.06	0.55	458	35	943	264
26	3.34	0.93	333	72	1112	393
27	5.49	1.47	907	146	4976	1554
28	1.40	0.37	378	43	527	152
29	3.63	0.97	3388	189	12296	3343
30	3.00	1.10	1522	233	4561	1818
31	1.44	0.39	212	21	305	88

Table 5.1: Further measured properties for each clump. For each cloud is listed: number density, uncertainty in the number density, maximum kinetic temperature of coldest component fit with ROHSA, uncertainty in the maximum kinetic temperature, thermal pressure, uncertainty in the thermal pressure.

Then the uncertainty in P/k is:

$$\sigma_{P/k} = P/k \sqrt{\left(\frac{\sigma_n}{n}\right)^2 + \left(\frac{\sigma_T}{T}\right)^2} \quad (5.18)$$

The values for each clump are shown in Table 5.1.

These values can be used to fit to models of the phase diagram outlined in Section 5.1, varying certain parameters. I chose to vary the value of G'_0 which is set by the stellar population in a given area, thus will likely vary from the centre of the SMC to the outskirts. The SMC is a very spatially and kinematically disturbed system and that extends to the stellar structure. It is possible that the inhomogeneous stellar distributions could enhance the ionising radiation in a given area. Additionally the FUV radiation field strength is shown to decrease as the radius of the Milky Way increases (Wolfire et al., 2003) and so I would expect the strength of the FUV field to be lower on average further from the SMC main body. With this in mind I vary the values of G'_0 as $-2 < \log_{10}(G'_0) < 2$.

I mentioned in Section 5.1 that the equations describing the heating due to X-rays are tuned to the local ISM

conditions. This heating rate is scaled by N_W and n_e in a non-linear way, so there is no obvious adjustment to this equation for the SMC environment. This heating rate is orders of magnitude smaller the photoelectric heating rate (seen in Figure 1.1), particularly at CNM densities and in [Wolfire et al. \(1995\)](#) they show that changing the value of N_W does not affect the shape of the phase diagram to a degree comparable to the effects of changing G'_0 , Z_g or Z_d . This equation is likely not applicable to SMC conditions, but in the absence of measurements of how the X-ray ionisation rate changes across the SMC, I use the standard value of $N_W = 10^{19}$ used in [Wolfire et al. \(1995\)](#).

As the numerical solver is computationally expensive and using a minimiser that calculates a phase diagram at each iteration would take a prohibitively long time, I precomputed the solutions over the values for G'_0 defined above in tenths of each log step range. This gives 37 phase diagrams to use for comparison to the measured points from the SMC clump catalogue. To make a comparison of the values measured for the clumps and the phase diagrams at a given set of conditions, I used a least squared method to minimise the distance between the point that represents the clump and the phase diagram curve.

Additionally, $T_{k,max}$ is an upper limit on the temperature of the HI measured, as part of the dispersion is due to turbulent motions. To explore the case for the case where Mach number (\mathcal{M}) is 1, the derived temperatures are divided by a factor of $\sqrt{2}$ and these values are also fit. The results for the $\mathcal{M} = 0$ and $\mathcal{M} = 1$ cases are shown below in Figure 5.4

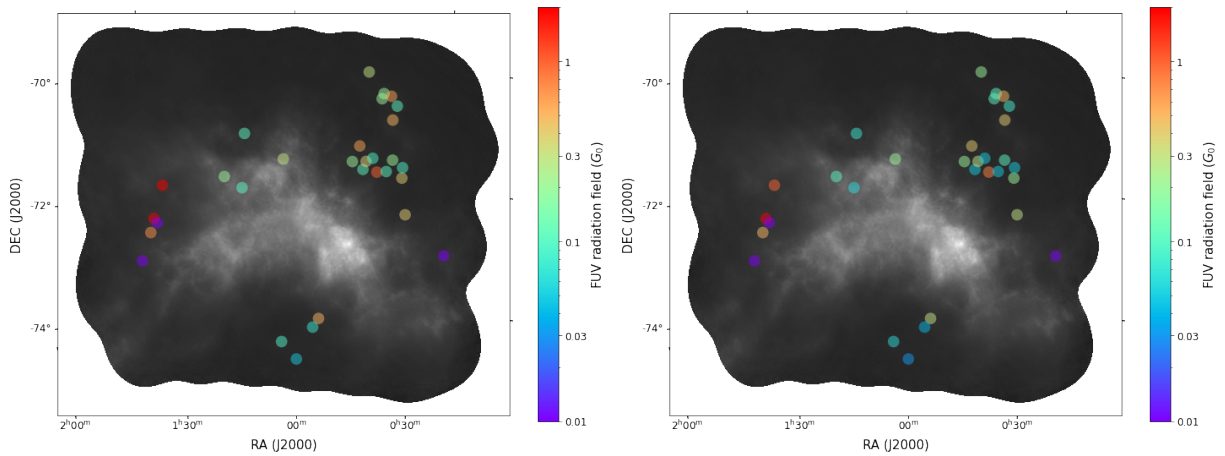


Figure 5.4: Distribution of the best fit values for G'_0 when $\mathcal{M} = 0$ (left) and when $\mathcal{M} = 1$ (right), overlaid on the total density map of the SMC.

What is evident from Figure 5.4 is the large variability of the values of G'_0 . This is normalised to the value of the Habing field ([Habing 1968](#)), for which the local value is 1.7. The maximum value fitted in this exploration was 2, so this suggests that the FUV radiation field strength in the outskirts of the SMC does not exceed the local value by a large amount and is overwhelming weaker than that in most areas.

There are some consistent values in the clustering around the north-west group of clumps, but still a variation of over an order of magnitude in clumps that are spatially very near to each other. It is important to remember the

ambiguity of the depth of the line of sight for this dataset. In the HI data there is only radial velocity information for these HI clumps, which does not definitively correspond to a distance along the line of sight. Additionally the SMC is very disturbed in all three dimensions, so components that are at the same velocity are not necessarily physically close to each other either. This is to say that variations in the values of the radiation field are feasible if these clumps are dispersed over multiple kpc. Clearly the relative value of the radiation field changes around the periphery of the SMC, which could be due to anomalous stellar populations at distances or velocities quite removed from the SMC disc. [Martínez-Delgado et al. \(2019\)](#) and [Piatti \(2022\)](#) identify stellar groupings in the north-eastern periphery, where some of the highest values for G'_0 are fitted in Figure 5.4, so the stellar population around the SMC could well be producing strong local FUV radiation fields in the SMC periphery.

In terms of the distribution of fitted value in the cases of no turbulence and transonic turbulence, the relative distribution of values is very similar. The fitted G'_0 in the transonic case are all smaller than in the no turbulence case. The distribution of the values over the range shown does change slightly, shown in Figure 5.5

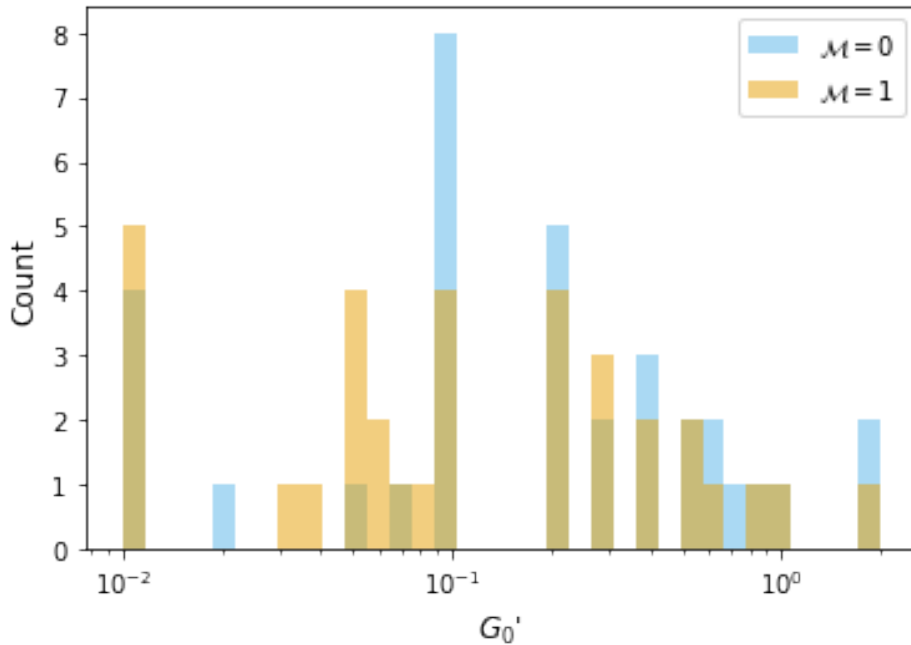


Figure 5.5: Distribution of the best fit values of all clumps and large clouds for G'_0 when $\mathcal{M} = 0$ (blue) and when $\mathcal{M} = 1$ (orange).

The distribution in the no turbulence regime has a strong peak at $G'_0 = 0.1$, whereas in transonic regime the distribution is comparatively flat. The sample size of the distribution is quite small and does not evenly sample the spatial extent of the SMC periphery so this peak likely does not represent a typical value for G'_0 of the SMC periphery.

Cloud	\tilde{W} (pc)	σ_W (pc)	$N_{H, \max}$ 10^{20} cm^{-2}	n (cm^{-3})	σ_n (cm^{-3})	$T_{k, \max}$ (K)	σ_T (K)	P/k	$\sigma_{P/k}$
Alpha	23.1	2.2	3.66	5.14	1.38	519	74	2666	811
Hook	33.7	3.6	1.83	1.76	0.48	246	37	432	135
Gamma	20.6	5.6	0.81	1.28	0.47	228	63	292	135

Table 5.2: Measured properties for each large cloud. For each cloud is listed: width of cloud, uncertainty in the width, maximum column density of coldest component fit with ROHSA, number density, uncertainty in the number density, maximum kinetic temperature of coldest component fit with ROHSA, uncertainty in the maximum kinetic temperature, thermal pressure, uncertainty in the thermal pressure.

5.1.2 Measuring the thermal properties of the three large clouds

So far I have conducted this analysis on just the small scale structures. To explore more of the SMC periphery I make the same measurements for the larger structures that were characterised in Chapter 3. I measure the density with the same process outlined in Section 4.4.2 and measure the temperature with the same process outlined in Section 5.1.1. Using the same contour levels as are outlined in Section 4.4.2 were too prohibitively small when applied to these clouds, which did not allow radfil to build a spine. So, to account for this, I adjusted the contour level that is used to define the mask the spine is built upon, to be the third highest level, rather than the second. Evidently the density enhancements are very compact. The positions at which the peak column densities were taken to calculate the densities are shown in Figure 5.6 and the results from this fitting are listed in Table 5.2.

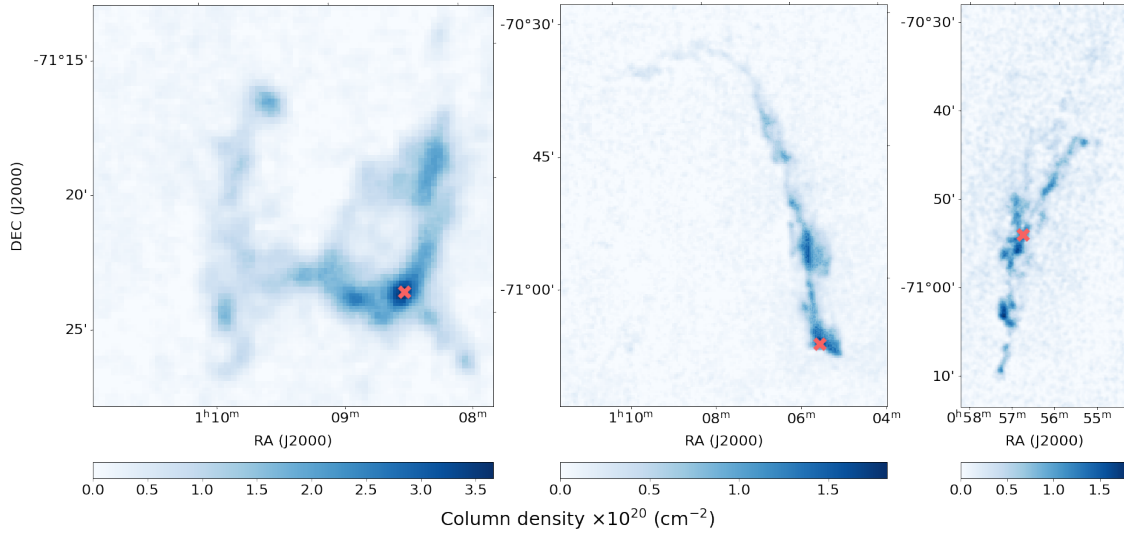


Figure 5.6: Location of the position at which the peak column density was measured (red marker) from the CNM column density map of each large cloud; Alpha (left), Hook (centre), and Gamma (right).

Adding these to the rest of the small scale clumps gives the whole picture from the catalogue of clouds of the SMC, shown in Figure 5.7.

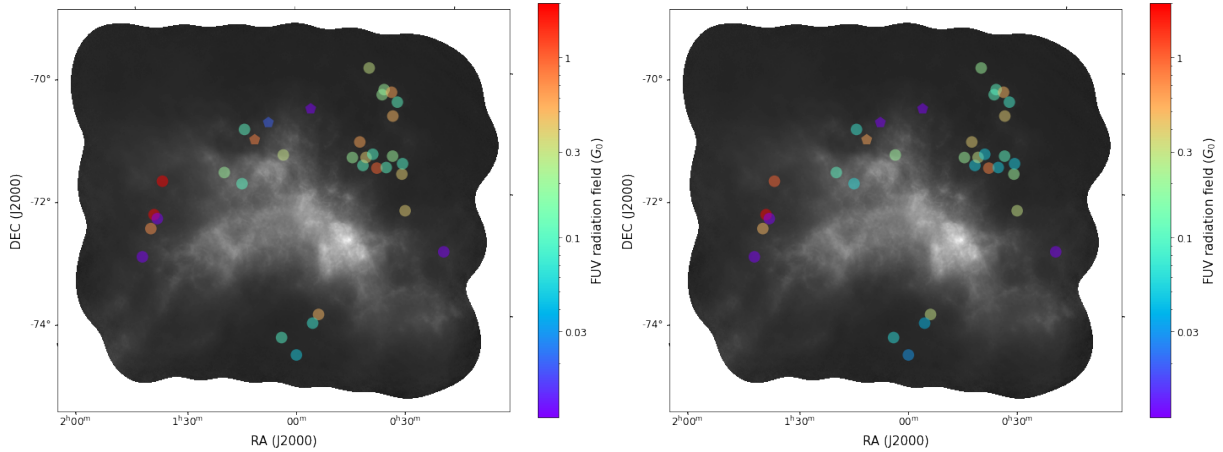


Figure 5.7: Distribution of the best fit values of all clumps (circles) and large clouds (pentagons) for G'_0 when $\mathcal{M} = 0$ (left) and when $\mathcal{M} = 1$ (right), overlaid on the total column density map of the SMC.

5.1.3 Caveats on ISM environment

There are some caveats to the phase diagram fitting that should be mentioned. In Table 5.1 it is clear that the uncertainties on some of the density and thermal pressure values are quite large, especially clump 2, which was explained in Section 4.4.2. The large errorbars on these measurements make it difficult to have a conclusively narrow phase diagram result because of the degeneracy of the solutions. The largest source of error when looking at the relative uncertainty is generally the uncertainty on the column density, but is often the uncertainty on the width.

This column density uncertainty could be refined by implementing a version of the bootstrapping method that accounts for the column density uncertainty on each component. This becomes difficult when fitting fields with a higher number of components, because there is no guarantee that the components will divide the signal in roughly the same way each time, it could be completely different each time. This is why the method used in Section 2.3 only measures the uncertainty in the specific phases of the HI.

The uncertainty in the width and the measurement of the widths in general is something that is limited by the resolution of this dataset. Within the Milky Way, where there is better physical resolution as the structures observed are much closer, the typical size of the cold filaments are less than a pc (McClure-Griffiths et al. 2006, Kalberla et al. 2016). A study similar to that in this work, with a smaller sample size of clouds in For et al. (2016) the measured angular sizes ranged between 0.028° and 0.144° which at the smallest distance to the clouds they consider, correspond to 12-63 pc respectively. This is more consistent with the sizes I find in this work. In their work they have a varying image resolution depending on the cloud region, so their smallest angular size of a cloud has a resolution of 0.056° , so they measured clump width is in-fact below their resolution limit. They are limited in the widths they can measure for their clumps due to the angular resolution and the large distance to the clouds, which combine to limit the physical resolution.

This same hurdle is encountered in this work. The angular resolution for this data is $30''$, which corresponds to

a physical size of 9pc at the distance of the SMC. The smallest width that I fit here is 12.2pc for clump 18, which is above the limiting resolution for this data. This is evidently larger than the typical widths found in the Milky Way, and it is possible that these structures have substructure that is not resolved by ASKAP. A study of molecular HCO+ filamentary structure in the LMC (Tokuda et al. 2023), shows filament widths of ~ 0.1 pc. Another study by the same team (Ohno et al. 2023), who measured the CO clumps in the north-east region of the SMC, found clumps with widths on the order of 1pc after using a deconvolving scheme with their beam that had a physical resolution of ~ 2 pc. Evidently the widths presented in this work could be an overestimation on the true widths of HI filamentary structure in the SMC. However, the only way to investigate this is to image the SMC at a higher resolution, something that will be possible with future telescopes.

5.2 Further chemical modelling

If the uncertainties in the density and the thermal pressure can be reduced, with higher-resolution imaging, it would be instructive to conduct more comprehensive chemical modelling, such as in Marchal et al. (2021), where the chemical modelling from Gong et al. (2017) was used for a handful of ISM conditions and compared to the different phases of the ISM. This modelling is a 1D PDR model, which can also provide the abundances of a number of chemical species. For example, in Rybarczyk et al. (2022) they compared PDR models with different ionisation rates and radiation field strengths, like what is done in this work, to observed HCO+ abundances in the diffuse ISM, coupled with the HI emission-absorption observation from Murray et al. (2015a). As discussed in Section 3.4 there are CO(2-1) observations for two of these clouds, however these were targeted because of the HI structure discovery in McClure-Griffiths et al. (2018). Molecular surveys typically target high column density areas of the SMC, and so there are fewer measurements of multiple species as we move away from the centre of the SMC. Now that I have catalogued the HI structure detected with the GASKAP pilot data, this supports a case for observing in these regions for other molecular tracers of dense gas.

5.3 All cloud properties

Here I compile the measured properties from all the clouds analysed in Chapters 3 and 4. Additionally, I calculate the deviation velocities (v_{DEV}) as:

$$v_{\text{DEV}} = v_{\text{peak}} - v_{M1} \quad (5.19)$$

where v_{peak} is the peak velocity in Table 5.3 and v_{M1} is the first moment velocity for the field from Figure 3.6. At this stage, the M_1 velocity map was recalculated after masking the emission from all the clumps as well as the large clouds. I show the distribution of deviation velocities across the SMC in Figure 5.8. All of these clouds have deviation

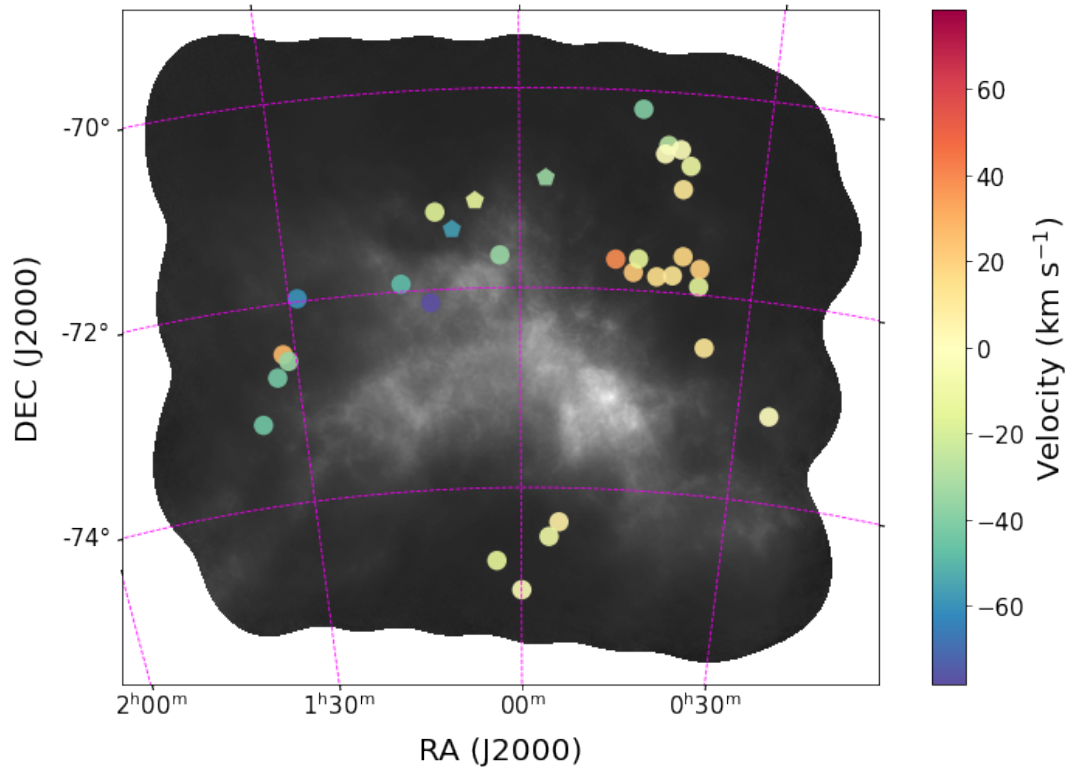


Figure 5.8: Distribution of the deviation velocities, indicated by the colourscale, overlaid on the total column density map of the SMC, shown in greyscale with white indicating maximum column density.

velocities that fall in the IVC classification, if using the definition for MW HVC/IVCs. Some clouds have deviation velocities near to zero, but these clouds are along low column density lines of sight, where the clump itself makes up a large part of the total column density, well separated from the main body of the SMC. So these very small deviation velocities should be considered with that in mind.

Cloud	RA centre (J2000)	DEC centre (J2000)	Peak velocity (km s ⁻¹)	v_{DEV} (km s ⁻¹)	$\overline{\sigma}_{\text{min}}$ (km s ⁻¹)	\overline{W} (pc)	σ_W (pc)	$N_{\text{H,max}} 10^{20}$ (cm ⁻²)	n (cm ⁻³)	σ_n (cm ⁻³)	$T_{k,\text{max}}$ (K)	σ_T (K)	P/k (K cm ⁻³)	σ_{PK} (K cm ⁻³)
Alpha	00:56:18	-70:50:43	106.4	-56.4	2.0	23.1	2.2	3.66	5.14	1.38	519	74	2666	811
Hook	01:09:21	-71:20:48	133.5	-13.8	1.3	33.7	3.6	1.83	1.76	0.48	246	37	432	135
Gamma	01:07:40	-70:51:21	110.7	-34.8	1.3	20.6	5.6	0.81	1.28	0.47	228	63	292	135
1	00:25:29	-73:08:16	124.9	0.0	1.5	66.4	13.9	2.92	1.09	0.32	298	39	326	105
2	00:35:23	-72:31:58	145.5	11.2	2.1	21.1	41.4	2.24	3.18	6.04	569	185	1811	3489
3	00:36:47	-71:44:39	142.5	20.6	1.8	21.3	6.2	1.53	2.33	0.79	292	53	679	261
4	00:36:50	-72:00:01	109.3	-17.1	2.2	25.5	3.1	2.69	3.42	0.91	492	130	1682	633
5	00:39:10	-70:43:27	120.1	-10.8	2.2	21.3	2.1	1.39	2.11	0.56	565	44	1193	329
6	00:39:16	-71:36:56	145.5	18.3	1.9	14.0	2.0	1.27	2.93	0.80	390	70	1143	372
7	00:39:54	-70:51:49	138.6	11.6	3.0	18.5	1.9	1.86	3.25	0.86	837	210	2717	992
8	00:40:11	-71:48:27	136.7	11.9	1.4	18.1	2.1	1.98	3.54	0.94	201	23	713	207
9	00:40:29	-70:34:29	132.8	-0.8	2.3	15.4	1.7	1.61	3.39	0.90	835	202	2831	1018
10	00:41:34	-70:37:13	129.8	-0.2	1.5	22.7	3.6	2.29	3.26	0.90	327	42	1066	325
11	00:41:56	-70:31:09	108.3	-28.0	2.0	23.2	4.3	2.03	2.84	0.81	470	71	1336	430
12	00:42:36	-71:51:08	142.5	15.0	3.7	24.0	3.4	2.93	3.96	1.07	1351	304	5349	1885
13	00:43:00	-71:33:29	159.2	32.0	1.6	26.4	12.8	2.88	2.69	0.75	297	75	799	300
14	00:44:59	-71:41:25	118.1	-14.7	1.9	39.0	12.9	4.61	3.83	1.40	467	72	1790	710
15	00:45:10	-70:10:53	109.3	-39.8	2.3	17.7	2.7	1.52	2.77	0.76	570	85	1580	492
16	00:45:32	-71:50:00	155.2	22.3	1.4	15.3	2.6	1.56	3.29	0.92	216	33	712	227
17	00:47:10	-71:25:09	163.1	19.3	1.7	15.7	1.4	2.34	4.83	1.28	468	54	2261	652
18	00:47:55	-71:41:19	180.6	38.6	2.1	12.2	1.9	1.05	2.79	0.77	365	80	1018	359
19	00:53:11	-74:26:09	157.2	7.9	2.5	13.9	4.6	3.09	3.39	0.93	711	103	2411	750
20	00:56:35	-74:28:21	141.6	-10.8	1.9	26.4	12.2	2.92	1.69	0.50	502	71	850	280
21	00:59:57	-75:01:32	153.0	-2.5	1.3	23.0	6.1	1.35	1.90	0.61	361	140	686	347
22	01:02:38	-71:41:02	122.0	-34.9	1.9	19.0	1.9	1.87	3.18	0.84	443	23	1410	380
23	01:03:26	-74:46:32	142.5	-14.4	1.9	29.5	6.0	1.72	1.89	0.55	541	51	1024	315
24	01:11:14	-71:15:54	143.5	-16.6	1.9	20.1	2.6	1.68	2.72	0.73	328	36	892	259
25	01:11:41	-72:07:48	91.7	-78.3	1.9	18.5	2.4	1.18	2.06	0.55	458	35	943	264
26	01:15:59	-71:56:28	135.7	-48.8	1.6	26.3	4.4	2.72	3.34	0.93	333	72	1112	393
27	01:28:51	-71:57:58	126.9	-59.9	2.9	13.6	1.6	2.30	5.49	1.47	907	146	4976	1554
28	01:31:00	-72:34:20	135.7	-35.1	2.0	20.2	2.0	0.87	1.40	0.37	378	43	527	152
29	01:31:45	-72:30:28	144.5	-27.7	5.2	24.0	2.7	2.68	3.63	0.97	3388	189	12296	3343
30	01:32:40	-72:42:29	128.9	-43.4	3.5	24.6	8.3	2.28	3.00	1.10	1522	233	4561	1818
31	01:35:45	-73:10:33	126.9	-44.6	1.3	28.4	3.7	1.26	1.44	0.39	212	21	305	88

Table 5.3: Properties of all clouds, compiled from Tables 2.1, 3.1, 4.1, 4.2, 5.2, 5.1. Columns: (1) Cloud name/number, (2) Field centre Right Ascension, (3) Field centre Declination, (4) Peak velocity of cloud, (5) Deviation velocity of cloud from the SMC, (6) Mean velocity dispersion of coldest component fit with ROHSA, (7) Width of cloud as fit in Section 4.4.2, (8) Uncertainty in the width, (9) Maximum column density of coldest component fit with ROHSA, (10) Number density as calculated in Section 4.4.2, (11) Uncertainty in the number density, (12) Maximum kinetic temperature of coldest component fit with ROHSA from 3.1, (13) Uncertainty in the maximum kinetic temperature, (14) Thermal pressure from 5.16, (15) Uncertainty in the thermal pressure. Peak velocities for the large clouds are calculated as the closest velocity channel to the mean of the velocities reported for each component related to each cloud in 3.1. Deviation velocities (v_{DEV}) are defined by Equation 5.19.

Chapter 6

Conclusions and perspectives

6.1 Conclusions

The periphery of the SMC is rich in small scale structure. In this work I have analysed three previously known clouds and identified 31 smaller clouds with new high resolution ASKAP data from the GASKAP survey (Pingel et al. 2022). Gaussian decomposition, a powerful tool for examining the dynamics and phases of the HI, was conducted with ROHSA to analyse this data. ROHSA, a regularised Gaussian decomposition algorithm, was fine-tuned to the specifics of this data to reach stable, spatially-coherent solutions of the HI spectra across the field. The velocity and phases of the HI were extracted from these solutions and analysed in relation to each other. This is the first time ROHSA has been used to trace the multiphase medium in isolated clouds, rather than extended HI along large velocity ranges. In analysing the phases of each large cloud, I find the CNM fraction for the Alpha and Gamma clouds is highest towards the SMC and decreases away from the SMC, with the Gamma cloud completely CNM-dominated at its southern-most point. For the Hook cloud I find a higher CNM fraction along the western edge, reaching values of ~ 0.5 . The Alpha and Gamma clouds both exhibit velocity gradients, becoming more blue-shifted at increasing distance from the SMC. When analysing the velocities of the separate phases of the Alpha cloud, this velocity trend is dominated by the CNM component, while the WNM component exhibits no coherent velocity gradient along the same direction. There is also a small offset in the velocity of the CNM and WNM (a few km s^{-1}) for the Alpha cloud at the closest point to the SMC. There is a hint of this offset in the Gamma cloud as well, but the CNM and WNM do not overlap spatially at high enough column densities to be confident in this trend.

These results indicate that the morphology of all three clouds have been shaped by compressive forces. The Hook cloud seems not to have been shaped by a force coming from the direction of the main body of the SMC, as it has retained its WNM envelope around the southern base. Its filamentary CNM structure is more consistent with an expansive force from the centre of this field that has swept up and compressed the HI into a shell-like structure. The Alpha and Gamma clouds look to be experiencing a wind force that has stripped the WNM away from the denser

CNM cores, creating elongation away from the SMC and a velocity offset between the CNM and WNM. This indicates that the periphery structure of the SMC is strongly influenced by dynamic processes in SMC itself and not just the tidal interaction between the LMC and SMC. The tidal interaction is thought to have created the Magellanic Bridge and Stream, which pulled large amounts of gas out of the SMC, but evidently the existence of these clouds suggest that the SMC is pushing out its own gas as well. Follow-up observations of the $H\alpha$ south of the Alpha and Gamma clouds will help constrain the exact origins of the compressive force.

This work has also uncovered a number of small, dense and mostly cold clumps around the edges of the SMC. I catalogue 31 clouds through a spatial dispersion method that used clustering to identify outlying structures. I use the same method on recently observed field for the LMC, Bridge and Stream and find fewer similar clumps. The SMC clumps all have CNM number densities of $1-6 \text{ cm}^{-3}$ consistent with the lower end of the CNM density range. These densities may be limited by the angular resolution of the data, as structures can only be resolved down to the scale of 9pc in the SMC, whereas cold filaments in the Milky Way have widths of $\sim 0.1 \text{ pc}$. I used the densities, in conjunction with the maximum kinetic temperatures derived from the fit linewidths, to estimate the properties of the ISM using the main ISM heating and cooling equations. The results of this showed a wide variation in the FUV radiation field strength around the SMC, which may be indicative of local enhancements in the radiation strength due to stellar structure in periphery, or the fact that a group of clumps, while appearing to be close in this projection are actually separated by a large distance along the line of sight. This would mean the changes in radiation field strength from clump to clump may not be happening over as small of distances as they appear to be, due to projection effects.

In summation, I have now improved our understanding of the phase composition of three large IVCs at the north-eastern edge of the SMC. I have catalogued and characterised a further 31 clumps of HI around the SMC and used all the clouds to estimate the relative radiation field strength distribution in the SMC periphery.

6.2 Perspectives

The work conducted in this thesis has uncovered a wealth of structure around the SMC which should be followed up on. It shows that more and more structures become apparent in the Magellanic System as successive HI surveys improve in resolution and sensitivity. The GASKAP survey is currently in the pilot data collection phase (Pingel et al. 2022) and the data used in this work is from a 20hr integration of the SMC. The full survey imaging for the SMC will have an integration time of 200hr (Dickey et al. 2013) and achieve a theoretical noise level of 0.48 K in each 1 km s^{-1} channel at 30" resolution, which is 3 times lower than the median value for the noise I find in this work. Observations at higher sensitivity give the ability to confidently resolve fainter gas clouds, and could lead to more HVC/IVC detections around the SMC. Assuming a linewidth of 2 km s^{-1} , an upper limit for the CNM, at this sensitivity, gives a 3σ detection of column densities down to $1.3 \times 10^{19} \text{ cm}^{-2}$. The ability of Gaussian decomposition to trace distinct, multiphase structures along neighbouring lines of sight, could be applied to large section of the SMC. This

may help in building a picture, in conjunction with distance measurements from stellar objects, of the 3D structure of the HI phases along the elongated depth of the SMC. This would allow for more direct comparisons with the models proposed to explain the evolution of the Magellanic System over its lifetime. The 200hr integration time will also be afforded to the LMC. Another search for small scale structure can be done on those fields to investigate the HVC/IVC population of the SMC's companion.

ASKAP is a pathfinder for the Square Kilometre Array (SKA). The mid-frequency component of the SKA (SKA-MID) will be well equipped to observe the HI 21cm line and has an anticipated angular resolution of 0.4" at this wavelength (Braun et al. 2019), which is around a factor of 100 better than ASKAP's resolution of 30". This will allow for the resolution of structures in the SMC closer to the typical cold filament width of 0.1pc. With this spatial resolution, a full characterisation of the small scale structures will be possible for not just the SMC, but the whole Magellanic system. This will also allow for a more accurate estimate of number density in cold clouds. SKA-MID will drastically improve our understanding of how the HI is being transported around the Magellanic system.

The work done in this thesis also builds the foundation for regularised Gaussian decomposition of large survey data. The amount of data we will have will only increase between now with ASKAP and 5-10 years in the future with the SKA. Development of these techniques on small fields of the data, with the Gaussian decomposition tools currently available, lays the groundwork for using them on larger datasets.

CO observations towards the Gamma cloud and all the other small clumps, as well as the full extent of the Hook cloud will give a more complete picture of the relationship between the CO and HI in small cold clumps. The CO is not necessarily the best tracer of molecular hydrogen at low metallicities, so observations of OH or HCO+ may aid in uncovering the CO-dark molecular hydrogen in these regions as well. Follow-up CO observations could be achieved with the APEX telescope, which was used to obtain the initial CO observations of Di Teodoro et al. (2019a).

H α spectroscopy towards the base of the Gamma and Alpha clouds will be able to constrain the radial velocity of the H α emission seen near these clouds. If there are associations found, these could be used to further understand the time-scales and processes acting of the HI. On the other hand, if there is no radial velocity association, then the origin of the force acting on these clouds must be investigated further. Further constraints on the stellar population associated with the H α emission identified for the Gamma cloud are not currently available with the most recent Gaia data release (Vallenari et al. 2023), however future data releases may have radial velocity information that could confirm or rule out the relationship of this cluster to the Gamma cloud.

Finally, numerical modelling of the clouds analysed in this work will help to better understand their origin. Modelling the the different scenarios of infall through an SMC halo or ram-pressure stripping of a cloud in the SMC environment will help understand the origin of these clouds. This will require a dedicated suite of HVC simulations in the multiphase ISM specific to the SMC.

Appendix A

Column density maps of all small clouds

Here I provide the column density maps and mean spectra for each clump identified in Chapter 4. The column density maps are integrated over the spectral range of the cloud, indicated by the pink band in each spectrum. The mean spectra are weighted by the column density over the range indicated, that is, the clump column density.

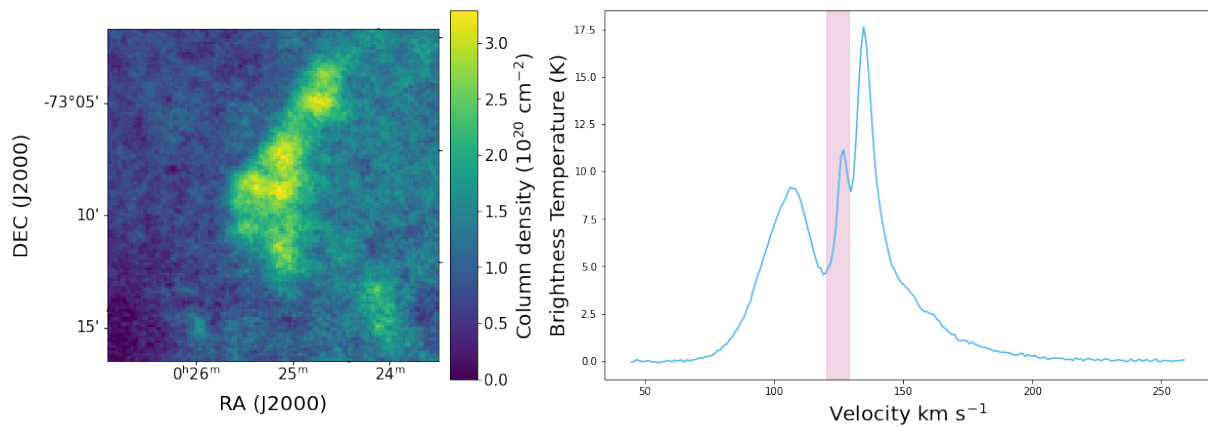


Figure A.1: Left: Integrated column density of Clump 1. Right: Column density weighted, mean spectrum across the field of Clump 1. Pink window indicates the velocity range the column density is integrated over.

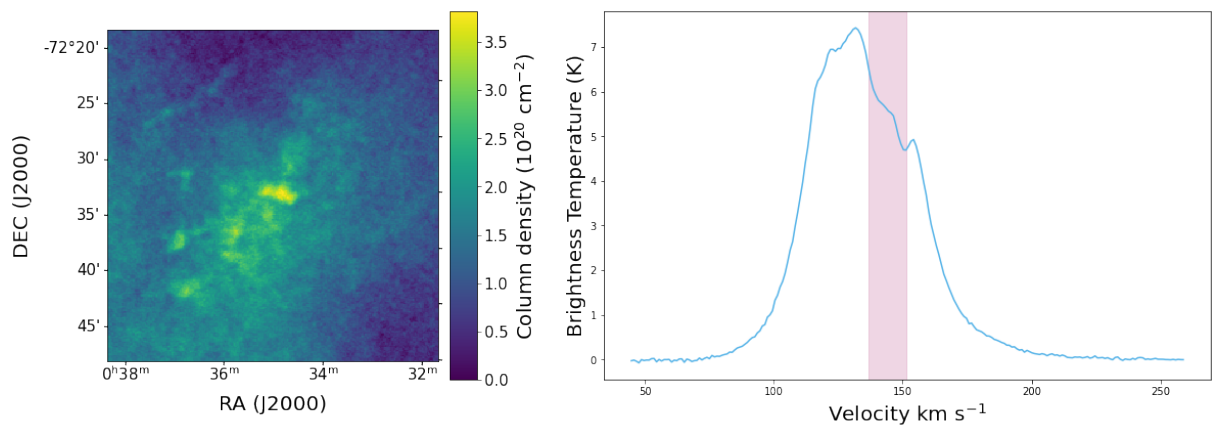


Figure A.2: Same as Figure A.1 but for Clump 2.

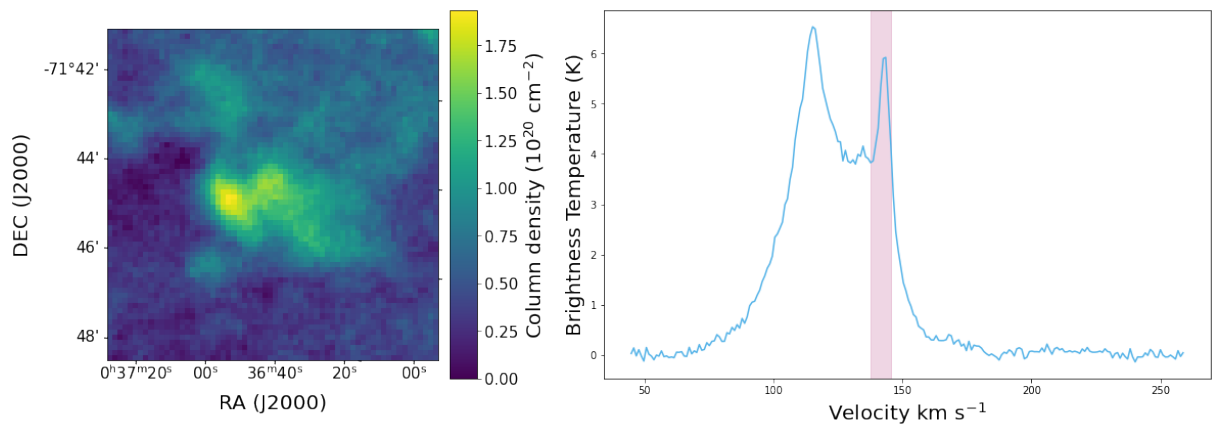


Figure A.3: Same as Figure A.1 but for Clump 3.

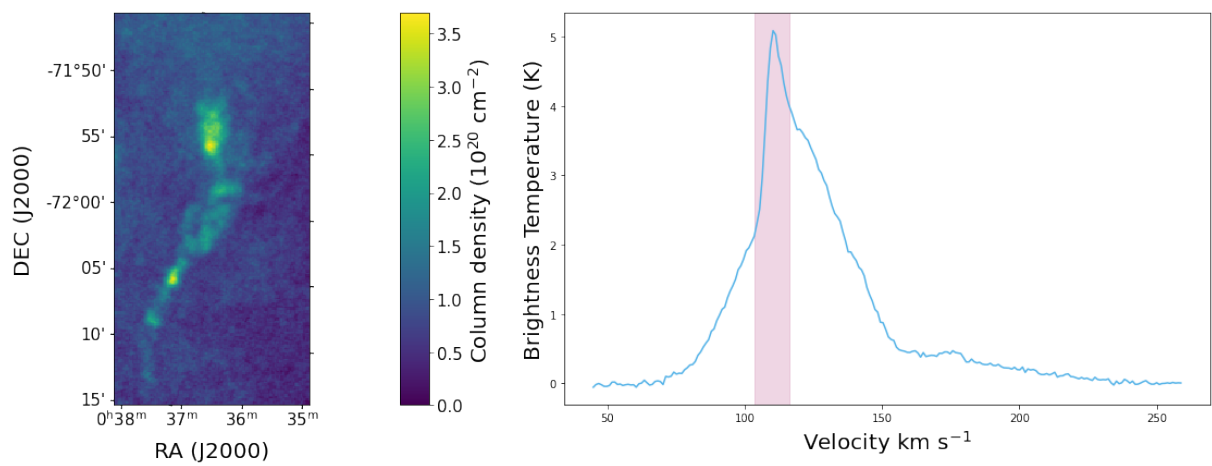


Figure A.4: Same as Figure A.1 but for Clump 4.

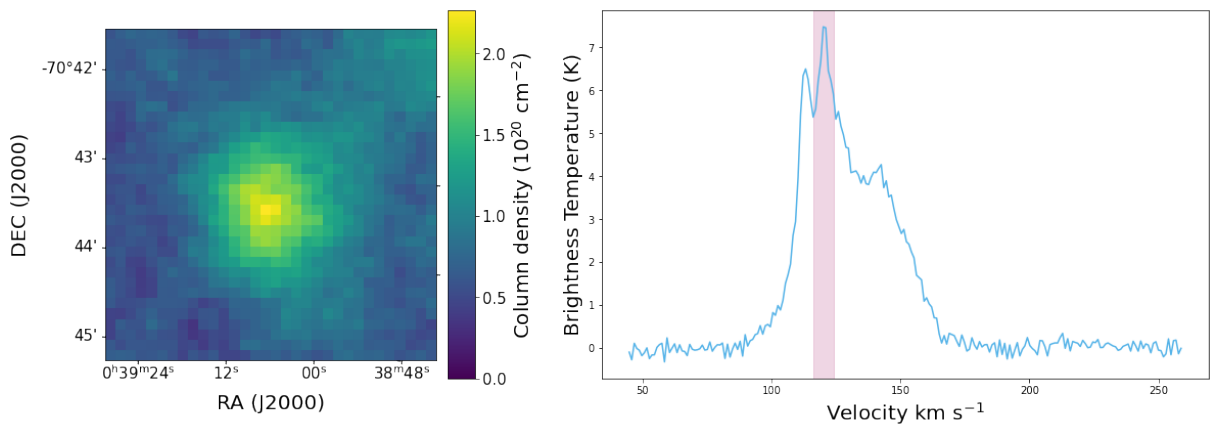


Figure A.5: Same as Figure A.1 but for Clump 5.

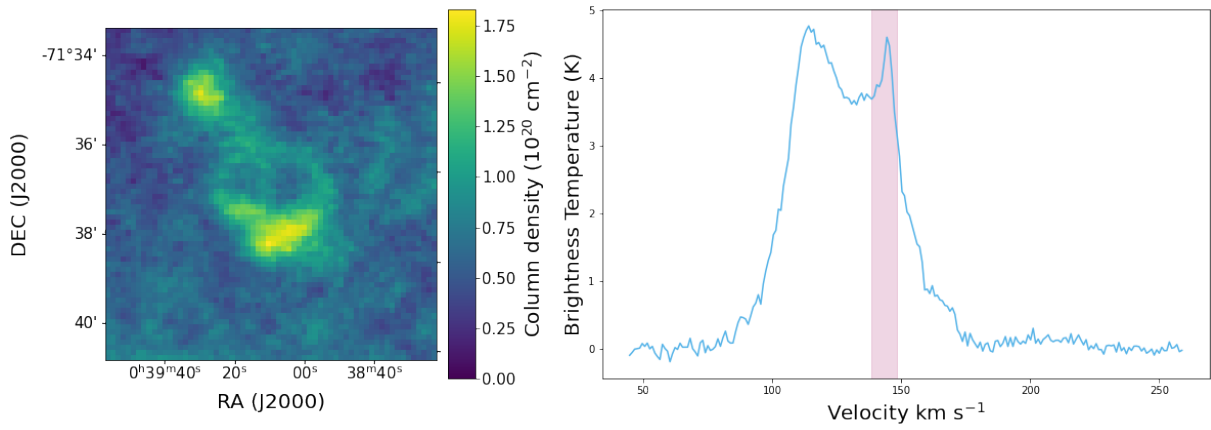


Figure A.6: Same as Figure A.1 but for Clump 6.

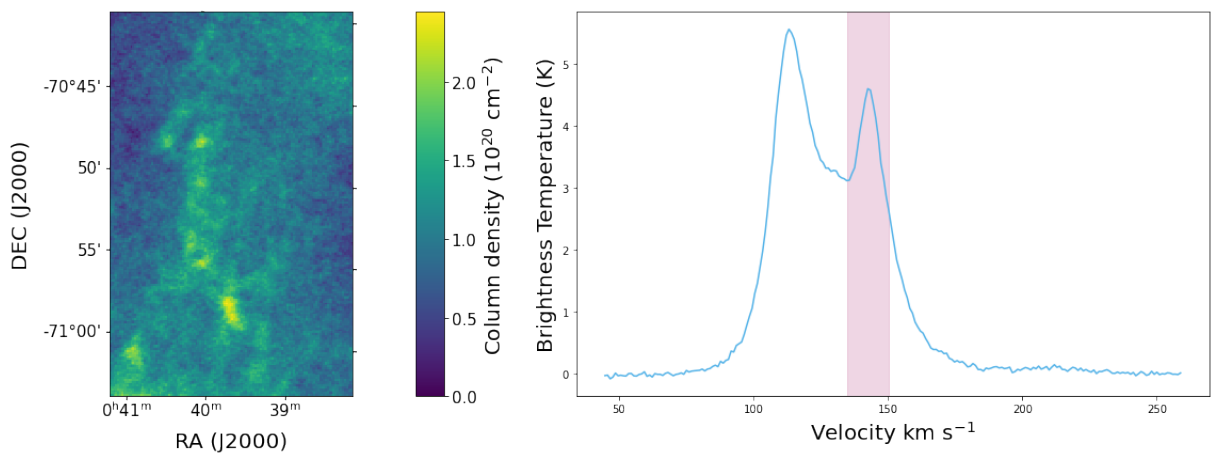


Figure A.7: Same as Figure A.1 but for Clump 7.

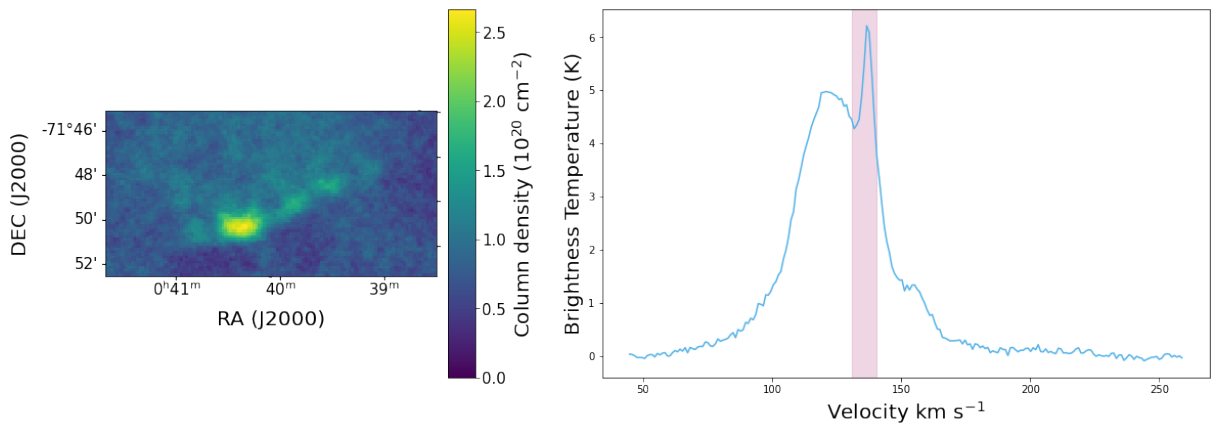


Figure A.8: Same as Figure A.1 but for Clump 8.

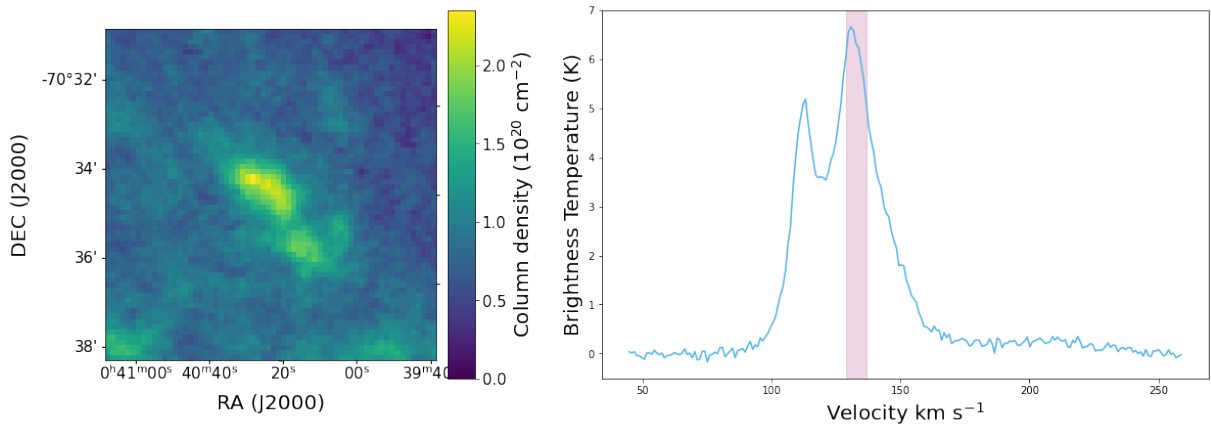


Figure A.9: Same as Figure A.1 but for Clump 9.

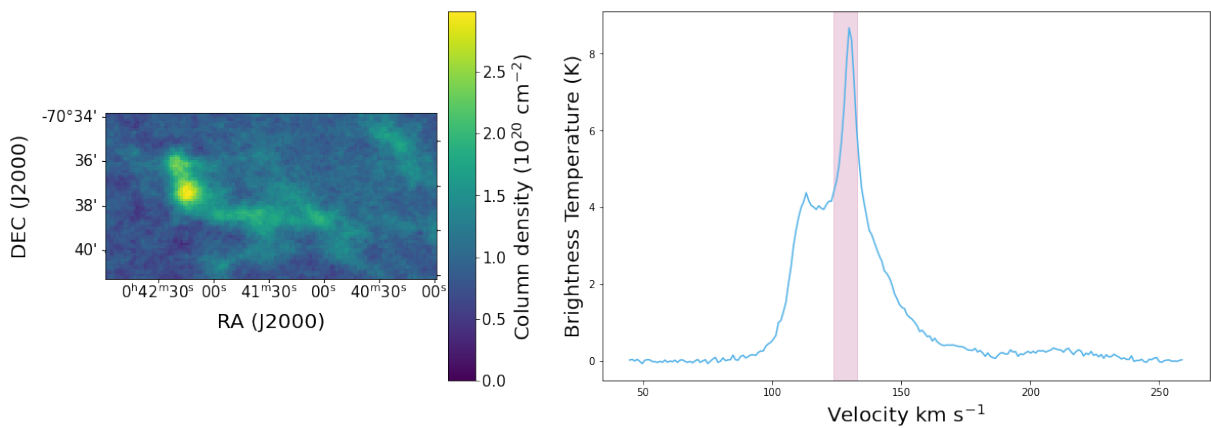


Figure A.10: Same as Figure A.1 but for Clump 10.

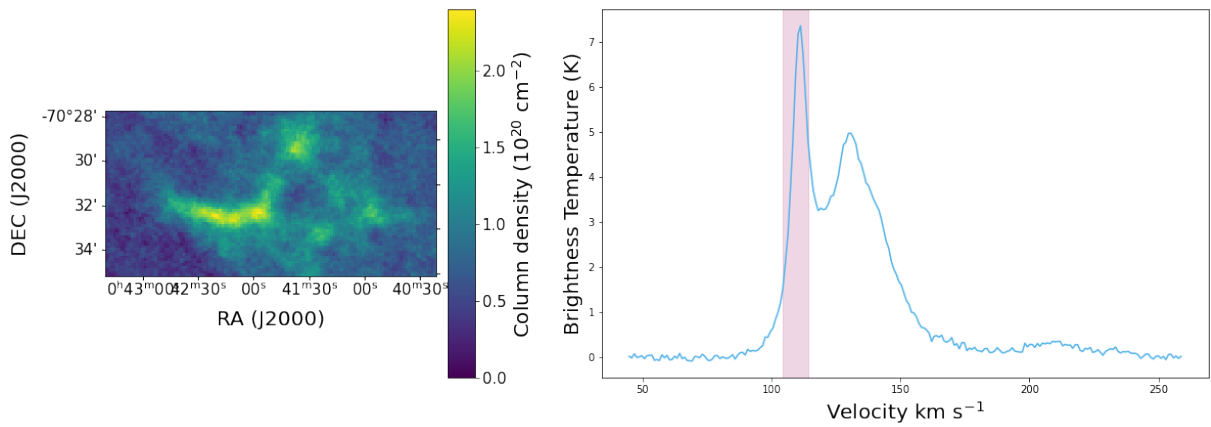


Figure A.11: Same as Figure A.1 but for Clump 11.

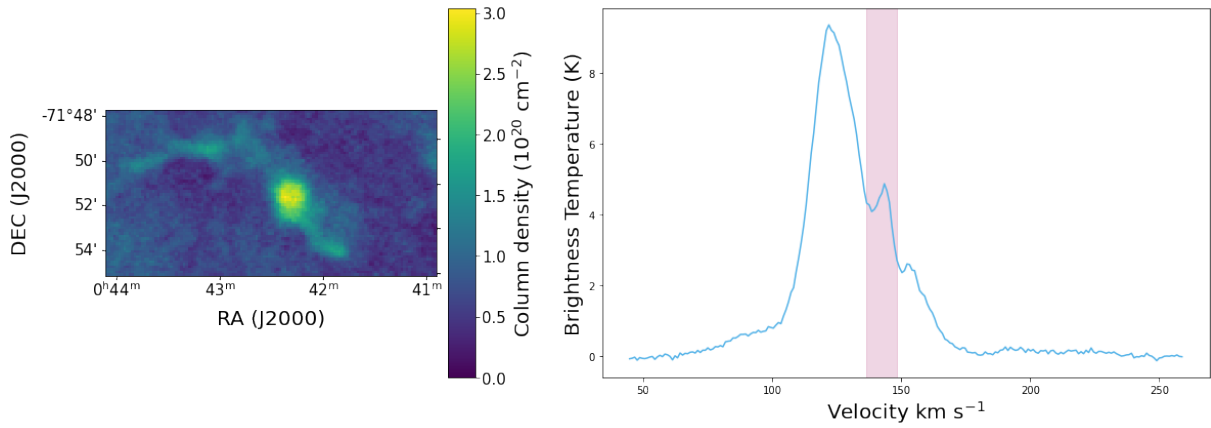


Figure A.12: Same as Figure A.1 but for Clump 12.

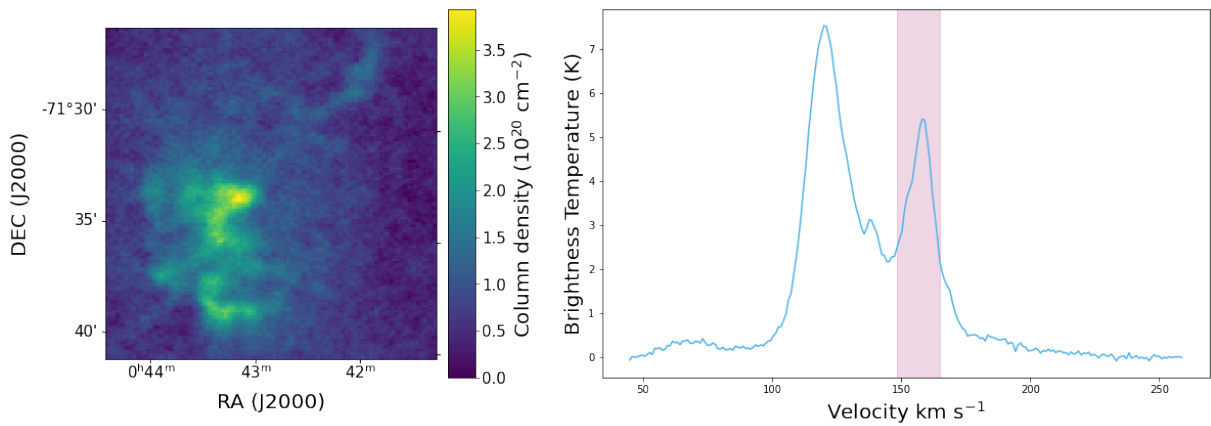


Figure A.13: Same as Figure A.1 but for Clump 13.

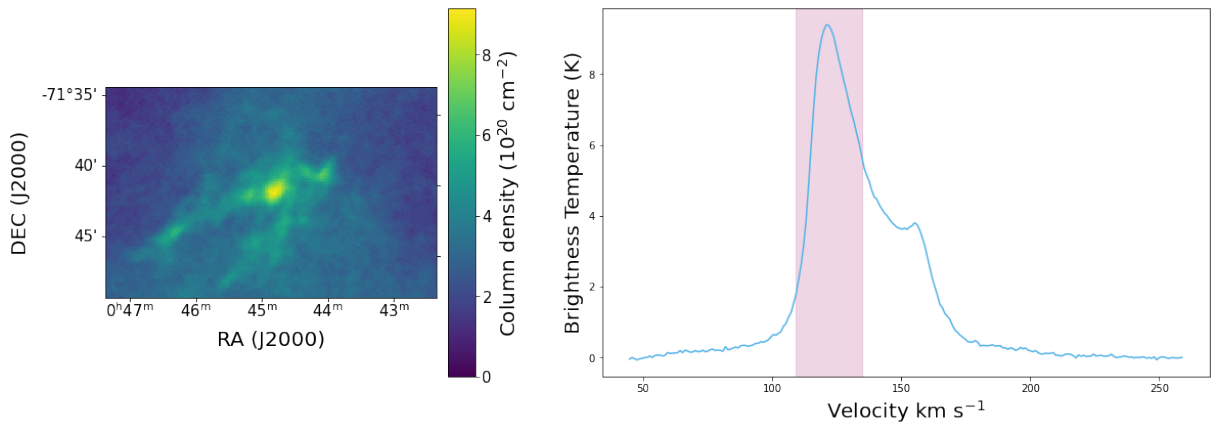


Figure A.14: Same as Figure A.1 but for Clump 14.

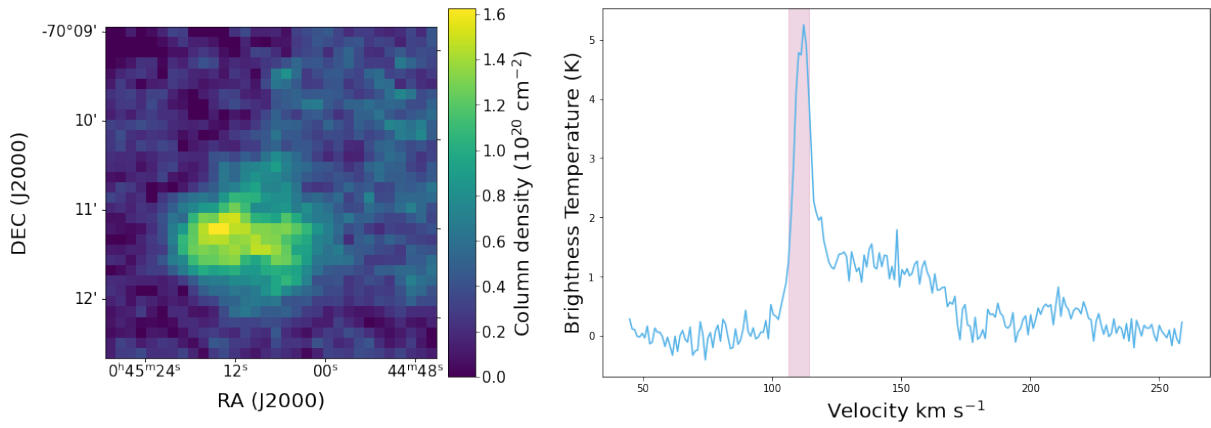


Figure A.15: Same as Figure A.1 but for Clump 15.

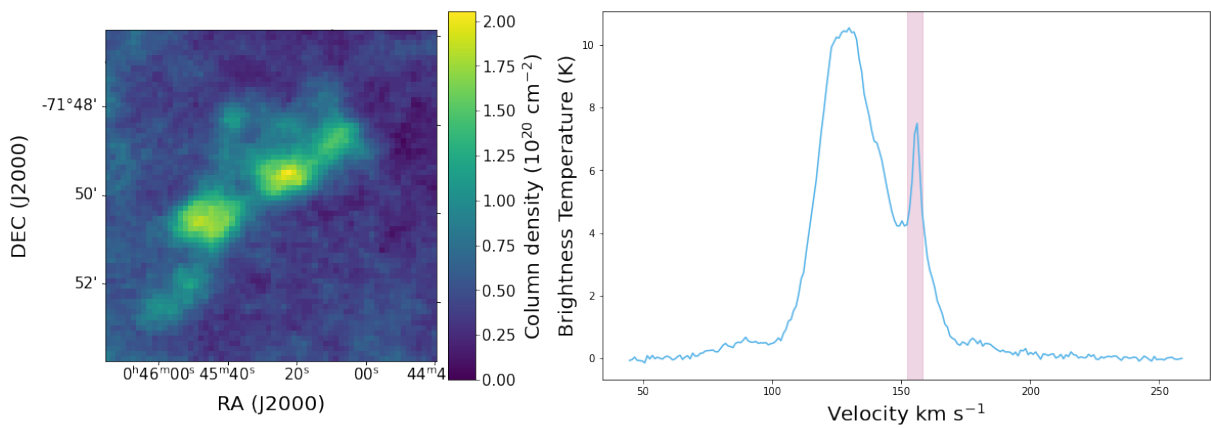


Figure A.16: Same as Figure A.1 but for Clump 16.

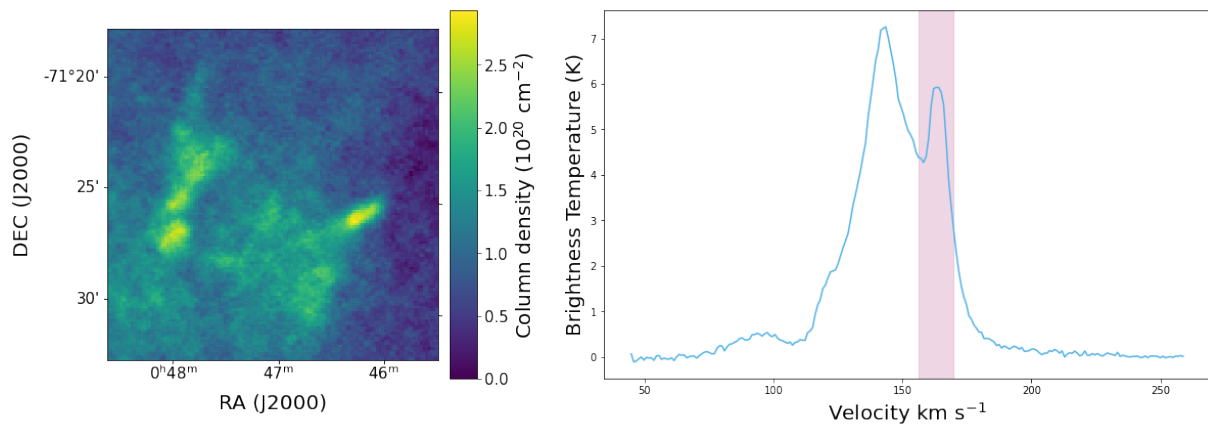


Figure A.17: Same as Figure A.1 but for Clump 17.

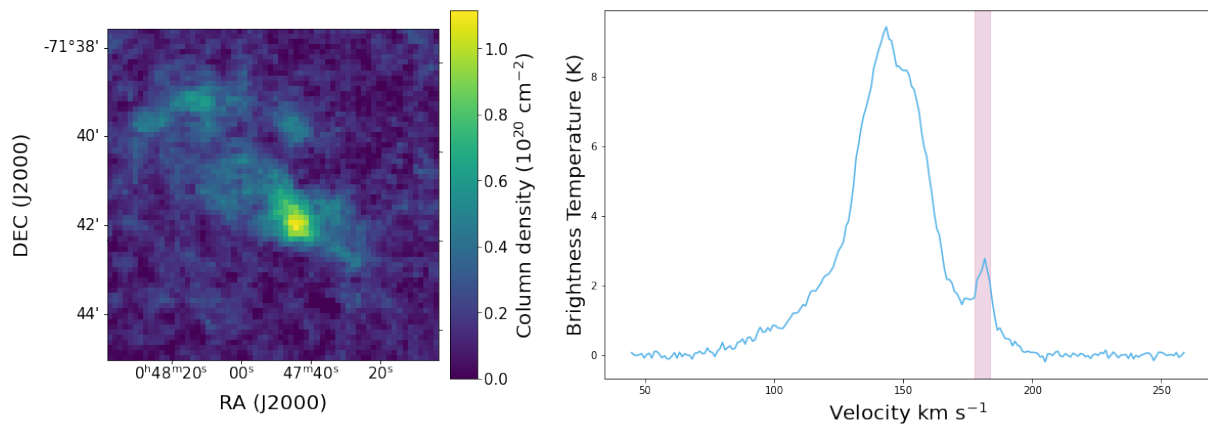


Figure A.18: Same as Figure A.1 but for Clump 18.

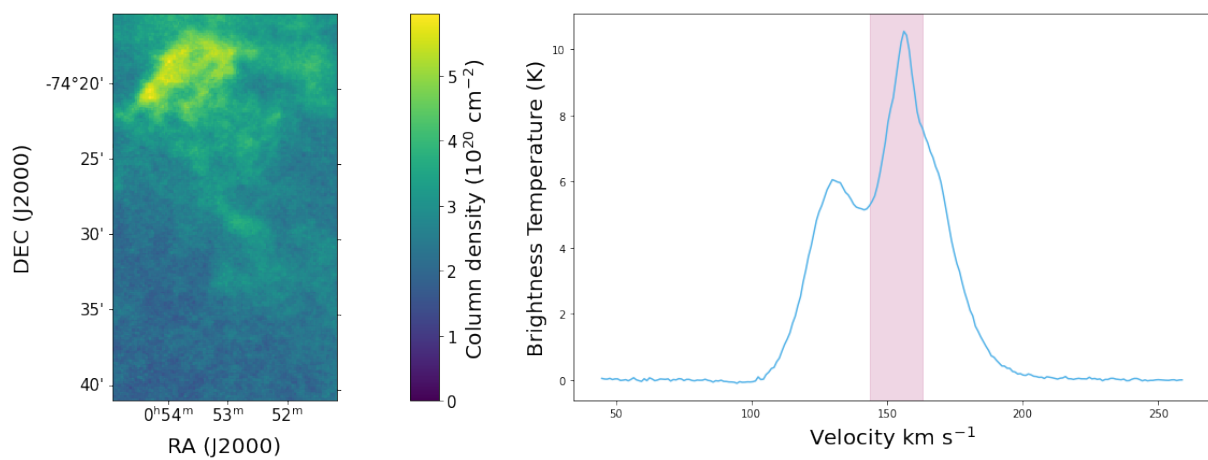


Figure A.19: Same as Figure A.1 but for Clump 19.

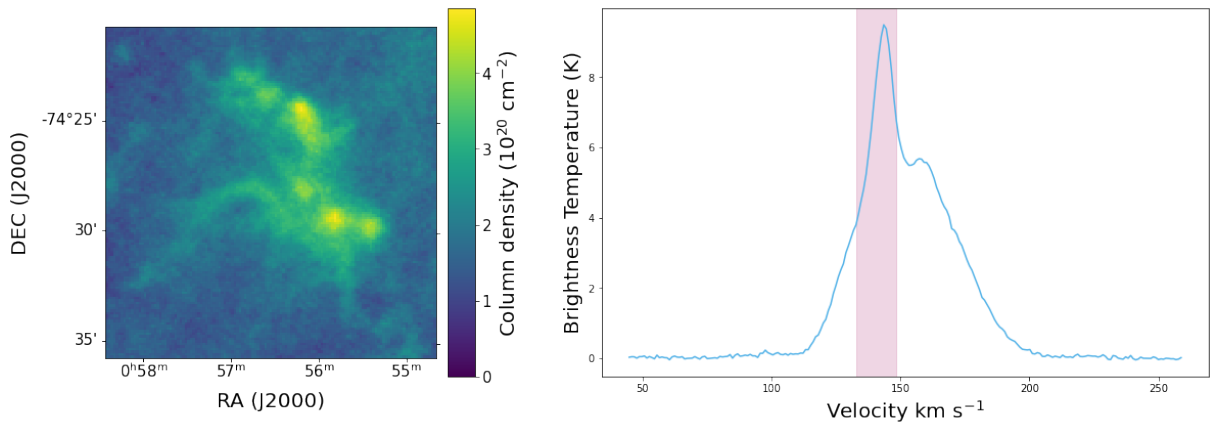


Figure A.20: Same as Figure A.1 but for Clump 20.

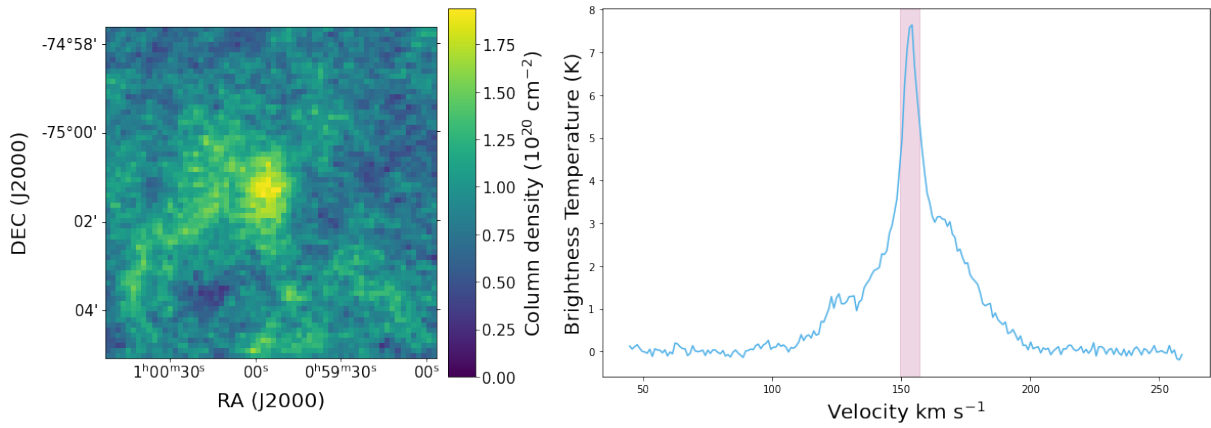


Figure A.21: Same as Figure A.1 but for Clump 21.

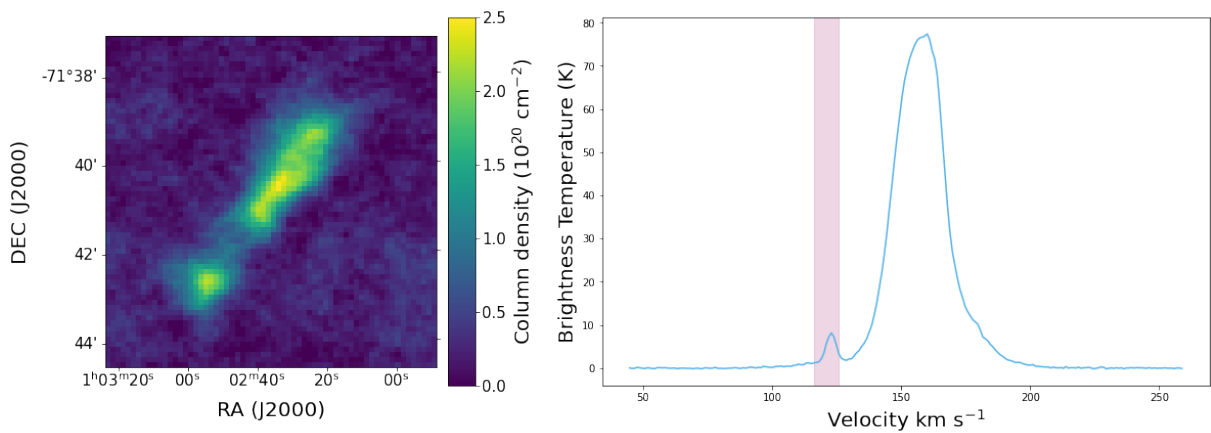


Figure A.22: Same as Figure A.1 but for Clump 22.

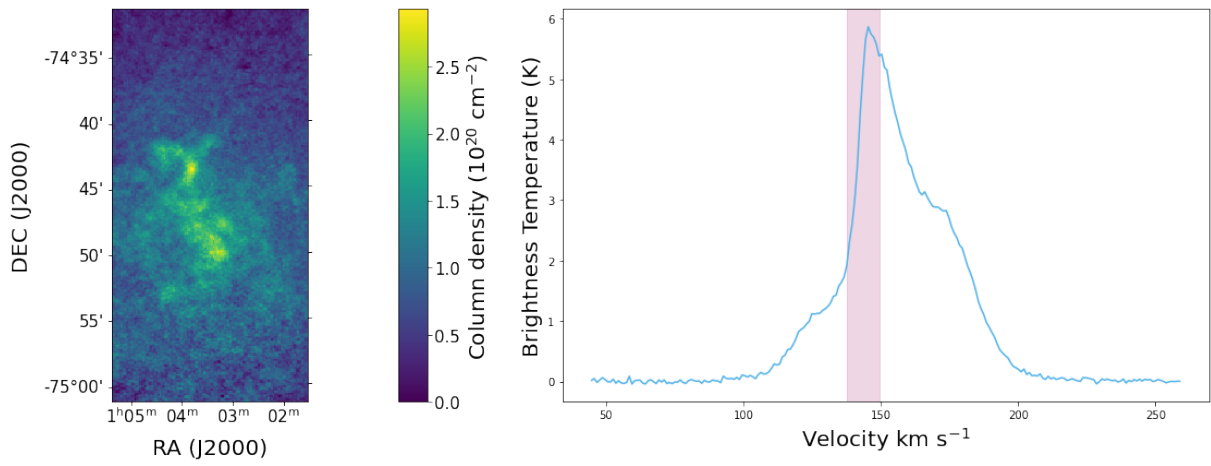


Figure A.23: Same as Figure A.1 but for Clump 23.

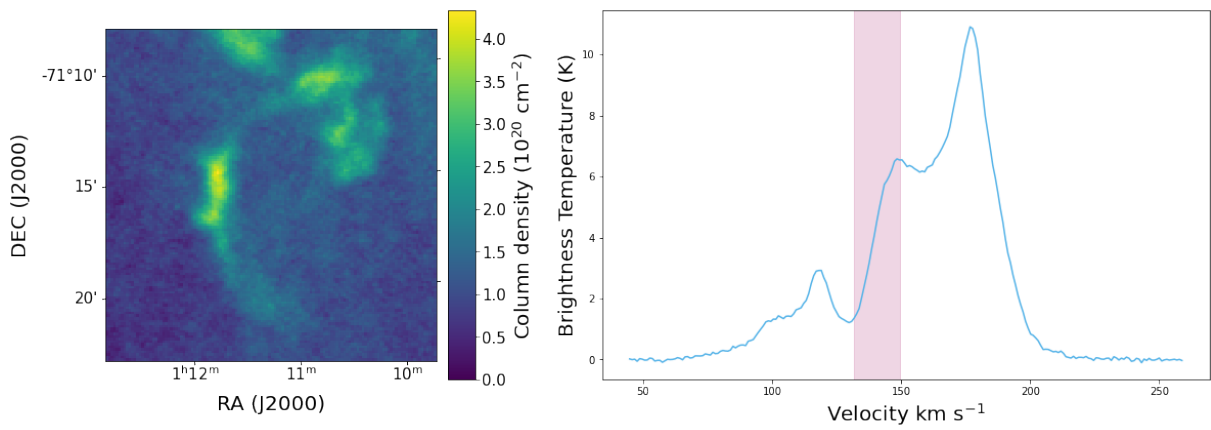


Figure A.24: Same as Figure A.1 but for Clump 24.

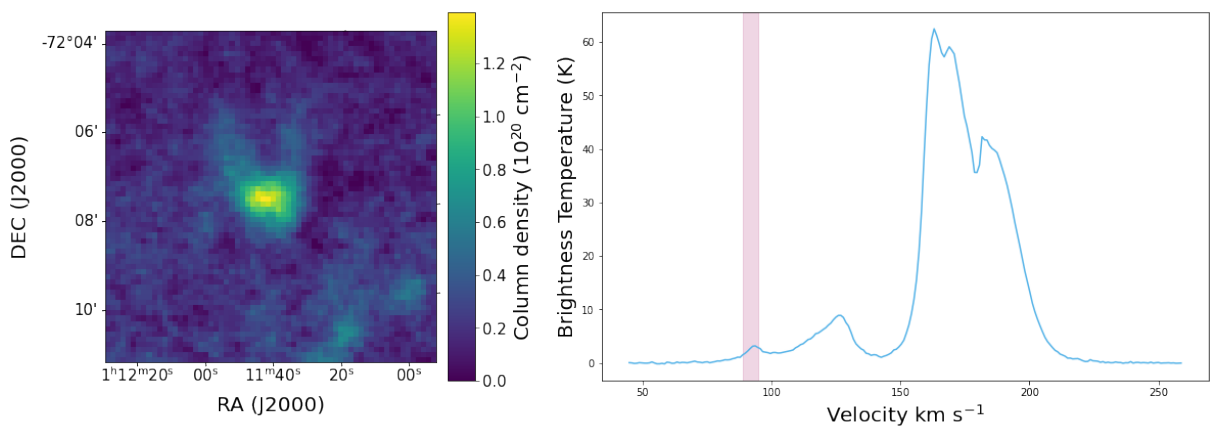


Figure A.25: Same as Figure A.1 but for Clump 25.

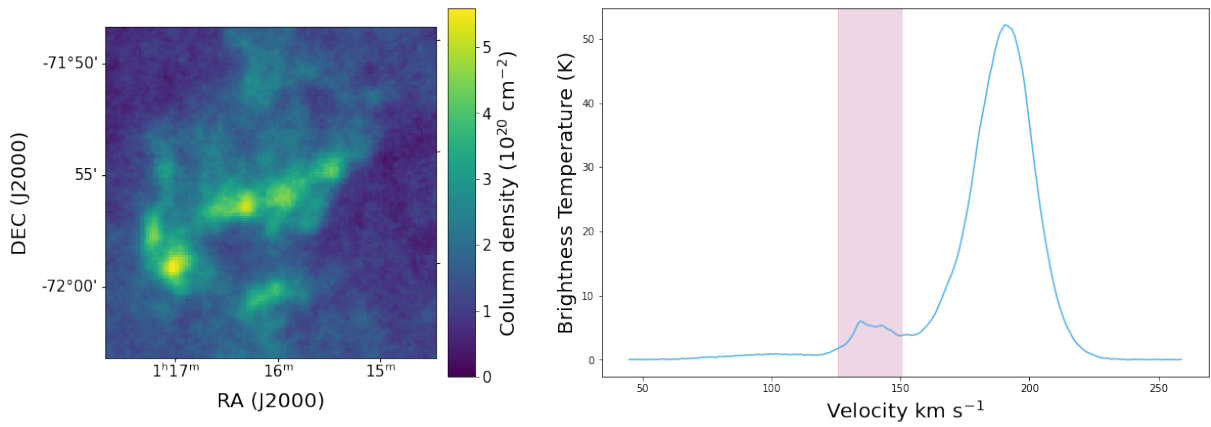


Figure A.26: Same as Figure A.1 but for Clump 26.

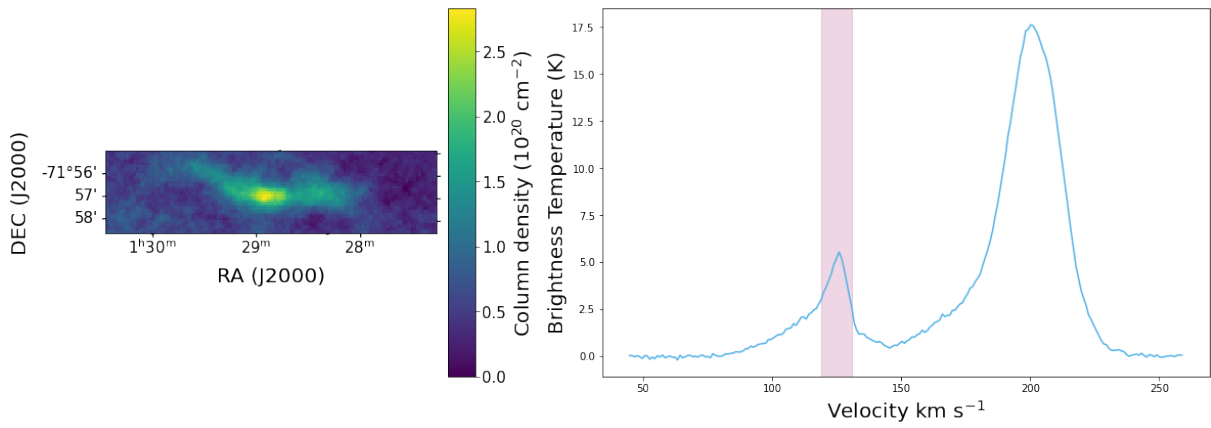


Figure A.27: Same as Figure A.1 but for Clump 27.

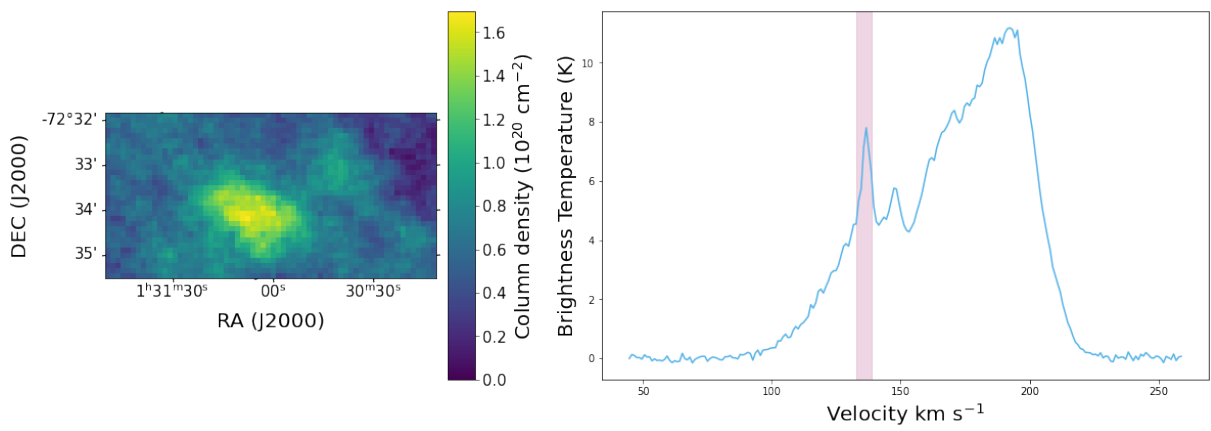


Figure A.28: Same as Figure A.1 but for Clump 28.

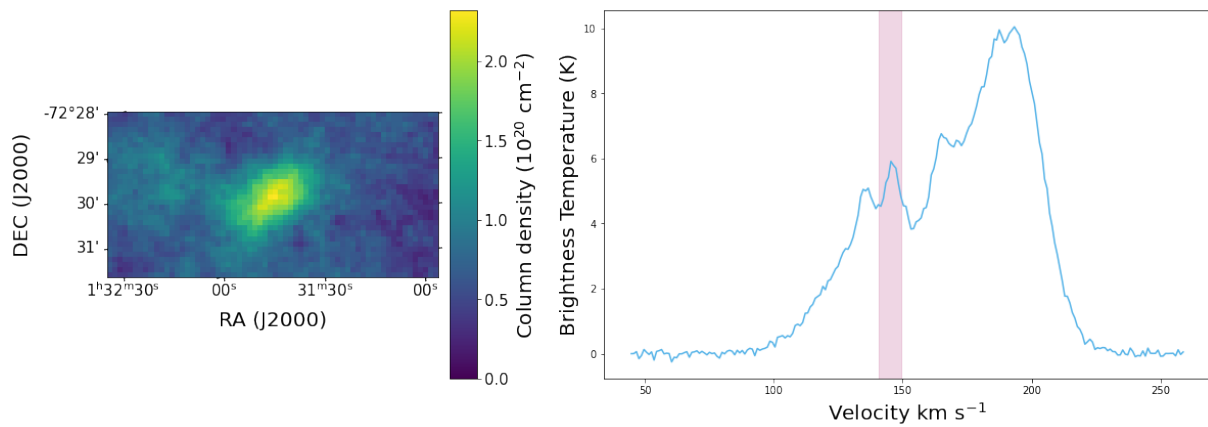


Figure A.29: Same as Figure A.1 but for Clump 29.

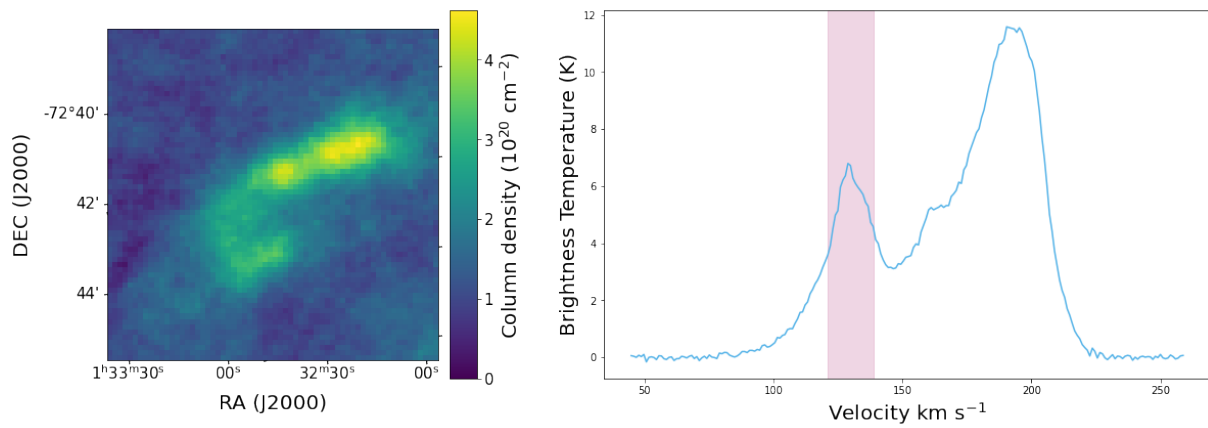


Figure A.30: Same as Figure A.1 but for Clump 30.

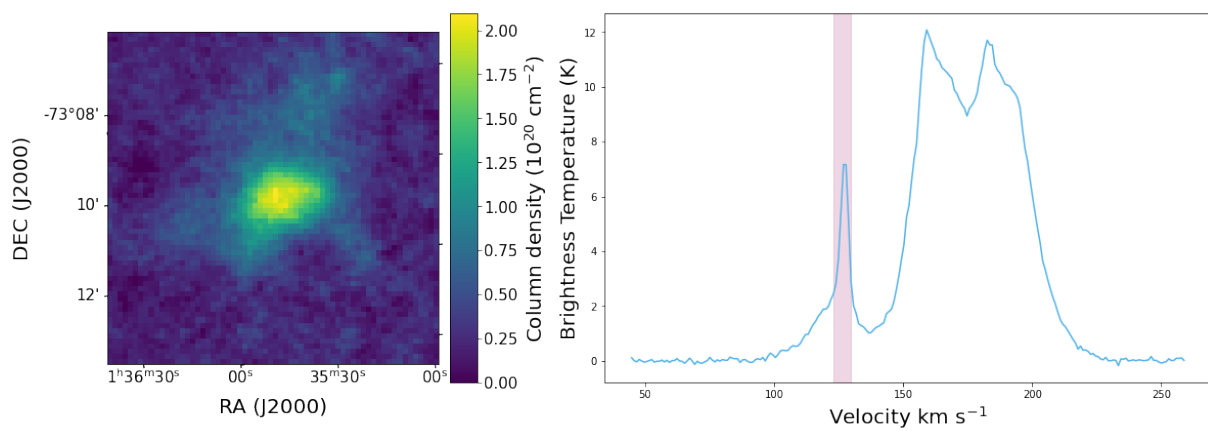


Figure A.31: Same as Figure A.1 but for Clump 31.

Résumé (en Français)

Dans ce travail, je présente les résultats d'une analyse des spectres d'émission HI du Petit Nuage de Magellan (SMC) avec le nouvel interféromètre SKA pathfinder, l'Australian Square Kilometre Array Pathfinder (ASKAP).

Cette thèse porte sur l'analyse des données spectrales d'émission HI du Petit Nuage de Magellan (SMC) obtenues avec le nouveau télescope interférométrique SKA pathfinder, l'Australian Square Kilometre Array Pathfinder (ASKAP). Ces données ont été obtenues dans le cadre de la collaboration Galactic ASKAP (GASKAP) qui mène une vaste étude de l'émission de la raie à 21 cm du HI à travers le système de Magellan et le plan galactique. Cette étude fournira à la communauté les observations d'émission HI à la plus haute résolution de ces régions du ciel à ce jour. Le SMC a été l'une des premières régions traitées par l'équipe de l'étude et constitue la base de cette thèse.

La raie à 21 cm est cruciale pour tracer le gaz d'hydrogène neutre à travers les galaxies. L'émission de cette raie provient du gaz HI à différentes températures. Typiquement, le HI est catégorisé en tant que milieu neutre chaud (WNM) ou milieu neutre froid (CNM) en fonction de la température mesurée du gaz. La décomposition spectrale de la raie d'émission de HI permet de distinguer les profils d'émission provenant de chacune de ces phases. Les profils d'absorption peuvent également être utilisés pour analyser le contenu en CNM le long de lignes de visée spécifiques. Les études d'absorption précédentes du SMC ont trouvé des fractions variables de CNM par rapport à WNM, 13

Le premier projet principal de ce travail a porté sur trois grandes structures au nord de la galaxie. Ces structures ont été identifiées pour la première fois comme des flux sortants potentiels dans les données de mise en service obtenues par la collaboration GASKAP. Ces structures sont longues de plusieurs centaines de pcs en projection sur le ciel et s'étendent sur plusieurs faisceaux ASKAP. Dans ce travail, j'ai analysé ces structures à l'aide d'un programme de décomposition en gaussiennes récemment introduit, ROHSA. Les décompositions en gaussiennes permettent de modéliser un spectre par une somme de fonctions gaussiennes dont les propriétés sont liées à celles du gaz HI. ROHSA utilise ce principe et impose également que les solutions des lignes de visée voisines soient spatialement cohérentes. L'utilisation de cet algorithme sur ces structures permet de créer des cartes pour chaque phase HI en vue d'une analyse plus approfondie. À partir de ces cartes, je constate que chaque nuage présente une distribution différente de la fraction de CNM. Pour tous les nuages, la fraction CNM diminue au fur et à mesure que les structures s'éloignent de la galaxie. Il y a également des preuves que les vitesses CNM et WNM changent en fonction de la distance de la galaxie et dans certaines zones de deux nuages, il y a des preuves de décalage dans la vitesse de

chaque phase. De plus, les mesures de CO de deux de ces nuages montrent que les amas de CO identifiés suivent préférentiellement la vitesse du CNM dans ces structures.

La deuxième partie de ce travail a consisté à trouver de plus petits nuages de HI autour des bords de la SMC. Les nuages devaient être décalés d'au moins 250pc ou 20 km/s de toute autre structure à petite échelle et ne pas être entourés d'émissions diffuses. J'ai utilisé une méthode automatisée initiale et une vérification visuelle pour identifier 31 nuages dans ce travail. Pour analyser les nuages, j'ai à nouveau utilisé ROHSA pour décomposer chaque nuage en ses composantes CNM et WNM. Les nuages ont ensuite été caractérisés à l'aide d'un logiciel de définition des filaments, en utilisant une fonction gaussienne pour modéliser la taille de chaque nuage afin d'obtenir une estimation de la densité. En combinant cela avec la température de chaque nuage, dérivée des largeurs de raies de décomposition spectrale, la variance des conditions du milieu interstellaire (ISM) autour du SMC a été explorée. J'ai trouvé une variance de plusieurs ordres de grandeur dans le champ de rayonnement FUV autour de chaque nuage, en supposant une chimie d'équilibre modélisée à partir des principaux processus de chauffage et de refroidissement dans l'ISM.

Au total, ce travail présente les premiers résultats de la décomposition de phase des données du programme GASKAP SMC avec ROHSA. Ce programme observera d'autres parties du système de Magellan et de la Voie Lactée, et ce travail démontre le potentiel de la décomposition en gaussiennes pour découvrir les phases du HI. Ce travail jette les bases de la décomposition de champs plus importants des données à venir. Ce travail a mis en évidence une méthodologie et des conditions pour trouver la meilleure solution de décomposition en gaussiennes régularisée avec ROHSA, qui pourra être appliquée à d'autres champs. Les résultats de cette analyse soulignent la capacité d'analyser les structures HI de façon très détaillée, donnant un aperçu de la façon dont le gaz est transporté autour de la galaxie et survit dans la périphérie du SMC. La dynamique extrême du SMC, causée par la forte interaction de marée avec son partenaire, le Grand Nuage de Magellan (LMC), fait qu'il n'est pas surprenant que du gaz HI froid et à petite échelle soit détecté à des distances significatives du corps principal de la galaxie. Ce travail propose un premier catalogue de structures à la périphérie de la galaxie. GASKAP observera le SMC pendant 200 heures au total, produisant une observation de meilleure qualité qui, avec d'autres traceurs de gaz et la dynamique stellaire, donnera une image plus claire de la structure galactique du SMC dans l'espace physique 3D et dans l'espace des vitesses.

Le chapitre 1 présente les motivations et l'état actuel des connaissances. Au chapitre 2, je détaille les processus utilisés pour déterminer les solutions les mieux adaptées avec l'algorithme ROHSA. Dans le chapitre 3, je discute de l'analyse des résultats des ajustements ROHSA sur 3 grandes structures et je les compare aux observations existantes à d'autres longueurs d'onde. Dans le chapitre 4, je discute de la semi-automatisation de la recherche de nuages supplémentaires autour du SMC et d'autres champs dans GASKAP, de l'ajustement ROHSA de ces nuages et des méthodes de mesure de leur densité. Dans le chapitre 5, je discute de l'analyse des nuages en les comparant aux diagrammes de phase HI calculés avec des conditions interstellaires variables. Dans le chapitre 6, je discute de

mes conclusions et des perspectives d'avenir.

Bibliography

- Arnal, E. M., Bajaja, E., Larrarte, J. J., Morras, R., and Pöppel, W. G. L. (2000). A high sensitivity HI survey of the sky at $\delta \leq -25^\circ$. *Astronomy and Astrophysics Supplement Series*, 142:35–40.
- Audit, E. and Hennebelle, P. (2004). Thermal condensation in a turbulent atomic hydrogen flow. *Astronomy and Astrophysics*, 433(1):1–13.
- Bajaja, E., Cappa de Nicolau, C. E., Cersosimo, J. C., Martin, M. C., Loiseau, N., Morras, R., Olano, C. A., and Poeppel, W. G. L. (1985). A new general survey of high-velocity neutral hydrogen in the southern hemisphere. *The Astrophysical Journal Supplement Series*, 58:143–165.
- Barbani, F., Pascale, R., Marinacci, F., Sales, L. V., Vogelsberger, M., Torrey, P., and Li, H. (2023). Galactic coronae in Milky Way-like galaxies: The role of stellar feedback in gas accretion. *Monthly Notices of the Royal Astronomical Society*, 524:4091–4108.
- Barnes, D. G., Staveley-Smith, L., de Blok, W. J. G., Oosterloo, T., Stewart, I. M., et al. (2001). The HI Parkes All Sky Survey: Southern observations, calibration and robust imaging. *Monthly Notices of the Royal Astronomical Society*, 322:486–498.
- Barrett, A. H. (1964). High-resolution Microwave Spectra of H and OH Absorption Lines of Cassiopeia A. *Nature*, 202:475–476.
- Besla, G., Kallivayalil, N., Hernquist, L., van der Marel, R. P., Cox, T. J., and Kereš, D. (2012). The role of dwarf galaxy interactions in shaping the Magellanic System and implications for Magellanic Irregulars. *Monthly Notices of the Royal Astronomical Society*, 421:2109–2138.
- Blaauw, A. and Tolbert, C. R. (1966). Intermediate-velocity features in the local hydrogen layer. *Bulletin of the Astronomical Institutes of the Netherlands*, 18:405.
- Bland-Hawthorn, J., Veilleux, S., Cecil, G. N., Putman, M. E., Gibson, B. K., and Maloney, P. R. (1998). The Smith cloud: HI associated with the SGR dwarf? *Monthly Notices of the Royal Astronomical Society*, 299:611–624.

- Blitz, L., Spergel, D. N., Teuben, P. J., Hartmann, D., and Burton, W. B. (1999). High-Velocity Clouds: Building Blocks of the Local Group. *The Astrophysical Journal*, 514:818–843.
- Bok, B. J. (1966). Magellanic Clouds. *Annual Review of Astronomy and Astrophysics*, 4:95.
- Boothroyd, A. I., Blagrove, K., Lockman, F. J., Martin, P. G., Gonçalves, D. P., and Srikanth, S. (2011). Accurate galactic 21-cm H I measurements with the NRAO Green Bank Telescope. *Astronomy and Astrophysics*, 536:1–23.
- Braun, R., Bonaldi, A., Bourke, T., Keane, E., and Wagg, J. (2019). Anticipated Performance of the Square Kilometre Array – Phase 1 (SKA1). *arXiv e-prints*, page arXiv:1912.12699.
- Bregman, J. N. (2004). The Origin of the High-Velocity Clouds. In van Woerden, H., Wakker, B. P., Schwarz, U. J., and de Boer, K. S., editors, *High Velocity Clouds*, volume 312 of *Astrophysics and Space Science Library*, page 341.
- Bruens, C., Kerp, J., Staveley-Smith, L., Mebold, U., Putman, M. E., Haynes, R. F., Kalberla, P. M. W., Muller, E., and Filipovic, M. D. (2005). The Parkes HI Survey of the Magellanic System. *Astronomy & Astrophysics*, 432(1):45–67.
- Brüns, C., Kerp, J., Kalberla, P. M. W., and Mebold, U. (2000). The head-tail structure of high-velocity clouds. A survey of the northern sky. *Astronomy and Astrophysics*, 357:120–128.
- Brüns, C., Kerp, J., and Pagels, A. (2001). Deep H I observations of the compact high-velocity cloud {HVC 125+41-207}. *Astronomy and Astrophysics*, 370:L26–L30.
- Burkhart, B., Stanimirović, S., Lazarian, A., and Kowal, G. (2010). Characterizing Magnetohydrodynamic Turbulence in the Small Magellanic Cloud. *The Astrophysical Journal*, 708:1204–1220.
- Casetti-Dinescu, D. I., Moni Bidin, C., Girard, T. M., Méndez, R. A., Vieira, K., Korchagin, V. I., and van Altena, W. F. (2014). Recent Star Formation in the Leading Arm of the Magellanic Stream. *The Astrophysical Journal*, 784:L37.
- Cashman, F. H., Fox, A. J., Savage, B. D., Wakker, B. P., Krishnarao, D., Benjamin, R. A., Richter, P., Ashley, T., Jenkins, E. B., Lockman, F. J., Bordoloi, R., and Kim, T.-S. (2021). Molecular Gas within the Milky Way's Nuclear Wind. *The Astrophysical Journal*, 923:L11.
- Chakraborty, A. and Roy, N. (2023). Detection of H I 21 cm emission from a strongly lensed galaxy at $z \sim 1.3$. *Monthly Notices of the Royal Astronomical Society*, 519(3):4074–4081.
- Chandra, V., Naidu, R. P., Conroy, C., Bonaca, A., Zaritsky, D., Cargile, P. A., Caldwell, N., Johnson, B. D., Han, J. J., and Ting, Y.-S. (2023). Discovery of the Magellanic Stellar Stream Out to 100 kpc. *The Astrophysical Journal*, 956(2):110.
- Chepurnov, A., Burkhart, B., Lazarian, A., and Stanimirovic, S. (2015). The Turbulence Velocity Power Spectrum of Neutral Hydrogen in the Small Magellanic Cloud. *The Astrophysical Journal*, 810:33.

- Christiansen, W. N. and Hindman, J. V. (1952). A Preliminary Survey of 1420 Mc/s. Line Emission from Galactic Hydrogen. *Australian Journal of Scientific Research A Physical Sciences*, 5:437.
- Clark, S. E., Peek, J. E. G., and Putman, M. E. (2014). Magnetically Aligned H I Fibers and the Rolling Hough Transform. *The Astrophysical Journal*, 789:82.
- Clutton-Brock, M. (1972). How Are Intergalactic Filaments Made? *Astrophysics and Space Science*, 17:292–324.
- Condon, J. J. and Ransom, S. M. (2016). *Essential Radio Astronomy*. Princeton University Press.
- Cullinane, L. R., Mackey, A. D., Da Costa, G. S., Erkal, D., Koposov, S. E., and Belokurov, V. (2022). The Magellanic Edges Survey - III. Kinematics of the disturbed LMC outskirts. *Monthly Notices of the Royal Astronomical Society*, 512:4798–4818.
- de Grijs, R., Wicker, J. E., and Bono, G. (2014). Clustering of Local Group distances: Publication bias or correlated measurements? I. The Large Magellanic Cloud. *The Astronomical Journal*, 147(5):122.
- de Heij, V., Braun, R., and Burton, W. B. (2002). An automated search for compact high-velocity clouds in the Leiden/Dwingeloo Survey. *Astronomy and Astrophysics*, 391:159–178.
- De Vaucouleurs, G. and Freeman, K. (1972). Structure and dynamics of barred spiral galaxies, in particular of the Magellanic type. *Vistas in Astronomy*, 14:163–294.
- Dempsey, J., McClure-Griiths, N. M., Murray, C., Dickey, J. M., Pingel, N. M., Jameson, K., Dénes, H., Van Loon, J. T., Leahy, D., Lee, M.-Y., Stanimirović, S., Breen, S., Buckland-Willis, F., Gibson, S. J., Imai, H., Lynn, C., and Tremblay, C. D. (2022). GASKAP-HI Pilot Survey Science III: An unbiased view of cold gas in the Small Magellanic Cloud. *Publications of the Astronomical Society of Australia*, 39:1–19.
- Desert, F. X., Bazell, D., and Blitz, L. (1990). CO and IRAS Detection of an Intermediate-Velocity Cloud. *The Astrophysical Journal*, 355:L51.
- Dessauges-Zavadsky, M., Combes, F., and Pfenniger, D. (2007). Molecular gas in high-velocity clouds: Revisited scenario. *Astronomy and Astrophysics*, 473(3):863–870.
- Di Teodoro, E. M., McClure-Griffiths, N. M., Breuck, C. D., Armillotta, L., Pingel, N. M., Jameson, K. E., Dickey, J. M., Rubio, M., Stanimirović, S., and Staveley-Smith, L. (2019a). Molecular Gas in the Outflow of the Small Magellanic Cloud. *The Astrophysical Journal*, 885(2):L32.
- Di Teodoro, E. M., McClure-Griffiths, N. M., Jameson, K. E., Dénes, H., Dickey, J. M., Stanimirović, S., Staveley-Smith, L., Anderson, C., Bunton, J. D., Chippendale, A., Lee-Waddell, K., MacLeod, A., and Voronkov, M. A. (2019b). On the dynamics of the Small Magellanic Cloud through high-resolution ASKAP H I observations. *Monthly Notices of the Royal Astronomical Society*, 483(1):392–406.

- Di Teodoro, E. M., McClure-Griffiths, N. M., Lockman, F. J., and Armillotta, L. (2020). Cold gas in the Milky Way's nuclear wind. *Nature*, 584:364–367.
- Di Teodoro, E. M., McClure-Griffiths, N. M., Lockman, F. J., Denbo, S. R., Endsley, R., Ford, H. A., and Harrington, K. (2018). Blowing in the Milky Way Wind: Neutral Hydrogen Clouds Tracing the Galactic Nuclear Outflow. *The Astrophysical Journal*, 855:33.
- Diaz, J. and Bekki, K. (2011). Constraining the orbital history of the Magellanic Clouds: A new bound scenario suggested by the tidal origin of the Magellanic Stream. *Monthly Notices of the Royal Astronomical Society*, 413:2015–2020.
- Dickey, J. M. and Lockman, F. J. (1990). H I in the galaxy. *Annual Review of Astronomy and Astrophysics*, 28:215–261.
- Dickey, J. M., McClure-Griffiths, N., Gibson, S. J., Gómez, J. F., Imai, H., et al. (2013). GASKAP-The Galactic ASKAP Survey. *Publications of the Astronomical Society of Australia*, 30:e003.
- Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., and Green, A. J. (2003). Fitting Together the H I Absorption and Emission in the Southern Galactic Plane Survey. *The Astrophysical Journal*, 585:801–822.
- Dickey, J. M., Mebold, U., Stanimirovic, S., and Staveley-Smith, L. (2000). Cold Atomic Gas in the Small Magellanic Cloud. *The Astrophysical Journal*, 536:756–772.
- Dieter, N. H. (1965). Neutral hydrogen near the galactic poles. *The Astronomical Journal*, 70:552.
- D'Onghia, E. and Fox, A. J. (2016). The Magellanic Stream: Circumnavigating the Galaxy. *Annual Review of Astronomy and Astrophysics*, 54(1):363–400.
- Draine, B. T. (2011). *Physics of the Interstellar and Intergalactic Medium*. Princeton University Press.
- Ewen, H. I. and Purcell, E. M. (1951). Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec. *Nature*, 168:356.
- Ewen, H. I. and Purcell, E. M. (1982). *Observation of a Line in the Galactic Radio Spectrum*, pages 328–330. Springer Netherlands, Dordrecht.
- Field, G. B. (1965). Thermal Instability. *The Astrophysical Journal*, 142:531.
- For, B. Q., Staveley-Smith, L., McClure-Griffiths, N. M., Westmeier, T., and Bekki, K. (2016). The distance and properties of hydrogen clouds in the Leading Arm of the Magellanic System. *Monthly Notices of the Royal Astronomical Society*, 461(1):892–907.
- Fox, A. J., Richter, P., Wakker, B. P., Lehner, N., Howk, J. C., Bekhti, N. B., Bland-Hawthorn, J., and Lucas, S. (2013). The COS/UVES Absorption Survey of the Magellanic Stream: I. One-Tenth Solar Abundances along the Body of the Stream. *The Astrophysical Journal*, 772(2):110.

- Fraternali, F. (2017). *Gas Accretion via Condensation and Fountains*, pages 323–353. Springer International Publishing.
- Fraternali, F., Marasco, A., Armillotta, L., and Marinacci, F. (2015). Galactic hail: The origin of the high-velocity cloud complex C. *Monthly Notices of the Royal Astronomical Society*, 447:L70–L74.
- Fraternali, F., Oosterloo, T., Sancisi, R., and van Moorsel, G. (2001). A New, Kinematically Anomalous H I Component in the Spiral Galaxy NGC 2403. *The Astrophysical Journal*, 562:L47–L50.
- Fraternali, F., van Moorsel, G., Sancisi, R., and Oosterloo, T. (2002). Deep H I Survey of the Spiral Galaxy NGC 2403. *The Astronomical Journal*, 123:3124–3140.
- Giovanelli, R., Haynes, M. P., Kent, B. R., Perillat, P., Saintonge, A., et al. (2005). The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy. *The Astronomical Journal*, 130:2598–2612.
- Giovanelli, R., Verschuur, G. L., and Cram, T. R. (1973). High resolution studies of high velocity clouds of neutral hydrogen. *Astronomy and Astrophysics Supplement Series*, 12:209–262.
- Glover, S. C. O. and Mac Low, M. M. (2011). On the relationship between molecular hydrogen and carbon monoxide abundances in molecular clouds. *Monthly Notices of the Royal Astronomical Society*, 412:337–350.
- Gong, M., Ostriker, E. C., and Wolfire, M. G. (2017). A Simple and Accurate Network for Hydrogen and Carbon Chemistry in the Interstellar Medium. *The Astrophysical Journal*, 843:38.
- Grisdale, K., Agertz, O., Romeo, A. B., Renaud, F., and Read, J. I. (2017). The impact of stellar feedback on the density and velocity structure of the interstellar medium. *Monthly Notices of the Royal Astronomical Society*, 466:1093–1110.
- Habing, H. J. (1968). The interstellar radiation density between 912 Å and 2400 Å. *Bulletin of the Astronomical Institutes of the Netherlands*, 19:421.
- Hagen, J. P., McClain, E. F., and Hepburn, N. (1954). Radio discrete sources and galactic absorption at 21 CM wavelength. *The Astronomical Journal*, 59:323.
- Hartmann, D. and Burton, W. B. (1997). *Atlas of Galactic Neutral Hydrogen*. Cambridge University Press.
- Haud, U. (2008). Gaussian decomposition of H I surveys. IV. Galactic intermediate- and high-velocity clouds. *Astronomy and Astrophysics*, 483(2):461–469.
- Hayakawa, T. and Fukui, Y. (2022). Dust-to-neutral gas ratio of the intermediate velocity HI clouds derived based on the sub-mm dust emission for the whole sky. *arXiv e-prints*, page arXiv:2208.13406.
- Heeschen, D. S. (1955). Some Features of Interstellar Hydrogen in the Section of the Galactic Center. *The Astrophysical Journal*, 121:569.

- Heiles, C. (1967). Observations of the Spatial Structure of Interstellar Hydrogen. I. High-Resolution Observations of a Small Region. *The Astrophysical Journal Supplement Series*, 15:97.
- Heiles, C. and Troland, T. H. (2003a). The Millennium Arecibo 21 Centimeter Absorption-Line Survey. I. Techniques and Gaussian Fits. *The Astrophysical Journal Supplement Series*, 145(2):329–354.
- Heiles, C. and Troland, T. H. (2003b). The Millennium Arecibo 21 Centimeter Absorption-Line Survey. II. Properties of the Warm and Cold Neutral Media. *The Astrophysical Journal*, 586(2):1067–1093.
- Heitsch, F., Marchal, A., Miville-Deschênes, M. A., Shull, J. M., and Fox, A. J. (2022). Mass, morphing, metallicities: the evolution of infalling high velocity clouds. *Monthly Notices of the Royal Astronomical Society*, 509(3):4515–4531.
- Heitsch, F. and Putman, M. E. (2009). The Fate of High-Velocity Clouds: Warm or Cold Cosmic Rain? *The Astrophysical Journal*, 698:1485–1496.
- Hennebelle, P. (2013). On the origin of non-self-gravitating filaments in the ISM. *Astronomy and Astrophysics*, 556:A153.
- Hennebelle, P., Audit, E., and Miville-Deschênes, M. A. (2007). On the structure of the turbulent interstellar atomic hydrogen. II. First comparison between observation and theory. Are the characteristics of molecular clouds determined early in the turbulent 2-phase atomic gas? *Astronomy and Astrophysics*, 465:445–456.
- HI4PI Collaboration et al. (2016). HI4PI: A full-sky H I survey based on EBHIS and GASS. *Astronomy and Astrophysics*, 594:A116.
- Hindman, J. V., Kerr, F. J., and McGee, R. X. (1963). A Low Resolution Hydrogen-line Survey of the Magellanic System. II. Interpretation of Results. *Australian Journal of Physics*, 16:570.
- Hulsbosch, A. N. M. and Raimond, E. (1966). Neutral Hydrogen with High Velocities at High Galactic Latitudes. *Bulletin of the Astronomical Institutes of the Netherlands*, 18:413–420.
- Hulsbosch, A. N. M. and Wakker, B. P. (1988). A deep, nearly complete, survey of northern high-velocity clouds. *Astronomy and Astrophysics Supplement Series*, 75:191–236.
- Ibata, R. A., Gilmore, G., and Irwin, M. J. (1994). A dwarf satellite galaxy in Sagittarius. *Nature*, 370:194–196.
- Inoue, T. and Inutsuka, S.-i. (2016). Formation of HI Clouds in Shock-compressed Interstellar Medium: Physical Origin of Angular Correlation Between Filamentary Structure and Magnetic Field. *The Astrophysical Journal*, 833(1):10.
- Jackson, J. M., Rathborne, J. M., Shah, R. Y., Simon, R., Bania, T. M., Clemens, D. P., Chambers, E. T., Johnson, A. M., Dormody, M., Lavoie, R., and Heyer, M. H. (2006). The Boston University-Five College Radio Astronomy Observatory Galactic Ring Survey. *The Astrophysical Journal Supplement Series*, 163:145–159.

- Jacyszyn-Dobrzeniecka, A. M., Skowron, D. M., Mróz, P., Skowron, J., Soszyński, I., Udalski, A., Pietrukowicz, P., Kozłowski, S., Wyrzykowski, Ł., Poleski, R., Pawlak, M., Szymański, M. K., and Ulaczyk, K. (2016). OGLE-ing the Magellanic System: Three-Dimensional Structure of the Clouds and the Bridge Using Classical Cepheids. *Acta Astronomica*, 66(2):149–196.
- Jameson, K. E., McClure-Griffiths, N. M., Liu, B., Dickey, J. M., Staveley-Smith, L., Stanimirović, S., Dempsey, J., Dawson, J. R., Dénes, H., Bolatto, A. D., Li, D., and Wong, T. (2019). An ATCA Survey of H I Absorption in the Magellanic Clouds. I. H I Gas Temperature Measurements in the Small Magellanic Cloud. *The Astrophysical Journal Supplement Series*, 244(1):7.
- Jung, S. L., Grønnow, A., and McClure-Griffiths, N. M. (2023). Magnetic field draping around clumpy high-velocity clouds in galactic halo. *Monthly Notices of the Royal Astronomical Society*, 522:4161–4180.
- Juvela, M., Malinen, J., and Lunttila, T. (2012). Profiles of interstellar cloud filaments: Observational effects in synthetic sub-millimetre observations. *Astronomy & Astrophysics*, 544:A141.
- Kahn, F. D. (1994). Galactic Fountains. *Astrophysics and Space Science*, 216:325–332.
- Kalberla, P. M. W., Burton, W. B., Hartmann, D., Arnal, E. M., Bajaja, E., Morras, R., and Pöppel, W. G. L. (2005). The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI. Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections. *Astronomy and Astrophysics*, 440:775–782.
- Kalberla, P. M. W. and Haud, U. (2006). Global properties of the HI high velocity sky. A statistical investigation based on the LAB survey. *Astronomy and Astrophysics*, 455(2):481–498.
- Kalberla, P. M. W., Kerp, J., Haud, U., Winkel, B., Bekhti, N. B., Floer, L., and Lenz, D. (2016). Cold Milky Way HI gas in filaments. *The Astrophysical Journal*, 821(2):117.
- Kerp, J., Winkel, B., Ben Bekhti, N., Flöer, L., and Kalberla, P. M. W. (2011). The Effelsberg Bonn H I Survey (EBHIS). *Astronomische Nachrichten*, 332:637.
- Kerr, F. J. and Hindman, J. V. (1953). Preliminary report on a survey of 21 CM radiation from the Magellanic Clouds. *The Astronomical Journal*, 58:218.
- Kobulnicky, H. A. and Dickey, J. M. (1999). Detection of Cold Atomic Clouds in the Magellanic Bridge. *The Astronomical Journal*, 117:908–919.
- Koch, E. W. and Rosolowsky, E. W. (2015). Filament identification through mathematical morphology. *Monthly Notices of the Royal Astronomical Society*, 452:3435–3450.
- Konz, C., Brüns, C., and Birk, G. T. (2002). Dynamical evolution of high velocity clouds in the intergalactic medium. *Astronomy and Astrophysics*, 391:713–723.

- Lenz, D., Flöer, L., and Kerp, J. (2016). Dust in a compact, cold, high-velocity cloud: A new approach to removing foreground emission. *Astronomy and Astrophysics*, 586:A121.
- Li, A., Fraternali, F., Marasco, A., Trager, S. C., Pezzulli, G., Mancera Piña, P. E., and Verheijen, M. A. W. (2023). Fountain-driven gas accretion feeding star formation over the disc of NGC 2403. *Monthly Notices of the Royal Astronomical Society*, 520:147–160.
- Lindblad, P. O. (1966). Dwingeloo Atlas of 21-cm profiles, Part III, with a representation in Gaussian components. *Bulletin of the Astronomical Institutes of the Netherlands Supplement Series*, 1:177.
- Lindner, R. R., Vera-Ciro, C., Murray, C. E., Stanimirović, S., Babler, B., Heiles, C., Hennebelle, P., Goss, W. M., and Dickey, J. (2015). Autonomous Gaussian Decomposition. *The Astronomical Journal*, 149:138.
- Livingston, J. D., McClure-Griffiths, N. M., Mao, S. A., Ma, Y. K., Gaensler, B. M., Heald, G., and Seta, A. (2021). A Radio Polarisation Study of Magnetic Fields in the Small Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society*, 510(1):260–275.
- Luri, X., Chemin, L., Clementini, G., Delgado, H. E., McMillan, P. J., et al. (2021). Gaia Early Data Release 3: Structure and properties of the Magellanic Clouds. *Astronomy and Astrophysics*, 649:A7.
- Ma, Y. K., McClure-Griffiths, N. M., Clark, S. E., Gibson, S. J., van Loon, J. T., Soler, J. D., Putman, M. E., Dickey, J. M., Lee, M. Y., Jameson, K. E., Uscanga, L., Dempsey, J., Dénes, H., Lynn, C., and Pingel, N. M. (2023). H I filaments as potential compass needles? Comparing the magnetic field structure of the Small Magellanic Cloud to the orientation of GASKAP-H I filaments. *Monthly Notices of the Royal Astronomical Society*, 521(1):60–83.
- Madden, S. C., Cormier, D., Hony, S., Lebouteiller, V., Abel, N., Galametz, M., De Looze, I., Chevance, M., Polles, F. L., Lee, M. Y., Galliano, F., Lambert-Huyghe, A., Hu, D., and Ramambason, L. (2020). Tracing the total molecular gas in galaxies: [CII] and the CO-dark gas. *Astronomy and Astrophysics*, 643:A141.
- Magnani, L. and Smith, A. J. (2010). Intermediate-velocity Molecular Gas at High Northern Galactic Latitudes. *The Astrophysical Journal*, 722:1685–1690.
- Maia, F. F., Piatti, A. E., and Santos, J. F. (2013). Mass distribution and structural parameters of Small Magellanic Cloud star clusters. *Monthly Notices of the Royal Astronomical Society*, 437(2):2005–2016.
- Mao, S. A. and Ostriker, E. C. (2018). Galactic Disk Winds Driven by Cosmic Ray Pressure. *The Astrophysical Journal*, 854:89.
- Marasco, A., Fraternali, F., Lehner, N., and Howk, J. C. (2022). Intermediate- and high-velocity clouds in the Milky Way - II. Evidence for a Galactic fountain with collimated outflows and diffuse inflows. *Monthly Notices of the Royal Astronomical Society*, 515:4176–4190.

- Marchal, A., Martin, P. G., and Gong, M. (2021). Resolving the Formation of Cold H I Filaments in the High-velocity Cloud Complex C. *The Astrophysical Journal*, 921:11.
- Marchal, A., Miville-Deschênes, M. A., Orioux, F., Gac, N., Soussen, C., Lesot, M. J., D'Allonnes, A. R., and Salomé, Q. (2019). ROHSA: Regularized Optimization for Hyper-Spectral Analysis. *Astronomy and Astrophysics*, 626(1965).
- Martínez-Delgado, D., Katherina Vivas, A., Grebel, E. K., Gallart, C., Pieres, A., et al. (2019). Nature of a shell of young stars in the outskirts of the Small Magellanic Cloud. *Astronomy and Astrophysics*, 631(2017).
- Marx-Zimmer, M., Herbstmeier, U., Dickey, J. M., Zimmer, F., Staveley-Smith, L., and Mebold, U. (2000). A study of the cool gas in the Large Magellanic Cloud. I. Properties of the cool atomic phase - a third H I absorption survey. *Astronomy and Astrophysics*, 354:787–801.
- Mastropietro, C., Moore, B., Mayer, L., Wadsley, J., and Stadel, J. (2005). The gravitational and hydrodynamical interaction between the Large Magellanic Cloud and the Galaxy. *Monthly Notices of the Royal Astronomical Society*, 363:509–520.
- Mathewson, D. S., Cleary, M. N., and Murray, J. D. (1974). The Magellanic Stream. *The Astrophysical Journal*, 190:291–296.
- McClure-Griffiths, N. M., Dénes, H., Dickey, J. M., Stanimirović, S., Staveley-Smith, L., Jameson, K., Di Teodoro, E., Allison, J. R., Collier, J. D., Chippendale, A. P., Franzen, T., Gürkan, G., Heald, G., Hotan, A., Kleiner, D., Lee-Waddell, K., McConnell, D., Popping, A., Rhee, J., Riseley, C. J., Voronkov, M. A., and Whiting, M. (2018). Cold gas outflows from the Small Magellanic Cloud traced with ASKAP. *Nature Astronomy*, 2(11):901–906.
- McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., Green, A. J., and Haverkorn, M. (2006). Magnetically Dominated Strands of Cold Hydrogen in the Riegel-Crutcher Cloud. *The Astrophysical Journal*, 652(2):1339–1347.
- McClure-Griffiths, N. M., Pisano, D. J., Calabretta, M. R., Ford, H. A., Lockman, F. J., Staveley-Smith, L., Kalberla, P. M. W., Bailin, J., Dedes, L., Janowiecki, S., Gibson, B. K., Murphy, T., Nakanishi, H., and Newton-McGee, K. (2009). GASS: The Parkes Galactic All-Sky Survey. I. Survey description, goals, and initial data release. *The Astrophysical Journal Supplement Series*, 181(2):398.
- McClure-Griffiths, N. M., Stanimirović, S., and Rybarczyk, D. R. (2023). Atomic Hydrogen in the Milky Way: A Stepping Stone in the Evolution of Galaxies. *Annual Review of Astronomy and Astrophysics*, 61(1):19–63.
- McGee, R. X. and Newton, L. M. (1986). HI profiles in the bridge region of the Magellanic Clouds. *Publications of the Astronomical Society of Australia*, 6:471–500.
- McKee, C. F. and Ostriker, J. P. (1977). A theory of the interstellar medium: Three components regulated by supernova explosions in an inhomogeneous substrate. *The Astrophysical Journal*, 218:148–169.

- Mebold, U. (1972). On the Intercloud Hi-Gas. *Astronomy and Astrophysics*, 19:13.
- Miller, E. D., Bregman, J. N., and Wakker, B. P. (2009). High-Velocity Clouds in the Nearby Spiral Galaxy M 83. *The Astrophysical Journal*, 692:470–491.
- Misawa, T., Charlton, J. C., Kobulnicky, H. A., Wakker, B. P., and Bland-Hawthorn, J. (2009). The Magellanic Bridge as a Damped Lyman Alpha System: Physical Properties of Cold Gas Toward PKS 0312-770. *The Astrophysical Journal*, 695:1382–1398.
- Miville-Deschênes, M. A., Levrier, F., and Falgarone, E. (2003). On the Use of Fractional Brownian Motion Simulations to Determine the Three-dimensional Statistical Properties of Interstellar Gas. *The Astrophysical Journal*, 593:831–847.
- Mizuno, N., Muller, E., Maeda, H., Kawamura, A., Minamidani, T., Onishi, T., Mizuno, A., and Fukui, Y. (2006). Detection of Molecular Clouds in the Magellanic Bridge: Candidate Star Formation Sites in a Nearby Low-Metallicity System. *The Astrophysical Journal*, 643:L107–L110.
- Morras, R., Bajaja, E., Arnal, E. M., and Pöppel, W. G. L. (2000). A new survey for high velocity HI detections in the Southern Hemisphere. *Astronomy and Astrophysics Supplement Series*, 142:25–30.
- Muller, C. A. (1957). 21-CM Absorption Effects in the Spectra of Two Strong Radio Sources. *The Astrophysical Journal*, 125:830.
- Muller, C. A. (1959). 21-cm observations on the Coma cluster. *Bulletin of the Astronomical Institutes of the Netherlands*, 14:339.
- Muller, C. A. and Oort, J. H. (1951). Observation of a Line in the Galactic Radio Spectrum: The Interstellar Hydrogen Line at 1,420 Mc./sec., and an Estimate of Galactic Rotation. *Nature*, 168:357–358.
- Muller, C. A., Oort, J. H., and Raimond, E. (1963). Hydrogène neutre dans la couronne galactique? *Academie des Sciences Paris Comptes Rendus*, 257:1661–1662.
- Münch, G. and Zirin, H. (1961). Interstellar Matter at Large Distances from the Galactic Plane. *The Astrophysical Journal*, 133:11.
- Muraveva, T., Subramanian, S., Clementini, G., Cioni, M.-R. L., Palmer, M., van Loon, J. T., Moretti, M. I., de Grijs, R., Molinaro, R., Ripepi, V., Marconi, M., Emerson, J., and Ivanov, V. D. (2018). The VMC survey - XXVI. Structure of the Small Magellanic Cloud from RR Lyrae stars. *Monthly Notices of the Royal Astronomical Society*, 473(3):3131–3146.
- Murray, C. E., Peek, J. E. G., and Kim, C.-G. (2020). Extracting the Cold Neutral Medium from H I Emission with Deep Learning: Implications for Galactic Foregrounds at High Latitude. *The Astrophysical Journal*, 899(1):15.

- Murray, C. E., Stanimirović, S., Goss, W. M., Dickey, J. M., Heiles, C., Lindner, R. R., Babler, B., Pingel, N. M., Lawrence, A., Jencson, J., and Hennebelle, P. (2015a). The 21-sponge H i absorption survey. I. Techniques and initial results. *Astrophysical Journal*, 804(2).
- Murray, C. E., Stanimirović, S., Goss, W. M., Heiles, C., Dickey, J. M., Babler, B., and Kim, C.-G. (2018). The 21-SPONGE H i Absorption Line Survey. I. The Temperature of Galactic H i. *The Astrophysical Journal Supplement Series*, 238(2):14.
- Murray, C. E., Stanimirović, S., Heiles, C., Dickey, J. M., McClure-Griffiths, N. M., Lee, M. Y., Goss, W., and Killerby-Smith, N. (2021). The MACH HI Absorption Survey. I. Physical Conditions of Cold Atomic Gas outside of the Galactic Plane. *The Astrophysical Journal Supplement Series*, 256:37.
- Murray, C. E., Stanimirović, S., McClure-Griffiths, N. M., Putman, M. E., Liszt, H. S., Wong, T., Richter, P., Dawson, J. R., Dickey, J. M., Lindner, R. R., Babler, B. L., and Allison, J. R. (2015b). First Detection of HCO⁺ Absorption in the Magellanic System. *The Astrophysical Journal*, 808(1):41.
- Nidever, D. L., Majewski, S. R., Butler Burton, W., Nidever, D. L., Majewski, S. R., and Butler Burton, W. (2008). The Origin of the Magellanic Stream and Its Leading Arm. *The Astrophysical Journal*, 679(1):432–459.
- Nidever, D. L., Majewski, S. R., Butler Burton, W., and Nigra, L. (2010). The 200° Long Magellanic Stream System. *The Astrophysical Journal*, 723:1618–1631.
- Noon, K. (2022). *Astronomy: Sky Country*. Thames Hudson Australia.
- Noon, K. A., Krumholz, M. R., Di Teodoro, E. M., McClure-Griffiths, N. M., Lockman, F. J., and Armillotta, L. (2023). Direct observations of the atomic-molecular phase transition in the Milky Way's nuclear wind. *Monthly Notices of the Royal Astronomical Society*, 524(1):1258–1268.
- Ntormousi, E., Burkert, A., Fierlinger, K., and Heitsch, F. (2011). Formation of Cold Filamentary Structure from Wind Blown Superbubbles. *The Astrophysical Journal*, 731(1):13.
- Ohno, T., Tokuda, K., Konishi, A., Matsumoto, T., Sewiło, M., Kondo, H., Sano, H., Tsuge, K., Zahorecz, S., Goto, N., Neelamkudan, N., Wong, T., Fukushima, H., Takekoshi, T., Muraoka, K., Kawamura, A., Tachihara, K., Fukui, Y., and Onishi, T. (2023). An unbiased CO survey toward the northern region of the small magellanic cloud with the atacama compact array. ii. CO cloud catalog. *The Astrophysical Journal*, 949(2):63.
- Oliveira, R. A. P., Maia, F. F. S., Barbuy, B., Dias, B., Santos, J. F. C., Souza, S. O., Kerber, L. O., Bica, E., Sanmartim, D., Quint, B., Fraga, L., Armond, T., Minniti, D., Parisi, M. C., Katime Santrich, O. J., Angelo, M. S., Pérez-Villegas, A., and De Bórtoli, B. J. (2023). The VISCACHA survey - VII. Assembly history of the Magellanic Bridge and SMC Wing from star clusters. *Monthly Notices of the Royal Astronomical Society*, 524:2244–2261.

- Orchiston, W. (2000). A Polynesian Astronomical Perspective: The Maori of New Zealand. In Selin, H. and Xiaochun, S., editors, *Astronomy Across Cultures: The History of Non-Western Astronomy*, Science Across Cultures: The History of Non-Western Science, pages 161–196. Springer Netherlands, Dordrecht.
- Pellegrini, E. W., Oey, M. S., Winkler, P. F., Points, S. D., Smith, R. C., Jaskot, A. E., and Zastrow, J. (2012). The Optical Depth of H II Regions in the Magellanic Clouds. *The Astrophysical Journal*, 755:40.
- Piatti, A. E. (2022). Revisiting a detached stellar structure in the outer north-eastern region of the Small Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society*, 509(3):3462–3469.
- Pingel, N. M., Dempsey, J., McClure-Griffiths, N. M., Dickey, J. M., Jameson, K. E., et al. (2022). GASKAP-HI pilot survey science I: ASKAP zoom observations of HI emission in the Small Magellanic Cloud. *Publications of the Astronomical Society of Australia*, 39.
- Planck Collaboration, et al., P. (2011). Planck early results. XXIV. Dust in the diffuse interstellar medium and the Galactic halo. *Astronomy and Astrophysics*, 536:A24.
- Putman, M. E., de Heij, V., Staveley-Smith, L., Braun, R., Freeman, K. C., et al. (2002). HIPASS High-Velocity Clouds: Properties of the Compact and Extended Populations. *The Astronomical Journal*, 123:873–891.
- Putman, M. E., Gibson, B. K., Staveley-Smith, L., Banks, G., Barnes, D. G., Bhatla, R., Disney, M. J., Ekers, R. D., Freeman, K. C., Haynes, R. F., Henning, P., Jerjen, H., Kilborn, V., Koribalski, B., Knezek, P., Malin, D. F., Mould, J. R., Oosterloo, T., Price, R. M., Ryder, S. D., Sadler, E. M., Stewart, I., Stootman, F., Vaile, R. A., Webster, R. L., and Wright, A. E. (1998). Tidal disruption of the Magellanic Clouds by the Milky Way. *Nature*, 394:752–754.
- Putman, M. E., Saul, D. R., and Mets, E. (2011). Head-tail clouds: Drops to probe the diffuse Galactic halo. *Monthly Notices of the Royal Astronomical Society*, 418:1575–1586.
- Putman, M. E., Staveley-Smith, L., Freeman, K. C., Gibson, B. K., and Barnes, D. G. (2003). The Magellanic Stream, High-Velocity Clouds, and the Sculptor Group. *The Astrophysical Journal*, 586:170–194.
- Quilis, V. and Moore, B. (2001). Where are the high velocity clouds? *The Astrophysical Journal*, 555(2):L95–L98.
- Renaud, F., Bournaud, F., Kraljic, K., and Duc, P.-A. (2014). Starbursts triggered by intergalactic tides and interstellar compressive turbulence. *Monthly Notices of the Royal Astronomical Society: Letters*, 442(1):L33–L37.
- Richter, P. and de Boer, K. S. (2004). The Coldest Phase in Halo High-Velocity Gas: Dust and Molecules. In van Woerden, H., Wakker, B. P., Schwarz, U. J., and de Boer, K. S., editors, *High Velocity Clouds*, volume 312 of *Astrophysics and Space Science Library*, page 183.

- Richter, P., Fox, A. J., Wakker, B. P., Howk, J. C., Lehner, N., Barger, K. A., D'Onghia, E., and Lockman, F. J. (2018). New Constraints on the Nature and Origin of the Leading Arm of the Magellanic Stream. *The Astrophysical Journal*, 865:145.
- Richter, P., Fox, A. J., Wakker, B. P., Lehner, N., Howk, J. C., Bland-Hawthorn, J., Bekhti, N. B., and Fechner, C. (2013). The COS/UVES Absorption Survey of the Magellanic Stream: II. Evidence for a complex enrichment history of the Stream from the Fairall 9 sightline. *The Astrophysical Journal*, 772(2):111.
- Richter, P., Wakker, B. P., Savage, B. D., and Sembach, K. R. (2003). A Far Ultraviolet Spectroscopic Explorer Survey of Molecular Hydrogen in Intermediate-Velocity Clouds in the Milky Way Halo. *The Astrophysical Journal*, 586:230–248.
- Riener, M., Kainulainen, J., Henshaw, J. D., Orkisz, J. H., Murray, C. E., and Beuther, H. (2019). GAUSSPY+: A fully automated Gaussian decomposition package for emission line spectra. *Astronomy and Astrophysics*, 628:A78.
- Ripepi, V., Cioni, M.-R. L., Moretti, M. I., Marconi, M., Bekki, K., Clementini, G., de Grijs, R., Emerson, J., Groenewegen, M. A. T., Ivanov, V. D., Molinaro, R., Muraveva, T., Oliveira, J. M., Piatti, A. E., Subramanian, S., and van Loon, J. T. (2017). The VMC survey – XXV. The 3D structure of the Small Magellanic Cloud from Classical Cepheids. *Monthly Notices of the Royal Astronomical Society*, 472(1):808–827.
- Rolleston, W. R. J., Trundle, C., and Dufton, P. L. (2002). The present-day chemical composition of the LMC. *Astronomy and Astrophysics*, 396:53–64.
- Russell, S. C. and Dopita, M. A. (1992). Abundances of the Heavy Elements in the Magellanic Clouds. III. Interpretation of Results. *The Astrophysical Journal*, 384:508.
- Rybarczyk, D. R., Stanimirović, S., Gong, M., Babler, B., Murray, C. E., Gerin, M., Winters, J. M., Luo, G., Dame, T. M., and Steffes, L. (2022). The Role of Neutral Hydrogen in Setting the Abundances of Molecular Species in the Milky Way's Diffuse Interstellar Medium. I. Observational Constraints from ALMA and NOEMA. *The Astrophysical Journal*, 928(1):79.
- Saury, E., Miville-Deschênes, M.-A., Hennebelle, P., Audit, E., and Schmidt, W. (2014). The structure of the thermally bistable and turbulent atomic gas in the local interstellar medium. *Astronomy & Astrophysics*, 567:A16.
- Shuter, W. L. H. and Verschuur, G. L. (1964). A high resolution investigation of 21 cm absorption. *Monthly Notices of the Royal Astronomical Society*, 127:387.
- Skowron, D. M., Jacyszyn, A. M., Udalski, A., Szymański, M. K., Skowron, J., Poleski, R., Kozłowski, S., Kubiak, M., Pietrzyński, G., Soszyński, I., Mróz, P., Pietrukowicz, P., Ulaczyk, K., and Wyrzykowski, Ł. (2014). OGLE-ING the Magellanic System: Stellar Populations in the Magellanic Bridge. *The Astrophysical Journal*, 795:108.

- Smart, B. M., Haffner, L. M., Barger, K. A., Hill, A., and Madsen, G. (2019). The Diffuse Ionized Gas Halo of the Small Magellanic Cloud. *The Astrophysical Journal*, 887(1):16.
- Sparke, L. S. and Gallagher, John S., I. (2000). *Galaxies in the universe : an introduction*. Cambridge University Press.
- Spitzer, L. (1978). *Physical Processes in the Interstellar Medium*. A Wiley-Interscience Publication.
- Spitzer, Jr., L. (1956). On a Possible Interstellar Galactic Corona. *The Astrophysical Journal*, 124:20.
- Stanimirović, S., Staveley-Smith, L., and Jones, P. A. (2004). A New Look at the Kinematics of Neutral Hydrogen in the Small Magellanic Cloud. *The Astrophysical Journal*, 604(1):176–186.
- Staveley-Smith, L., Kim, S., Calabretta, M. R., Haynes, R. F., and Kesteven, M. J. (2003). A New Look at the Large-Scale HI Structure of the LMC. *Monthly Notices of the Royal Astronomical Society*, 339(1):87–104.
- Staveley-Smith, L., Sault, R. J., Hatzidimitriou, D., Kesteven, M. J., and McConnell, D. (1997). An HI aperture synthesis mosaic of the Small Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society*, 289:225–252.
- Syed, J., Soler, J. D., Beuther, H., Wang, Y., Suri, S., Henshaw, J. D., Riener, M., Bialy, S., Kh., S. R., Stil, J. M., Goldsmith, P. F., Rugel, M. R., Glover, S. C. O., Klessen, R. S., Kerp, J., Urquhart, J. S., Ott, J., Roy, N., Schneider, N., Smith, R. J., Longmore, S. N., and Linz, H. (2022). The "Maggie" filament: Physical properties of a giant atomic cloud. *Astronomy & Astrophysics*, 657:A1.
- Taank, M., Marchal, A., Martin, P. G., and Vujeva, L. (2022). Mapping the Thermal Condensation of Diffuse H i in the North Celestial Pole Loop. *The Astrophysical Journal*, 937(2):81.
- Takakubo, K. (1967). Neutral hydrogen at intermediate galactic latitudes. III. Local kinematical properties derived from Gaussian 21-cm profile components; comparison with Ca⁺ K-line data. *Bulletin of the Astronomical Institutes of the Netherlands*, 19:125.
- Takakubo, K. and van Woerden, H. (1966). Neutral Hydrogen at Intermediate Galactic Latitudes. *Bulletin of the Astronomical Institutes of the Netherlands*, 18:488–533.
- Tatton, B. L., van Loon, J. T., Cioni, M. R. L., Bekki, K., Bell, C. P. M., Choudhury, S., de Grijs, R., Groenewegen, M. A. T., Ivanov, V. D., Marconi, M., Oliveira, J. M., Ripepi, V., Rubele, S., Subramanian, S., and Sun, N. C. (2021). The VMC Survey - XL. Three-dimensional structure of the Small Magellanic Cloud as derived from red clump stars. *Monthly Notices of the Royal Astronomical Society*, 504:2983–2997.
- Tchernyshyov, K. (2022). A Detection of H₂ in a High-velocity Cloud toward the Large Magellanic Cloud. *The Astrophysical Journal*, 931(2):78.

- Thompson, A. R., Moran, J. M., and Swenson, Jr., G. W. (2017). *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. Springer.
- Tokuda, K., Harada, N., Tanaka, K. E. I., Inoue, T., Shimonishi, T., Zhang, Y., Sewiło, M., Kunitoshi, Y., Konishi, A., Fukui, Y., Kawamura, A., Onishi, T., and Machida, M. N. (2023). An alma glimpse of dense molecular filaments associated with high-mass protostellar systems in the large magellanic cloud. *The Astrophysical Journal*, 955(1):52.
- Valdivia-Mena, M. T., Rubio, M., Bolatto, A. D., Saldaño, H. P., and Verdugo, C. (2020). ALMA resolves molecular clouds in metal-poor Magellanic Bridge A. *Astronomy and Astrophysics*, 641:A97.
- Vallenari, A., Brown, A. G. A., Prusti, T., de Bruijne, J. H. J., Arenou, F., et al. (2023). Gaia Data Release 3 - Summary of the content and survey properties. *Astronomy & Astrophysics*, 674:A1.
- van de Hulst, H. C., Muller, C. A., and Oort, J. H. (1954). The spiral structure of the outer part of the Galactic System derived from the hydrogen emission at 21 cm wavelength. *Bulletin of the Astronomical Institutes of the Netherlands*, 12:117.
- van der Hulst, T. and Sancisi, R. (1988). High-Velocity Gas in M101. *The Astronomical Journal*, 95:1354.
- van Woerden, H., Takakubo, K., and Braes, L. L. E. (1962). Neutral hydrogen at intermediate galactic latitudes. I. 21-cm profiles in seven latitude zones and four special regions. *Bulletin of the Astronomical Institutes of the Netherlands*, 16:321.
- van Woerden, H. and Wakker, B. P. (2004). Distances and Metallicities of HVCS. In van Woerden, H., Wakker, B. P., Schwarz, U. J., and de Boer, K. S., editors, *High Velocity Clouds*, volume 312 of *Astrophysics and Space Science Library*, page 195.
- Venzmer, M. S., Kerp, J., and Kalberla, P. M. W. (2012). The four leading arms of the Magellanic Cloud system. Evidence for interaction with Milky Way disk and halo. *Astronomy and Astrophysics*, 547:A12.
- Vujeva, L., Marchal, A., Martin, P. G., and Taank, M. (2023). Mapping the Multiphase Structure of H I in the Low-latitude Intermediate-velocity Arch 1. *The Astrophysical Journal*, 951:120.
- Wakker, B. P. (2004). HVC/IVC Maps and HVC Distribution Functions. In van Woerden, H., Wakker, B. P., Schwarz, U. J., and de Boer, K. S., editors, *High Velocity Clouds*, volume 312 of *Astrophysics and Space Science Library*, page 25.
- Wakker, B. P., Murphy, E. M., van Woerden, H., and Dame, T. M. (1997). A Sensitive Search for Molecular Gas in High-Velocity Clouds. *The Astrophysical Journal*, 488(1):216–223.
- Wakker, B. P. and van Woerden, H. (1991). Distribution and origin of high-velocity clouds. III. Clouds, complexes and populations. *Astronomy and Astrophysics*, 250:509–532.

- Wakker, B. P. and van Woerden, H. (2013). High-Velocity Clouds. In Oswalt, T. D. and Gilmore, G., editors, *Planets, Stars and Stellar Systems: Volume 5: Galactic Structure and Stellar Populations*, pages 587–640. Springer Netherlands, Dordrecht.
- Wang, J., Hammer, F., and Yang, Y. (2022). Lessons from the Magellanic System and its modeling. *Monthly Notices of the Royal Astronomical Society*, 515(1):940–952.
- Wang, J., Hammer, F., Yang, Y., Ripepi, V., Cioni, M. R. L., Puech, M., and Flores, H. (2019). Towards a complete understanding of the Magellanic Stream Formation. *Monthly Notices of the Royal Astronomical Society*, 486(4):5907–5916.
- Wannier, P. and Wrixon, G. T. (1972). An Unusual High-Velocity Hydrogen Feature. *The Astrophysical Journal*, 173:L119.
- Welty, D. E., Lauroesch, J. T., Wong, T., and York, D. G. (2016). Thermal Pressures in the Interstellar Medium of the Magellanic Clouds. *The Astrophysical Journal*, 821(2):118.
- Westmeier, T. (2018). A new all-sky map of Galactic high-velocity clouds from the 21-cm HI4PI survey. *Monthly Notices of the Royal Astronomical Society*, 474:289–299.
- Westmeier, T., Brüns, C., and Kerp, J. (2005). Compact High-Velocity Clouds around the Galaxy and M31. In Braun, R., editor, *Extra-Planar Gas*, volume 331 of *Astronomical Society of the Pacific Conference Series*, page 105.
- Westmeier, T., Brüns, C., and Kerp, J. (2008). Relics of structure formation: Extra-planar gas and high-velocity clouds around the Andromeda Galaxy. *Monthly Notices of the Royal Astronomical Society*, 390:1691–1709.
- Winkler, P. F., Smith, R. C., Points, S. D., and MCELS Team (2015). The Interstellar Medium in the Small Magellanic Cloud: Results from MCELS. In Points, S. and Kunder, A., editors, *Fifty Years of Wide Field Studies in the Southern Hemisphere: Resolved Stellar Populations of the Galactic Bulge and Magellanic Clouds*, volume 491 of *Astronomical Society of the Pacific Conference Series*, page 343.
- Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., and Bakes, E. L. O. (1995). The Neutral Atomic Phases of the Interstellar Medium. *The Astrophysical Journal*, 443:152.
- Wolfire, M. G., McKee, C. F., Hollenbach, D., and Tielens, A. G. G. M. (2003). Neutral Atomic Phases of the Interstellar Medium in the Galaxy. *The Astrophysical Journal*, 587(1):278–311.
- Zucker, C. and Chen, H. H.-H. (2018). RadFil : A Python Package for Building and Fitting Radial Profiles for Interstellar Filaments. *The Astrophysical Journal*, 864(2):152.
- Zucker, C., Goodman, A., Alves, J., Bialy, S., Koch, E. W., Speagle, J. S., Foley, M. M., Finkbeiner, D., Leike, R., Enßlin, T., Peek, J. E. G., and Edenhofer, G. (2021). On the Three-dimensional Structure of Local Molecular Clouds. *The Astrophysical Journal*, 919(1):35.