Astronomy & Astrophysics manuscript no. HIandCO
January 2, 2019

Histogram of oriented gradients: a technique for the study of molecular cloud formation.

J. D. Soler[1]⋆, H. Beuther[1], M. Rugel[1], Y. Wang[1], P. C. Clark[2], S. C. O. Glover[3], P. F. Goldsmith[4], M. Heyer[5], L. D. Anderson[6], A. Goodman[7], Th. Henning[1], J. Kainulainen[8], R. S. Klessen[3,9], S. N. Longmore[10], N. M. McClure-Griffiths[11], K. M. Menten[12], J. C. Mottram[1], J. Ott[13], S. E. Ragan[2], R. J. Smith[14], J. S. Urquhart[15], F. Bigiel[16,3], P. Hennebelle[17], N. Roy[18], P. Schilke[19].

1. Max Planck Institute for Astronomy, Königstuhl 17, 69117, Heidelberg, Germany.
2. School of Physics and Astronomy, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA, UK.
3. Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120, Heidelberg, Germany.
4. Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena CA, 91109, USA.
5. Department of Astronomy, University of Massachusetts, Amherst, MA 01003-9305, USA.
6. Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA.
7. Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS 42, Cambridge, MA 02138, USA.
8. Dept. of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, 439 92 Onsala, Sweden.
9. Universität Heidelberg, Interdiszipliäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany.
10. Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK.
11. Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT, Australia.
12. Max Planck Institute for Radio Astronomy, Auf dem Hügel 69, 53121 Bonn, Germany.
13. National Radio Astronomy Observatory, PO Box O, 1003 Lopezville Road, Socorro, NM 87801, USA.
14. Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK.
15. Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK.
16. Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany.
17. Laboratoire AIM, Paris-Saclay, CEA/IRFU/SAp - CNRS - Université Paris Diderot, 91191, Gif-sur-Yvette Cedex, France.
18. Department of Physics, Indian Institute of Science, 560012 Bangalore, India.
19. Physikalisches Institut der Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany.

Received 21 SEP 2018 / Accepted 28 DEC 2018

ABSTRACT

We introduce the histogram of oriented gradients (HOG), a tool developed for machine vision that we propose as a new metric for the systematic characterization of spectral line observations of atomic and molecular gas and the study of molecular cloud formation models. In essence, the HOG technique takes as input extended spectral-line observations from two tracers and provides an estimate of their spatial correlation across velocity channels.

We characterize HOG using synthetic observations of HI and ^{13}CO ($J = 1 \rightarrow 0$) emission from numerical simulations of magnetohydrodynamic (MHD) turbulence leading to the formation of molecular gas after the collision of two atomic clouds. We find a significant spatial correlation between the two tracers in velocity channels where $v_{\rm HI} \approx v_{\rm 13CO}$, almost independent of the orientation of the collision with respect to the line of sight.

Subsequently, we use HOG to investigate the spatial correlation of the HI, from The HI/OH/Recombination line survey of the inner Milky Way (THOR), and the ^{13}CO ($J = 1 \rightarrow 0$) emission from the Galactic Ring Survey (GRS), toward the portion of the Galactic plane $33.^{\circ}75 \leq l \leq 35.^{\circ}25$ and $|b| \leq 1.^{\circ}25$. We find a significant spatial correlation between the two tracers in extended portions of the studied region. Although some of the regions with high spatial correlation are associated with HI self-absorption (HISA) features, suggesting that it is produced by the cold atomic gas, the correlation is not exclusive to this kind of region.

The HOG results derived for the observational data indicate significant differences between individual regions: some show spatial correlation in channels around $v_{\rm HI} \approx v_{\rm 13CO}$ while others present spatial correlations in velocity channels separated by a few kilometers per second. We associate these velocity offsets to the effect of feedback and to the presence of physical conditions that are not included in the atomic-cloud-collision simulations, such as more general magnetic field configurations, shear, and global gas infall.

Key words. ISM: atoms – ISM: clouds – ISM: molecules – ISM: structure – galaxies: ISM – radio lines: ISM

⋆ Corresponding author, e-mail: soler@mpia.de

1. Introduction

Molecular clouds (MCs) are the main reservoir of cold gas from which stars are formed in the Milky Way and similar spiral galaxies (see for example, Bergin & Tafalla 2007; Dobbs et al. 2014; Molinari et al. 2014). Hence the study of the formation, evolution, and destruction of MCs is crucial for any understanding of the star formation process.

Much of the interstellar medium (ISM) in disk galaxies is in the form of neutral atomic hydrogen (HI), which is the matrix within which many MCs reside (Ferrière 2001; Dickey et al. 2003; Kalberla & Kerp 2009). Much of the HI is observed to be either warm neutral medium (WNM) with $T \approx 10^4$ K or cold neutral medium (CNM) with $T \approx 10^2$ K (Kulkarni & Heiles 1987; Dickey & Lockman 1990; Heiles & Troland 2003). The transition between the HI and the molecular gas is primarily driven by changes in the density and extinction (Reach et al. 1994; Draine & Bertoldi 1996; Glover & Mac Low 2011). Consequently, the first step for MC formation is the gathering of sufficient gas in one place to raise the column density above the value needed to provide effective shielding against the photodissociation produced by the interstellar radiation field (Krumholz et al. 2008, 2009; Sternberg et al. 2014). There are multiple processes that intervene in the accumulation of the parcels of gas out of the diffuse ISM to make dense MCs (for reviews see Hennebelle & Falgarone 2012; Klessen & Glover 2016, and references therein). However, despite the increasing number of models and observations, it is still unclear what are the dominant processes that lead to MC formation and what are the observational signatures with which to identify them.

Some of the MC formation mechanisms that have been proposed are converging flows driven by feedback or turbulence, agglomeration of smaller clouds, gravitational instability and magneto-gravitational instability, and instability involving differential buoyancy (see Dobbs et al. 2014, and references therein). Each one of these processes produces morphological and kinematic imprints over different spatial and time scales. Some are related to the spatial distribution of the atomic and molecular emission (e.g., Dawson et al. 2013), some are associated with the relative velocity (e.g., Motte et al. 2014) or the spatial correlation between these two components (e.g., Gibson et al. 2005; Goldsmith & Li 2005). However, most of these imprints remain to be discovered.

An idealized spherical cloud of diffuse gas and dust immersed in a bath of isotropic interstellar radiation begins to form an MC when the column density gets sufficiently high that the gas/dust can self-shield, the HI converts to H_2, and the ^{13}CO appears toward the center. In this ideal cloud, it is expected that the HI and ^{13}CO emission match at exactly the same velocities, but that is not necessarily the case for a real MC, where the density and velocity structures are much more complex, the spectra of both tracers are affected by optical depth and self-absorption, and the simple inspection of the emission lines may not be sufficient to assess the association between the atomic and the molecular gas. Yet, there is important information about the dynamics of the MC formation process encoded in the relation between the extended emission from both tracers.

To systematically study the density and velocity information in extended spectral line observations and characterize the imprint of MC formation scenarios in numerical simulations, we introduce the histogram of oriented gradients (HOG), a technique developed for machine vision that we employ to study the spatial correlation between different tracers of the ISM. In a nutshell, HOG takes as input extended spectral line observations from two ISM tracers and provides an estimate of their spatial correlation across velocity channels. We use HOG to study three aspects of the correlation between atomic and molecular gas. First, we evaluate if there is a spatial correlation between the two tracers, which would indicate the relation between the MC and its associated atomic gas. Second, we evaluate the distribution of such a spatial correlation across velocity channels, which can reveal details about the kinematics of both gas phases. Third, we compare the spatial correlation and its distribution across velocity channels in different regions and evaluate if they are similar to the synthetic observations of one of the multiple MC formation scenarios.

In this work, we characterize HOG using a set of synthetic HI and ^{13}CO($J = 1 \rightarrow 0$) emission observations obtained from the numerical simulation of magnetohydrodynamic (MHD) turbulence and MC formation in the collision of two atomic clouds presented in Clark et al. (2018). Then, we apply HOGs to the observations of the 21-cm HI emission, from the HI/OH/Recombination line survey of the inner Milky Way (THOR, Beuther et al. 2016) and the ^{13}CO($J = 1 \rightarrow 0$) emission, from the Galactic Ring Survey (GRS, Jackson et al. 2006), toward a selected portion of the Galactic plane. Finally, we detail the results of HOG toward some of the MC candidates identified in the GRS observations presented in Rathborne et al. (2009). All of the routines used for the HOG analysis presented in this paper, including the example presented in Fig. 1 and other illustrative cases, are publicly available at `http://github.com/solerjuan/astrohog`.

This paper is organized as follows. Section 2 describes our implementation of the HOG technique. Section 3 presents the characterization of HOG using the colliding flow simulations. Section 4 introduces the HI and ^{13}CO($J = 1 \rightarrow 0$) observations used for this study. We report the results of the HOG analysis of the observations in Sec. 5. We discuss the origin of the spatial correlations and the MC characteristics revealed by HOG in Sec. 6. Finally, Sec. 7 presents our main conclusions and the future prospects of this approach. We reserve the technical details of the HOG technique for a set of appendices. Appendix A describes details of the HOG method, such as the calculation of the gradient and the circular statistics used to evaluate the HOG results. Appendix B presents a series of tests of the statistical significance of the HOG method. Finally, Appendix C presents further analysis of the synthetic observations of MHD simulations.

2. The histogram of oriented gradients

The histogram of oriented gradients (HOG) is a feature descriptor used in machine vision and image processing for object detection and image classification processes (McConnell 1986; Leonardis et al. 2006). A feature descriptor is a representation of an image or an image patch that simplifies the image by extracting one or more characteristics. In the case of HOG, the method is based on the assumption that the local appearance and shape of an object in an image can be well characterized by the distribution of local intensity gradients or edge directions, which are by definition perpendicular to the direction of the gradient. The HOG method is widely applied in the detection of objects in a variety of applications such as recognition of hand gestures (Freeman & Roth 1994), detection of humans (Zhu et al. 2006), and use of sketches for searching and indexing digital image libraries (Hu et al. 2010).

One of the simplest applications of the HOG method is quantifying the spatial correlation between two images. The HOG is

a representation of the occurrences of the relative orientations between local gradient orientations in the two images, thus it is a representation of how the edges in the images match each other. Given that we are interested in evaluating the correlation between observations of astronomical objects through different tracers, we do not need to match the scales of the images or assume a prior on the shape of the objects that we are investigating.

Although the maps of extended atomic and molecular emission are not dominated by sharp edges, the HOG systematically characterizes and correlates the intensity contours that human vision recognizes as their main features, such as clumps or filaments. We do not assume any physical interpretation for the origin of the velocity-channel map gradients, as it is the case in other gradient methods, such as those presented in the family of papers represented by Lazarian & Yuen (2018). We use the velocity-channel map gradients to compare systematically the intensity contours that might be common to two ISM tracers.

An application of HOG has been previously introduced in astronomical research in the study of the correlation between the column density structures and the magnetic field orientation in both synthetic observations of simulations of MHD turbulence and *Planck* polarization observations (Soler et al. 2013; Planck Collaboration Int. XXXV 2016). Other potential applications of the HOG technique in astronomy include, for example, characterizing the directionality of structures in an astronomical image, evaluating the morphological changes across velocity channels in a single PPV cube, and, in general, quantifying the spatial correlation between two or more ISM tracers.

2.1. Using HOG to quantify correlations between position-position-velocity cubes.

We use the HOG method to quantify the spatial correlation between maps of HI and ^{13}CO emission across radial velocities, better known in astronomy as position-position-velocity (PPV) cubes. Explicitly, we calculate the correlation between the two PPV cubes by following the steps described below.

2.1.1. Computation of the HOG

We align and re-project a pair of PPV cubes, $I^{\rm A}_{i,j,l}$ and $I^{\rm B}_{i,j,m}$, into a common spatial grid by using the `reproject` routine included in the `Astropy` package (Astropy Collaboration et al. 2013). Throughout this paper, the indexes i and j correspond to the spatial coordinates, Galactic longitude and latitude, and the indexes l and m correspond to the velocity channels in the respective PPV cube. Given that we are comparing the spatial gradients of each velocity channel map, the HOG technique does not require the same velocity resolution in the PPV cubes. For a pair of velocity-channel maps $I^{\rm A}_{i,j,l}$ and $I^{\rm B}_{i,j,m}$, we calculate the relative orientation angle ϕ between intensity gradients by evaluating

$$\phi_{i,j,lm} = \arctan\left(\frac{(\nabla I^{\rm A}_{i,j,l} \times \nabla I^{\rm B}_{i,j,m}) \cdot \hat{z}}{\nabla I^{\rm A}_{i,j,l} \cdot \nabla I^{\rm B}_{i,j,m}}\right), \quad (1)$$

where the differential operator ∇ corresponds to the gradient. The term $(\nabla I^{\rm A}_{i,j,l} \times \nabla I^{\rm B}_{i,j,m}) \cdot \hat{z} \equiv |\nabla I^{\rm A}_{i,j,l}||\nabla I^{\rm B}_{i,j,m}| \sin\phi_{i,j,lm}$ is the z-axis projection of the cross product. The term $\nabla I^{\rm A}_{i,j,l} \cdot \nabla I^{\rm B}_{i,j,m} \equiv |\nabla I^{\rm A}_{i,j,l}||\nabla I^{\rm B}_{i,j,m}| \cos\phi_{i,j,lm}$ is the scalar product of vectors, or dot product. We choose the representation in Eq. 1 because it is numerically better-behaved than the expression that would be obtained by using just the dot product and the arccos function. Equation 1 implies that the relative orientation angles are in the range $[-\pi/2, \pi/2)$, thus accounting for the orientation of the gradients and not their direction. The value of ϕ is only meaningful in regions when both $\nabla I^{\rm A}_{i,j,l}$ and $\nabla I^{\rm B}_{i,j,m}$ are significant, that is, their norm is greater than zero or above thresholds that are estimated according to the noise properties of the each PPV cube.

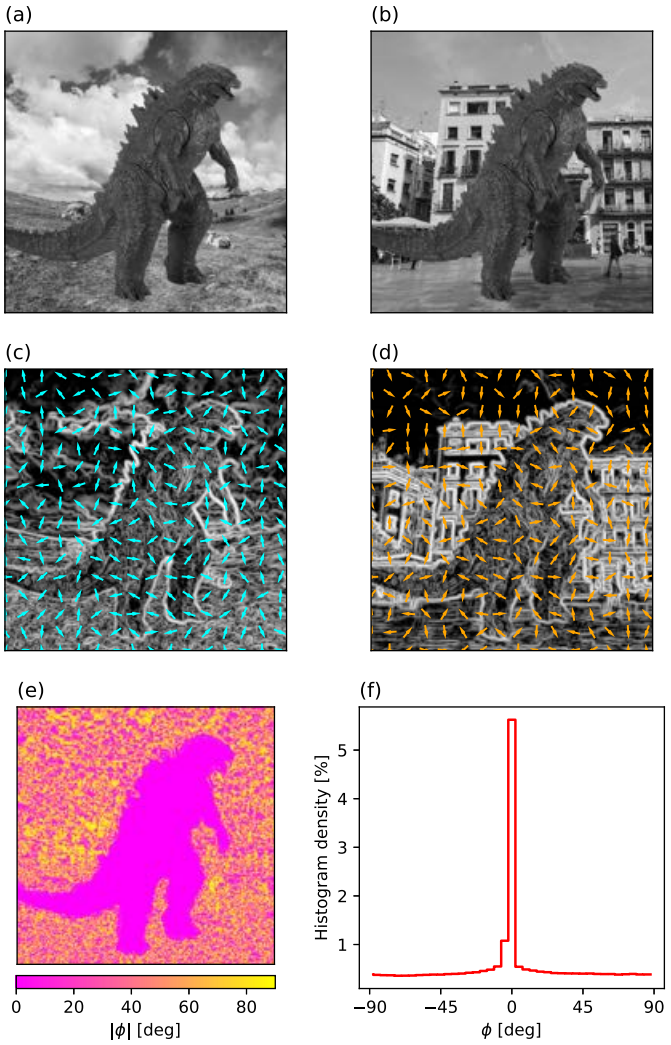

Fig. 1: Illustration of the histogram of oriented gradients (HOG) method. A pair of images, (*a*) and (*b*), are characterized by the norm and the orientation of their gradients, (*c*) and (*d*). The relative orientation angles between their gradients, (*e*), are summarized in the histogram of oriented gradients, (*f*). The number of histogram counts around 0° corresponds to the coincidence of gradient orientations in both images, which is quantified by using the mean resultant vector, Eq. (2), and the projected Rayleigh statistic, Eq. (3). Two completely uncorrelated images would produce a flat histogram while two identical images would produce a histogram having the form of a Dirac delta function centered at 0°.

We compute the gradients using Gaussian derivatives, explicitly, by applying the multidimensional Gaussian filter routines in the `filters` package of `Scipy`. The Gaussian derivatives are the result of the convolution of the image with the spatial derivative of a two-dimensional Gaussian function. The width of the Gaussian determines the area of the vicinity over which the gradient is calculated. Varying the width of the Gaussian kernel enables the sampling of different scales and re-

duces the effect of noise in the pixels (see Soler et al. 2013, and references therein).

For the sake of clarity, we illustrate the aforementioned procedure in a pair of mock velocity-channel maps presented in Fig. 1. We there present the two velocity-channel maps, panels (a) and (b); their corresponding gradients, panels (c) and (d); the relative orientation angles, ϕ, panel (e); and the histograms of oriented gradients, panel (f), which we evaluate by using the tools of circular statistics presented in the next section.

2.1.2. Evaluation of the correlation

Once we calculate the relative orientation angles $\phi_{i,j,lm}$ for a pair of channels centered on velocities $v_l^{\rm A}$ and $v_m^{\rm B}$, we summarize the spatial correlation contained in these angles by marginalizing over the spatial coordinates, indexes i and j. For that purpose, we use two tools from circular statistics: the mean resultant vector (r) and the projected Rayleigh statistic (V), both described in detail in Appendix A.3.

In our application, we use the definition of the mean resultant vector

$$r_{lm} = \frac{\left(\left[\sum_{ij} w_{ij,lm}\cos(2\phi_{ij,lm})\right]^2 + \left[\sum_{ij} w_{ij,lm}\sin(2\phi_{ij,lm})\right]^2\right)^{1/2}}{\sum_{ij} w_{ij,lm}}, \tag{2}$$

where the indexes i and j run over the pixel locations in the two spatial dimensions and $w_{ij,lm}$ is the statistical weight of each angle $\phi_{ij,lm}$. We account for the spatial correlations introduced by the telescope beam by choosing $w_{ij,lm} = (\delta x/\Delta)^2$, where δx is the pixels size and Δ is the diameter of the derivative kernel that we use to calculate the gradients. For pixels where the norm of the gradient is negligible or can be confused with the signal produced by noise, we choose $w_{ij,lm} = 0$ (see Appendix A for a description of the gradient selection).

The mean resultant vector, r, is a descriptive quantity that can be interpreted as the percentage of vectors pointing in a preferential direction. However, it does not provide any information on the shape of the angle distribution. The optimal statistic to test if the distribution of angles is non-uniform and peaked at 0° is the projected Rayleigh statistic

$$V_{lm} = \frac{\sum_{ij} w_{ij,lm}\cos(2\phi_{ij,lm})}{\sqrt{\sum_{ij}[(w_{ij,lm})^2/2]}}, \tag{3}$$

which follows the same conventions introduced in Eq. (2). Each value V_{lm} represents the likelihood test against a von Mises distribution, which is the circular normal distribution centered on 0°, or in other words, the likelihood that the gradients of the emission maps $I^{\rm A}_{ij,l}$ and $I^{\rm B}_{ij,m}$ are mostly parallel. The ensemble of V_{lm} values, which we denominate the correlation plane, represents the correlation between the emission maps centered on velocities $v_l^{\rm A}$ and $v_m^{\rm B}$. For the sake of simplicity, we designate the HOG correlation between tracers A and B as $V(v^{\rm A}, v^{\rm B})$, but this is just an approximation given that we can only estimate the discrete values of V_{lm}, which depend on the spectral resolution of the observations and the width of the velocity channels.

We present the results of our analysis in terms of both V and r. The values of the latter are only meaningful for our purposes when they are validated by V; because large values of the mean resultant vector only indicate a preferential orientation, not necessarily $\phi = 0°$. We note that the gradient vectors in each individual velocity-channel map are not statistically independent, that is, even if the observations were made with infinite angular resolution, the physical phenomena governing the ISM; that is, gravity, turbulence, and the magnetic fields; impose correlations across multiple spatial scales. And so it is not possible to draw conclusions from the values of V alone, but its statistical significance should be assessed by comparing its value to the values obtained in maps with similar statistical properties.

Given the difficulties in reproducing the statistical properties of each velocity-channel map, we use the mean value,

$$\langle V\rangle \equiv \frac{\sum_l^{[l_{\rm min},l_{\rm max}]}\sum_m^{[m_{\rm min},m_{\rm max}]} V_{lm}}{(l_{\rm max}-l_{\rm min})(m_{\rm max}-m_{\rm min})}, \tag{4}$$

and the population variance,

$$\varsigma_V^2 \equiv \frac{\sum_l^{[l_{\rm min},l_{\rm max}]}\sum_m^{[m_{\rm min},m_{\rm max}]} (V_{lm}-\langle V\rangle)^2}{(l_{\rm max}-l_{\rm min})(m_{\rm max}-m_{\rm min})}, \tag{5}$$

in the velocity ranges defined by the indexes $[l_{\rm min}, l_{\rm max}]$ and $[m_{\rm min}, m_{\rm max}]$ to assess the statistical significance of V. If we assume that most of the channel maps in a particular velocity range are uncorrelated, each of them would correspond to an independent realization of a scalar field with a spatial correlation given by the properties of the ISM and the angular resolution of the observations, ς_V would represent the chance correlation between those maps. Evidently that is not the case in reality, unless we consider channels separated by tens of km s^{-1} in a Galactic target, but still ς_V characterizes the V population variance within the selected range of velocities. There is of course a variance of V for each particular pair of velocity-channel maps, $(\sigma_V)_{lm}$, but it is in most cases smaller than ς_V, as shown in App. A.3.2.

In this work, we report the values of V_{lm} always in relation with the corresponding ς_V, as inferred from Eq. (5), in a particular velocity range. An alternative method for evaluating the statistical significance of V is based on estimating the population variance using velocity-channel maps that are uncorrelated by construction, for example, two PPV cubes that are not coincident in the sky or one PPV cube flipped with respect to the other in one of the spatial coordinates. This method is crucial for determining the validity of our method since the values of V in the cases mentioned above should be exclusively dominated by chance correlation, as we show in App. B.3. But the direct estimation of ς_V using these null-tests is computationally demanding and does not lead to significant differences with respect to the values obtained with Eq. 5.

3. HOG analysis of MHD simulations

We characterize HOG by analyzing a set of synthetic observations of HI and ^{13}CO emission from the numerical simulations of MC formation in a colliding flow presented in Clark et al. (2018). These simulations include a simplified treatment of the chemical and thermal evolution of the interstellar medium (ISM), which makes them well suited for obtaining synthetic observations of both tracers. Although the numerical setup and the chemistry treatment are not indisputable (see for example, Levrier et al. 2012), we use this simplified physical scenario to gain insight into the behavior of the HOG technique before we apply it to the observations.

3.1. Initial conditions

The simulations considered were carried out using the AREPO moving mesh code (Springel 2010). They represent two 38-pc-diameter atomic clouds with an initial particle density

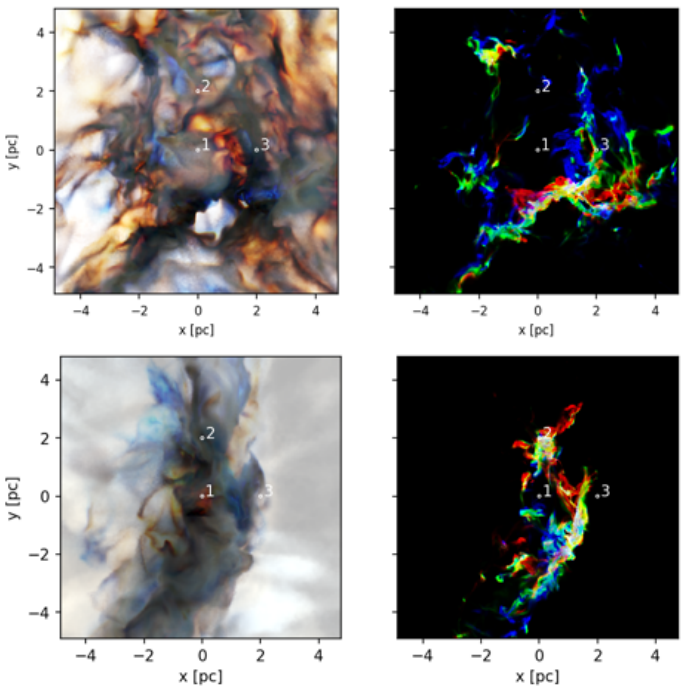

Fig. 2: Synthetic observations of HI with a 100 K background (left) and ^{13}CO (right) emission from the MHD turbulence simulation of two colliding atomic clouds presented in Clark et al. (2018). In both maps the colors represent the integrated intensities in three groups of velocity channels: red for $-0.6 \leq v_{\rm LSR} \leq -0.2\,{\rm km\,s^{-1}}$, green for $-0.2 \leq v_{\rm LSR} \leq 0.2\,{\rm km\,s^{-1}}$, and blue for $0.2 \leq v_{\rm LSR} \leq 0.6\,{\rm km\,s^{-1}}$. The numbers correspond to the positions of the spectra presented in Fig. 3. The top and bottom panels correspond to the synthetic observations made with the line of sight parallel (face-on) and perpendicular to the collision axis (edge-on), respectively.

$n_0^{\rm C} = 10\,{\rm cm^{-3}}$ that collide head-on along the x-axis of the simulation domain at $7.5\,{\rm km\,s^{-1}}$ with respect to each other. The clouds are given a turbulent velocity field with a $1\,{\rm km\,s^{-1}}$ amplitude and a $P(k) \propto k^{-4}$ scaling law. The simulation includes a uniform initial magnetic field $B_0 = 3\,\mu$G oriented along the x-axis, that is, parallel to the collision axis.

The clouds are initially set one cloud radius apart (19 pc) in a cubic computational domain of side 190 pc and initial number density $n_0^{\rm S} = 0.1\,{\rm cm^{-3}}$. The boundaries of the box are periodic, but self-gravity is not periodic. The initial cell mass is approximately $5 \times 10^{-3}\,{\rm M_\odot}$, both in the clouds and in the low-density surrounding medium. The cell refinement is set such that the thermal Jeans length is resolved by at least 16 AREPO cells at all times.

The simulations follow the thermal evolution of the gas using a cooling function based on Glover et al. (2010) and Glover & Clark (2012). The chemical evolution of the gas is modelled using a simplified H-C-O network based on Glover & Mac Low (2007) and Nelson & Langer (1999), updated as described in Glover et al. (2015). The effects of H_2 self-shielding and dust shielding are accounted for using the TreeCol algorithm (Clark et al. 2012).

The metallicity of the gas is taken to be solar with elemental abundances of oxygen and carbon set to $x_{\rm O} = 3.2 \times 10^{-4}$ and $x_{\rm C} = 1.4 \times 10^{-4}$ (Sembach et al. 2000). The three simulations presented in Clark et al. (2018) are designed to probe the effect of different interstellar radiation fields (ISRFs) and cosmic rate ionization rates (CRIRs). For the characterization of HOG we have chosen the simulation with ISRF $G_0 = 17$ and CRIR $= 3 \times 10^{-16}\,{\rm s^{-1}}$. This ISRF implies that the H_2 and the CO are found at higher column densities than in the other two simulations presented in Clark et al. (2018), but it does not imply any loss of generality in our results.

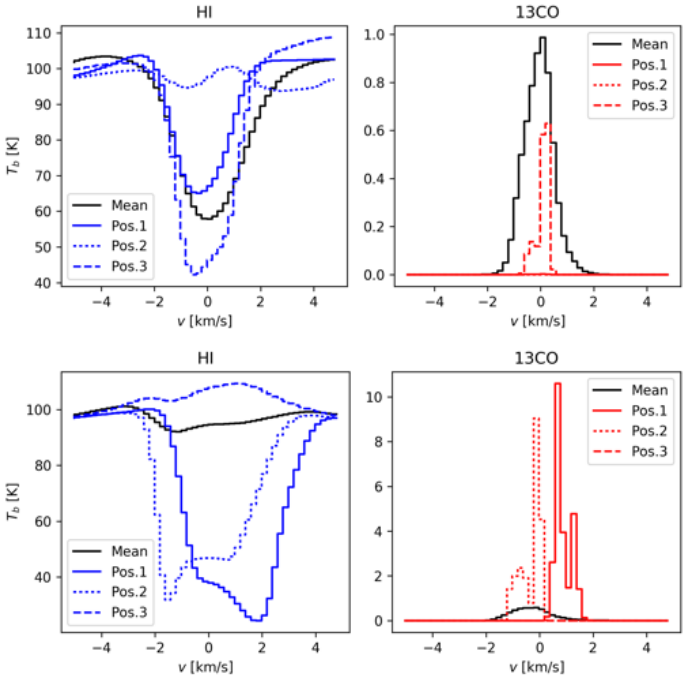

Fig. 3: Spectra from the synthetic observations of HI with a 100 K background (left) and ^{13}CO (right) emission presented in Fig. 2. The black lines correspond to the average spectra over the whole map. The solid, dashed and segmented colored lines correspond to the spectra toward the positions indicated in Fig. 2. The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

3.2. Synthetic observations

The radiative transfer (RT) post-processing of the simulations was made using the RADMC-3D code[1] following the procedures described in Clark et al. (2018). In brief, the HI emission is modelled assuming that the hyperfine energy levels are in local thermodynamic equilibrium (LTE), with a spin temperature $T_{\rm s}$ equal to the local kinetic temperature of the gas. This is a good approximation for the cold, dense atomic gas that dominates the emission signal in these simulations (e.g., Liszt 2001). For the ^{13}CO, we do not assume LTE, as some of the emission may be coming from regions with densities below the CO critical density. Instead, we use the large velocity gradient (LVG) module implemented in RADMC-3D by Shetty et al. (2011). In addition, as the Clark et al. (2018) simulations do not track ^{13}CO explicitly, it is necessary to compute the ^{13}CO abundance based on the ^{12}CO abundance. This is done using a fitting function for the ^{13}CO/^{12}CO ratio as a function of the ^{12}CO column density proposed by Szűcs et al. (2014). This column-density-dependent

[1] http://www.ita.uni-heidelberg.de/~dullemond/software/radmc-3d/

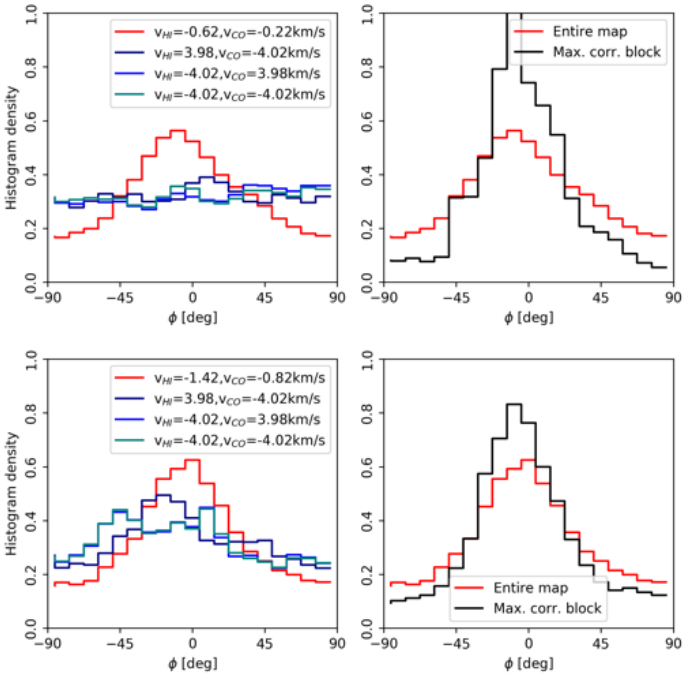

Fig. 4: *Left*. Histograms of oriented gradients (HOGs) corresponding to the pair of velocity-channel maps with the largest spatial correlation, as inferred from the V values shown in Fig. 6, and three pairs of arbitrarily selected velocity channels in the synthetic observations presented in Fig. 2. *Right*. For the pair of velocity-channel maps with the largest spatial correlation, HOGs corresponding to the entire map and just the block with the largest V indicated in Fig. 5. The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

conversion factor accounts for the effects of chemical fractionation and selective photodissociation of ^{13}CO and hence is more accurate than adopting a constant ^{13}CO/^{12}CO ratio.

The AREPO results are interpolated onto a regular cartesian grid. The grid covers a cubic region of 9.72 pc with 400 cells per side, corresponding to a spatial resolution of 0.024 pc. The synthetic spectra are initially calculated in 500 velocity channels covering the velocity range $[-5,5]$ km s^{-1}. We resample this original data into a velocity resolution of 0.2 km s^{-1} to match the channel width of the GRS data. The maps of the synthetic observations of HI and ^{13}CO and some selected corresponding spectra are presented in Fig. 2 and Fig. 3, respectively.

It is common at low Galactic latitudes that cold foreground clouds absorb the emission from gas behind. This effect is often called HI self-absorption (HISA), although it is not self-absorption in the normal radiative transfer sense, because the absorbing cloud may be spatially distant from the background HI emission, but sharing a common radial velocity (Gibson et al. 2005; Kavars et al. 2005). For that reason we use synthetic observations of HI that include a 100 K background emission. For the sake of completeness and discussion, we present the synthetic observations of HI without background emission in Appendix C.

We analyze two configurations of the aforementioned simulation: one with the line of sight parallel to the collision axis (face-on) and one with line of sight perpendicular to the collision axis (edge-on). Fig. 2 shows the clear differences between the two configurations. In the face-on configuration, the HI is distributed over the whole map in filamentary structures that appear dark against the bright background while the ^{13}CO appears more concentrated, but also filamentary in appearance. In the edge-on configuration, the HI appears concentrated in the shocked layer, which is clearly visible against the bright background, and the ^{13}CO is distributed in a couple of filamentary structures.

The spectra of the face-on and the edge-on synthetic observations, shown in Fig. 3, reveal two clear differences between these configurations. The face-on configuration presents a broad HI mean spectrum in absorption against the 100 K background and clearly centered at $v_{\rm LOS} \approx 0$ km s^{-1}. The ^{13}CO is also clearly centered at $v_{\rm LOS} \approx 0$ km s^{-1}. The edge-on configuration presents a flat HI mean spectrum at 100 K, resulting from the background emission that is dominant in most of the map, and absorption spectra with peaks at $v_{\rm LOS} \approx -2$ and 2 km s^{-1}. These two peaks are most likely the result of momentum conservation in the shocked layer, as we discuss in more detail in the next section. The ^{13}CO is clearly centered at $v_{\rm LOS} \approx 0$ km s^{-1}.

3.3. HOG analysis results

We run the HOG analysis of the two sets of synthetic observations (face-on and edge-on) following the procedure described in Sec. 2. We use a derivative kernel with a 0.12 pc (5 pixels) FWHM. Given that the synthetic observations do not include noise, we consider all non-zero gradients in the synthetic HI and ^{13}CO PPV cubes.

Figure 4 shows the HOGs corresponding to a selection of HI and ^{13}CO velocity channels. One fact that is evident from the shape of the HOGs is that, at least for some pairs of velocity channels, the distribution of relative orientation angles is not flat and it clearly peaks at $\phi = 0°$; this indicates that in these channel pairs, the HI and ^{13}CO have contours that are aligned and the two tracers are morphologically correlated. This can be visually confirmed in the gradient plots of the velocity-channel pairs with the highest V values, presented in Fig. 5, where it is evident that the ^{13}CO emission contours are adjacent to the contours of regions with a relative decrease of HI emission.

The behavior of the relative orientation trend is better visualized in the values of the mean resultant vector length r, defined in Eq. (2), and the projected Rayleigh statistic V, defined in Eq. (3), for all pairs of HI and ^{13}CO channels in the velocity range $-4.8 \leq v_{\rm LOS} \leq 4.8$ km s^{-1}, presented in Fig. 6. The distribution of V and r shows that the maximum spatial correlation between the HI and ^{13}CO emission appears at the same velocity in the two tracers, that is, along the diagonal of the correlation plane, where $v_{\rm HI} \approx v_{\rm 13CO}$. This observation is not entirely unexpected; if one considers the standard picture of a quiescent MC and its associated atomic envelope and the atomic gas and the molecular gas move together, then, the two tracers should appear approximately at the same velocity. However, it is worth remarking that this correlation indicates that the contours of the emission of the two tracers match across multiple velocity channels. This behavior is not exclusive to the case of HI with an emission background and ^{13}CO, it can also be seen when applying HOG to the analysis of synthetic observations of HI without background emission and ^{13}CO emission, as shown in Appendix C.

The HOG, however, does not reveal an unambiguous difference between the signals produced by the observations in the face-on and the edge-on configurations. To zeroth order, HOG is revealing the spatial coincidence of the two tracers, which does seem to be significantly affected by the orientation of the colliding flows with respect to the line of sight. In more detail, the face-on configuration presents homogenous high V values along $v_{\rm HI} \approx v_{\rm 13CO}$ in the velocity range $-2.5 < v_{\rm LOS} < 2.5$ km s^{-1},

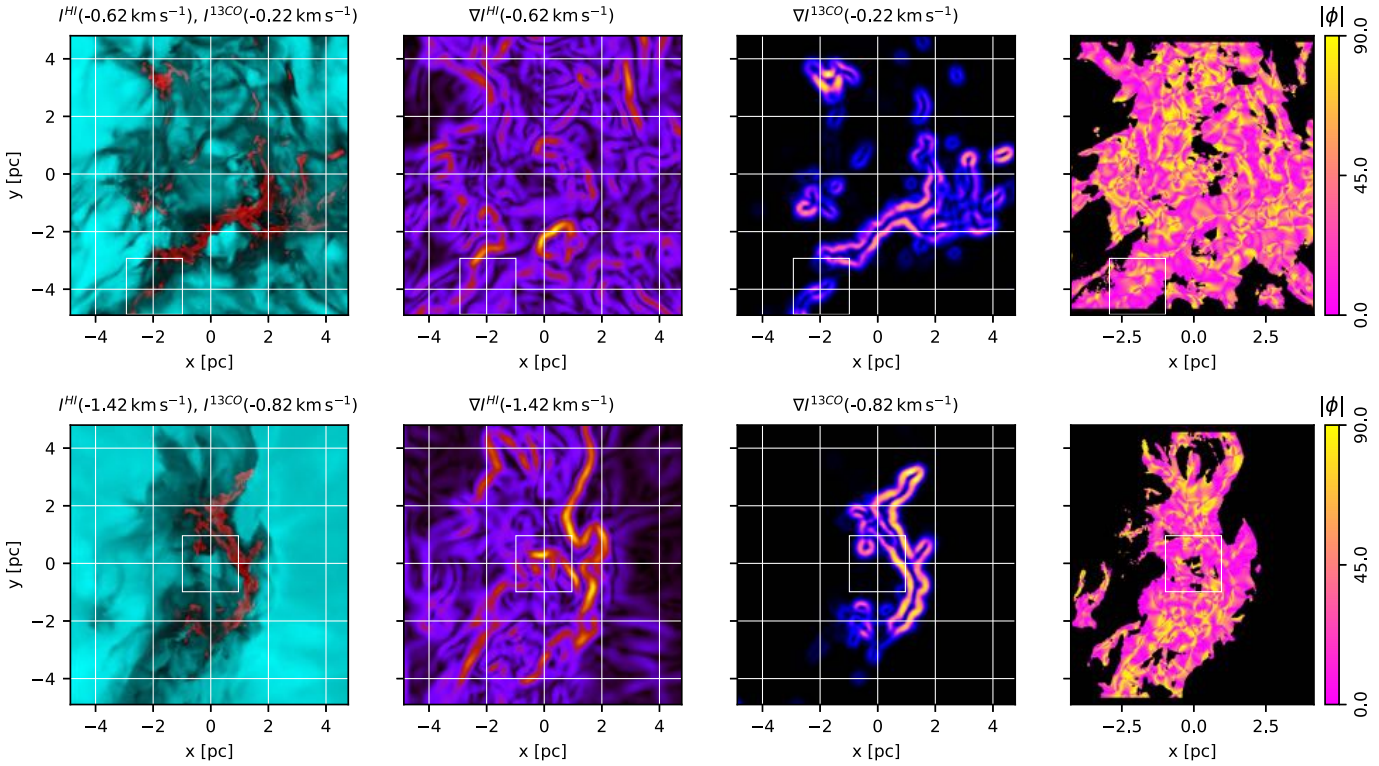

Fig. 5: Intensity, intensity gradients, and relative orientation angle maps from the synthetic observations presented in Fig. 2. *Left.* HI (teal) and ^{13}CO emission (red) in the velocity channels with the largest spatial correlation, as inferred from the V values shown in Fig. 6. *Middle left.* Norm of the gradient of the HI intensity map in the indicated velocity channel. *Middle right.* Norm of the gradient of the ^{13}CO intensity map in the indicated velocity channel. *Right.* Relative orientation angle ϕ, Eq. (1), between the gradients of the HI and ^{13}CO intensity maps in the indicated velocity channels. The white color in the ϕ map corresponds to areas with no significant gradient in either tracer. The square indicates the block, selected from a 7×7 spatial grid, with the largest values of V. The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

while in the edge-on configuration the high V values seem group around $v_{\rm LOS} \approx -2.0$ and $2.0\,{\rm km\,s^{-1}}$, but also close to $v_{\rm HI} \approx v_{\rm 13CO}$. In both configurations, these trends are produced by approximately 30% of the gradient pairs, as inferred from the values of r.

The difference between V in the face-on and the edge-on configurations can be understood in terms of the dynamics imposed by the colliding flow. In the face-on case, the ram pressure constrains both the cold HI and ^{13}CO to remain close to $v_{\rm LOS} \approx 0\,{\rm km\,s^{-1}}$. The molecular gas formed in the shocked interface does not inherit the structure of the colliding atomic clouds; consequently, we do not see high spatial correlation between the ^{13}CO at $v_{\rm LOS} \approx 0\,{\rm km\,s^{-1}}$ and the HI at the velocities of the colliding clouds. Given that the shocked interface is relatively thin, there is not much overlap of structures along the line of sight, which most likely explains the tight correlation around $v_{\rm HI} \approx v_{\rm 13CO}$ shown in the top panels of Fig. 6.

In the edge-on case, we are looking at the shocked interface in the direction that is not directly constrained by the ram pressure, where the parcels of HI and ^{13}CO have developed line-of-sight motions that are independent from the proper motion of the parental atomic clouds. In contrast with the face-on case, the larger values of V are centered on $v_{\rm LOS} \approx 2\,{\rm km\,s^{-1}}$, most likely due to the proper motion of the most dominant parcel of ^{13}CO formed in the shocked interface. The overlap of structures along the line of sight in the edge-on shocked interface is most likely producing the dispersion of high V values across velocity channels, however, it is still closely concentrated around $v_{\rm HI} \approx v_{\rm 13CO}$.

The HOG analysis reveals that the HI and the ^{13}CO emission from a colliding flow simulation appear morphologically correlated at roughly the same velocity, independently of the orientation of the primary flow with respect to the line of sight, which is both disappointing and encouraging. On the one hand, this implies that the results of the HOG analysis cannot unambiguously differentiate orientations of the cloud collision with respect to the line of sight. On the other hand, this implies that the HOG signal produced by the atomic cloud collision is not greatly affected by the orientation of the primary flow with respect to the line of sight and the HOG can be used to quantify any departures from this simple scenario. These departures are evident when we apply HOG to the observations.

4. Observations

For this first application of the histogram of oriented gradients (HOG) method, we choose the THOR HI and GRS ^{13}CO observations toward the portion of the Galactic plane defined by $33.^{\circ}75 < l < 35.^{\circ}25$ and $|b| < 1.^{\circ}25$, which are shown in Fig. 7 and Fig. 8. Given the need to describe the method in detail, we focus our analysis on this region because it contains a large diversity of objects, such as supernova remnants (SNR), HII regions, HI Self-Absorption (HISA) features (Bihr 2016; Wang et al. In prep.), and a wealth of MCs that have been identified in emission from ^{12}CO and ^{13}CO (Miville-Deschênes et al. 2017; Rathborne et al. 2009, respectively). We reserve the application of the HOG

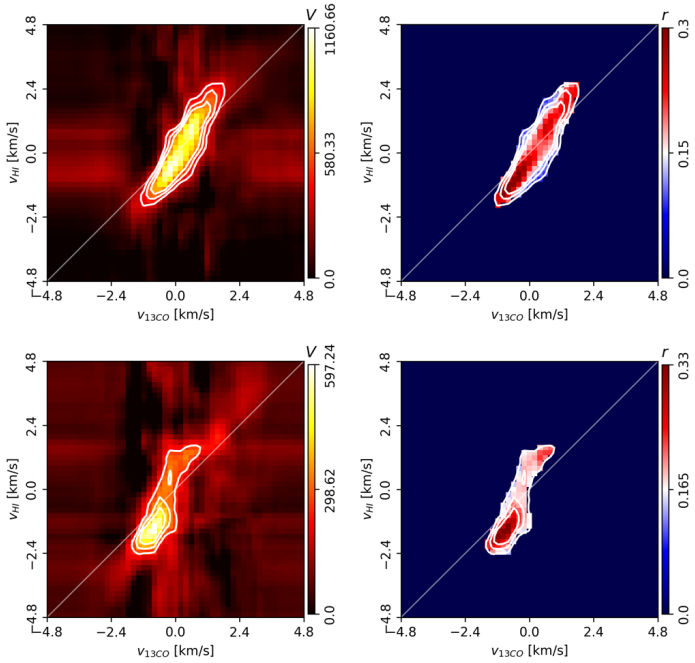

Fig. 6: Results of the HOG analysis of the HI and ^{13}CO synthetic observations presented in Fig. 2. *Left.* Projected Rayleigh statistic, $V(v_{13CO}, v_{HI})$, the HOG statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps, defined in Eq. (3). The contours indicate the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ levels in the corresponding velocity range. *Right.* Mean resultant vector length, $r(v_{13CO}, v_{HI})$, within the $3\varsigma_V$ confidence interval, a HOG metric that is roughly equivalent to the percentage of gradient pairs that imply the spatial correlation between the velocity-channel maps, defined in Eq. (2). The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

technique to the whole extent of both surveys for a subsequent publication (Soler et al. In prep.).

The selected region includes two SNRs that we identify using the catalogs presented in Green (2014) and Anderson et al. (2017). The most conspicuous of these two SNRs is Westerhout 44 (W44, Westerhout 1958), located around $[l, b] = [34.^\circ7, -0.^\circ4]$, which is shown in Fig. 7. Multi-wavelength observations of W44 show the presence of an elongated shell-like structure with a remarkable network of filaments and arcs across the face of this remnant suggesting the presence of shocked gas (Giacani et al. 1997; Reach et al. 2006).

The region also contains a plethora of HII regions, which we identify using the catalog produced using the *WISE* observations (Anderson et al. 2014). One of the most interesting objects in this catalog is the ultra-compact HII (UCHII) region G34.256+0.146, which produces a significant absorption feature that is clearly distiguishable in the HI longitude-velocity (LV) diagram presented in Fig. 8.

This region also includes portions of two giant molecular filaments (GMFs) in the sample presented in Ragan et al. (2014). First, 38.1-32.4a, a structure that extends across $33.^\circ4 \leq l \leq 37.^\circ1$ and $-0.^\circ4 \leq b \leq 0.^\circ6$ and is associated to ^{13}CO emission in the range $50 \leq v_{LSR} \leq 60$ km s^{-1}. Second, GMF38.1-32.4b, a structure that extends across $34.^\circ6 \leq l \leq 35.^\circ6$ and $-1.^\circ0 \leq b \leq 0.^\circ2$ and is associated to ^{13}CO emission in the range $43 \leq v_{LSR} \leq 46$ km s^{-1}.

4.1. Atomic hydrogen emission at 21 cm

We use the HI emission observations from The HI/OH/Recombination line survey of the inner Milky Way (THOR, Beuther et al. 2016). THOR comprises observations in eight continuum bands between 1 and 2 GHz made with Karl G. Jansky Very Large Array (VLA) in the C-array configuration covering the portion of the Galactic plane defined by $14.^\circ0 \leq l \leq 67.^\circ4$ and $|b| \leq 1.^\circ25$ at approximately 20″ resolution. As the survey name implies, the THOR frequency range includes the HI 21-cm emission line, four OH lines, and 19 Hα recombination lines.

The THOR HI data that are taken in C-array configuration are crucial for the study of absorption profiles against Galactic and extragalactic background sources. However, they do not recover the large-scale emission. In the present study, we use the data set resulting from the combination of the HI observations from THOR and the D-configuration VLA Galactic Plane Survey, (VGPS, Stil et al. 2006) combined with single-dish observations from the Green Bank Telescope (GBT).

The C-array configuration HI visibility data from the THOR survey were calibrated with the `CASA`[2] software package as described in Beuther et al. (2016). We used the multi-scale `CLEAN` routine in `CASA` to image the continuum-subtracted C-array configuration HI visibility together with the D-array configuration visibility from VGPS (Stil et al. 2006). We chose a pixel size of 4″, robust = 0.45, and a velocity resolution of 1.5 km s^{-1} in the velocity range $-50 \leq v_{LSR} \leq 150$ km s^{-1}. The resulting images were smoothed into a resolution of 40″ and feathered with the VGPS images (D+GBT) to recover the large-scale structure. Further details on the data reduction and imaging procedure are described in Beuther et al. (2016). The public release of this new HI data product is forthcoming (Wang et al. In prep.).

4.2. Carbon monoxide (CO) emission

We compare the HI emission observations with the ^{13}CO($J = 1 \rightarrow 0$) observations from The Boston University-Five College Radio Astronomy Observatory Galactic Ring Survey (GRS, Jackson et al. 2006). The GRS survey has 46″ angular resolution with an angular sampling of 22″. In this particular region, it covers the range $-5 \leq v_{LSR} \leq 135$ km s^{-1} at a resolution of 0.21 km s^{-1}. It has a typical root mean square (RMS) sensitivity of 0.13 K. We also make use of the catalog of MC and clump candidates identified in the GRS data (Rathborne et al. 2009).

We use ^{13}CO rather than ^{12}CO to minimize optical depth effects and facilitate the interpretation of the HOG analysis. The ^{12}CO emission is widespread towards the Galactic plane, just like HI, and only around 14% of the molecular gas mass traced by ^{12}CO emission is identified as part of molecular clouds in ^{13}CO (Roman-Duval et al. 2016). Compared to ^{12}CO, the ^{13}CO molecule is approximately 50 times less abundant and, thus, has a much lower optical depth (Wilson & Rood 1994). As a result, ^{13}CO is a much better tracer of column density and suffers less from line blending and self-absorption.

5. HOG analysis of observations

We apply the HOG analysis to the data products described in Sec. 4 using the method described in Sec. 2. We compute HOG exclusively using gradients that satisfy $I_{i,j,k} \geq 5I^N$ and $|\nabla I_{i,j,k}| \geq 5|\nabla I^N|$, where the noise intensity, I^N, and the noise

[2] https://casa.nrao.edu/

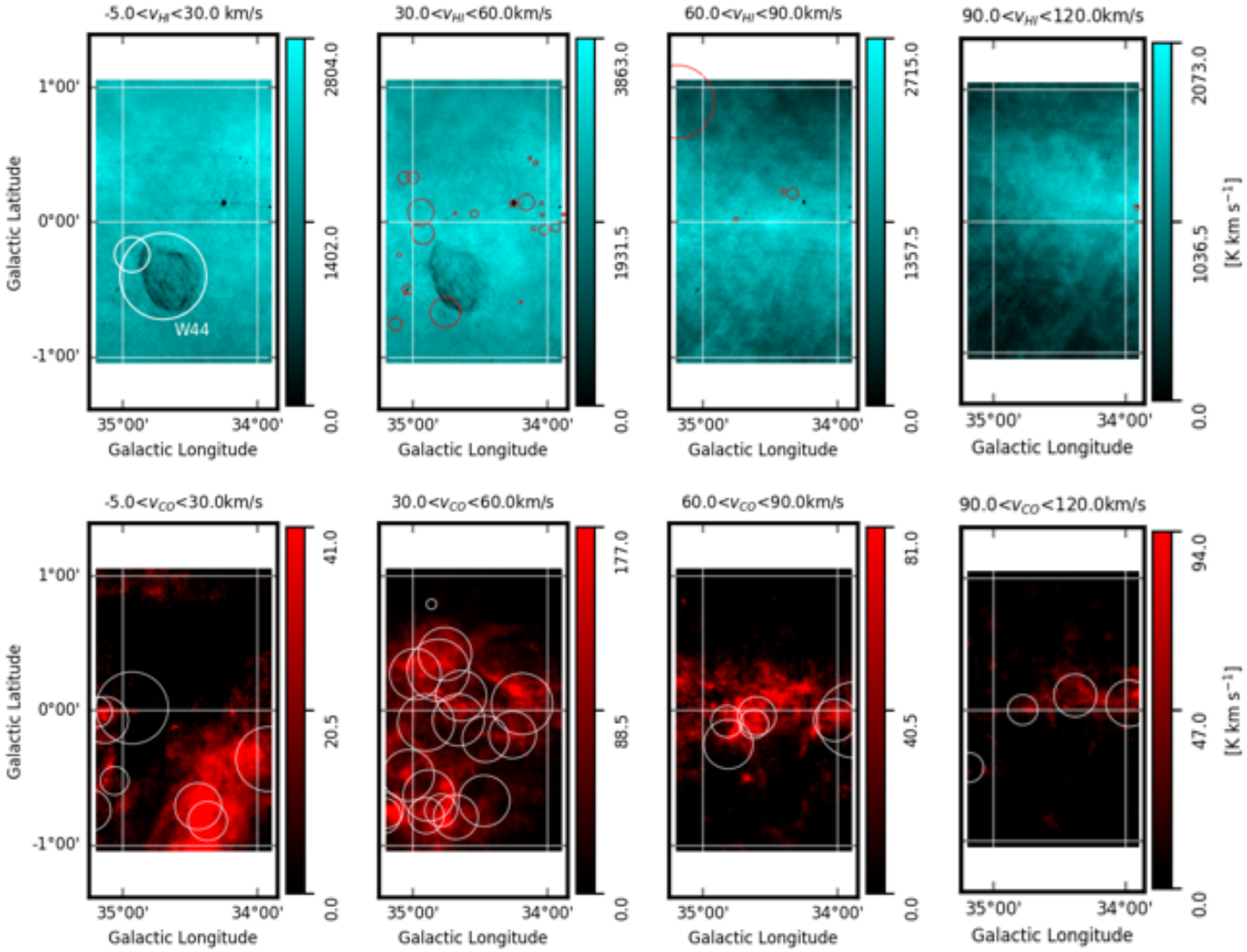

Fig. 7: Maps of the THOR HI and GRS ^{13}CO observations. *Top*. HI integrated emission from the THOR+VGPS+GBT observations (Beuther et al. 2016) in the indicated velocity ranges. The white circles correspond to the positions and effective sizes of identified supernovae remnants (Anderson et al. 2017; Green 2014) in each velocity range. The red circles correspond to the positions and effective sizes of identified HII regions (Anderson et al. 2014) in each velocity range. *Bottom*. ^{13}CO integrated emission from the GRS observations (Jackson et al. 2006) in the indicated velocity ranges. The white circles correspond to the positions and effective sizes of the molecular cloud candidates from the Rathborne et al. (2009) catalog in the indicated velocity ranges.

gradient norm, $|\nabla I^{\rm N}|$, are estimated following the procedure presented in Appendix A.2. Here we present and discuss the results obtained using a derivative kernel with a 90″ FWHM. This selection does not imply any loss of generality as described in Appendix A.4, where we discuss the results of using different derivative kernel sizes. The selection of $I_{i,j,k} \geq 5I^{\rm N}$ instead of $I_{i,j,k} \geq 3I^{\rm N}$ does not critically change the results of this analysis, as illustrated in Appendix B.1.

Figure 9 presents the values of the projected Rayleigh statistic, V, and the mean resultant vector length, r, corresponding to the HI and ^{13}CO emission for the velocity range $-5 \leq v_{\rm LSR} \leq 120\,{\rm km\,s^{-1}}$. It is clear from Fig. 9 that the spatial correlation between the HI and ^{13}CO emission is significant at the same velocity in the two tracers, that is, at $v_{\rm HI} \approx v_{\rm 13CO}$ or equivalently, along the diagonal of the V- and r-plane. As discussed in the previous section, if one considers a toy quiescent MC and its respective atomic envelope, the atomic gas and the molecular gas move together, then, the two tracers should appear approximately at the same $v_{\rm LSR}$. However, this result confirms the prediction from the analysis of the synthetic observations: there is a morphological correlation in the spatial distribution of HI and ^{13}CO. This spatial correlation is not the result of the concentration of emission around particular velocity channels nor the product of chance correlation, as we prove through the statistical tests presented in Appendix B.3. We discuss in detail this correlation around $v_{\rm HI} \approx v_{\rm 13CO}$ in Sec. 5.1 and particularly focus on the $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$ range in Sec. 5.2.

Figure 9 also shows some less-dominant correlation in velocity channels that are not necessary around $v_{\rm HI} \approx v_{\rm 13CO}$, such as that seen around $v_{\rm HI} \approx 10$ and $v_{\rm 13CO} \approx 55\,{\rm km\,s^{-1}}$ or less significantly around $v_{\rm HI} \approx 70$ and $v_{\rm 13CO} \approx 10\,{\rm km\,s^{-1}}$. This correlation appears associated to some vertical stripes in the V-plane, which can be interpreted as the spatial distribution of the ^{13}CO being correlated with the HI in many channels. We discuss this off-diagonal signal, in terms of its position in the V-plane, in Sec. 5.3.

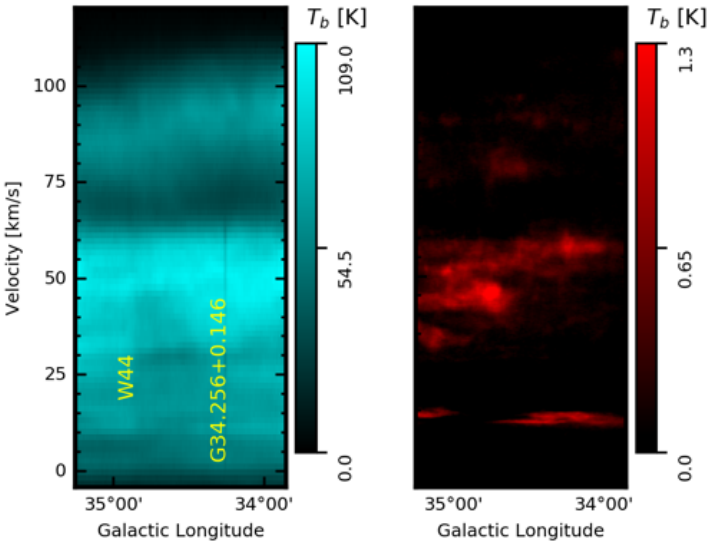

Fig. 8: Longitude-velocity (LV) diagrams of the HI emission (from THOR, left) and ^{13}CO emission (from GRS, right) toward the region presented in Fig. 7. In the HI LV diagram, the shadow around $0 \leq v_{\rm LSR} \leq 45\,{\rm km\,s^{-1}}$ and the vertical line around $0 \leq v_{\rm LSR} \leq 75\,{\rm km\,s^{-1}}$ correspond to absorption toward the supernova remnant (SNR) G34.7-0.4 (W44) and the HII region G34.256+0.146, respectively.

5.1. Interesting velocity ranges

Figure 9 reveals that the largest V values are grouped around roughly four values of $v_{\rm LSR}$; explicitly, $v_{\rm LSR} \approx 12$, 43, 55, and $75\,{\rm km\,s^{-1}}$. These velocities are related to the radial velocities of the individual parcels of HI and ^{13}CO that are morphologically correlated, thus, they are most likely associated with the rotation of the Galaxy and its spiral arm structure. Visual inspection of the spiral arm model presented in Reid et al. (2014) suggests that the ^{13}CO emission at $12\,{\rm km\,s^{-1}}$ might be associated with the Perseus arm, at 43 and $55\,{\rm km\,s^{-1}}$ with the far side of the Sagittarius arm, and at $75\,{\rm km\,s^{-1}}$ with the Aquila spur. However, establishing the association between the central velocities of this emission and the spiral arm structure is not straightforward and it is beyond the scope of this work.

In what follows we detail the HOG analysis around each of these central velocities to establish if the morphological correlation can be associated to a particular set of objects. For that purpose we focus our analysis both in the velocity ranges identified using the values of V and the MC candidates identified in catalogs presented in Rathborne et al. (2009) and Miville-Deschênes et al. (2017). For the sake of simplicity, we also identify the region with maximum V values in a Galactic longitude and latitude grid of 3×7 elements, which we call blocks following the vocabulary introduced in machine vision studies (for example, Zhu et al. 2006). This selection of grid is arbitrary and just aims to guide the eye to the areas of the maps where the ϕ distribution is more significantly peaked around 0°.

In the $-5 \leq v_{\rm LSR} \leq 30\,{\rm km\,s^{-1}}$ velocity range, the most conspicuous feature in V is centered on $v_{\rm LSR} \approx 12\,{\rm km\,s^{-1}}$. Figure 10 reveals that in the pair of HI and ^{13}CO velocity channels with the largest values of V, the gradients in the HI map are dominated by W44, but these do not have a particular correspondence with the ^{13}CO gradients. The maximum values of V correspond to the area in the southeast of W44, around $[l, b] = [35.^{\circ}0, 0.^{\circ}6]$, where an elongated ^{13}CO emission blob has a clear correspondence with the HI. This ^{13}CO emission feature is not among the objects identified in the Rathborne et al. (2009) cloud catalog or included within the effective radius of the objects identified in Miville-Deschênes et al. (2017).

In the $30 \leq v_{\rm LSR} \leq 60\,{\rm km\,s^{-1}}$ velocity range, the most significant features in V are centered at $v_{\rm LSR} \approx 43$ and $55\,{\rm km\,s^{-1}}$. The two velocities roughly correspond to those of the two giant molecular filaments (GMFs) identified in Ragan et al. (2014). The top panel of Fig. 11 shows that in the velocity channel maps corresponding to the largest V values, the HI gradients are still dominated by W44 and the largest correlation appears around the eastern edge of that SNR, around $[l, b] = [34.^{\circ}8, -0.^{\circ}4]$.

The studied area of the sky contains a large number of MC candidates from the Rathborne et al. (2009) and Miville-Deschênes et al. (2017) catalogs in this velocity range. One of the objects in the Rathborne et al. (2009) catalog, centered at $[l, b] = [35.^{\circ}0, -0.^{\circ}5]$, is coincident with the large-V region identified in the top panel of Fig. 11. Additionally, there is also large regions of coincident gradients in the Rathborne et al. (2009) MC candidates centered at $[l, b] = [34.^{\circ}6, 0.^{\circ}25]$ and $[34.^{\circ}6, -0.^{\circ}25]$, although there are extended regions with $\phi \approx 0°$ that do not correspond to any MC candidate.

In the $60 \leq v_{\rm LSR} \leq 90\,{\rm km\,s^{-1}}$ velocity range, the most significant features in V are centered at $v_{\rm LSR} \approx 75\,{\rm km\,s^{-1}}$. The middle panel of Figure 11 shows that in the velocity channel maps corresponding to the largest V values, the correlation between the gradients is concentrated in the region around $[l, b] = [34.^{\circ}5, 0.^{\circ}0]$, which is coincident with two Rathborne et al. (2009) and one Miville-Deschênes et al. (2017) MC candidates.

There is not a significant spatial correlation in the $90 \leq v_{\rm LSR} \leq 120\,{\rm km\,s^{-1}}$ velocity range when it is compared to the V values obtained in the full $-5 \leq v_{\rm LSR} \leq 120\,{\rm km\,s^{-1}}$ range, as illustrated in Fig. 9. However, when considering the pair of velocity channels with the maximum value of V in the $90 \leq v_{\rm LSR} \leq 120\,{\rm km\,s^{-1}}$ range, we find significant spatial correlation toward the Rathborne et al. (2009) and Miville-Deschênes et al. (2017) MC candidates centered on $[l, b] = [34.^{\circ}4, 0.^{\circ}15]$, as shown in the bottom panel of Fig. 11. There, the regions with $\phi \approx 0°$ seem to be less extended than those shown in the $30 \leq v_{\rm LSR} \leq 60\,{\rm km\,s^{-1}}$ and $60 \leq v_{\rm LSR} \leq 90\,{\rm km\,s^{-1}}$ ranges and they cover just a few small patches.

There is some interesting correlation between HI and ^{13}CO around $[l, b] = [34.^{\circ}2, -0.^{\circ}2]$, where there is a clear HISA feature correlated with a small patch of ^{13}CO emission, as it is evident in the gradients and the relative orientation angles presented in the bottom panel of Fig. 11. Nevertheless, this region is not coincident with any of the MC candidates in the Rathborne et al. (2009) and Miville-Deschênes et al. (2017) catalogs.

5.2. HOG in the $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$ range

Due to its large V values and the relatively low number of MC candidates, which facilitates our analysis, we devote special attention to velocity range around $v_{\rm LSR} \approx 55\,{\rm km\,s^{-1}}$. The distribution of the HI and ^{13}CO emission in this velocity range, shown in Fig. 12, suggests at first glimpse the correlation between a large scale HISA ring, seen as the shadows in the HI emission maps, and a ^{13}CO ring where the Rathborne et al. (2009) MC candidates are located. The detailed values of V in this velocity range, presented in Fig. 13, show a departure from the maxima along the $v_{\rm HI} \approx v_{\rm 13CO}$ range, although this behavior is below the $3\varsigma_V$ level in the $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$. We detail the individual behavior toward different portions of this region by making

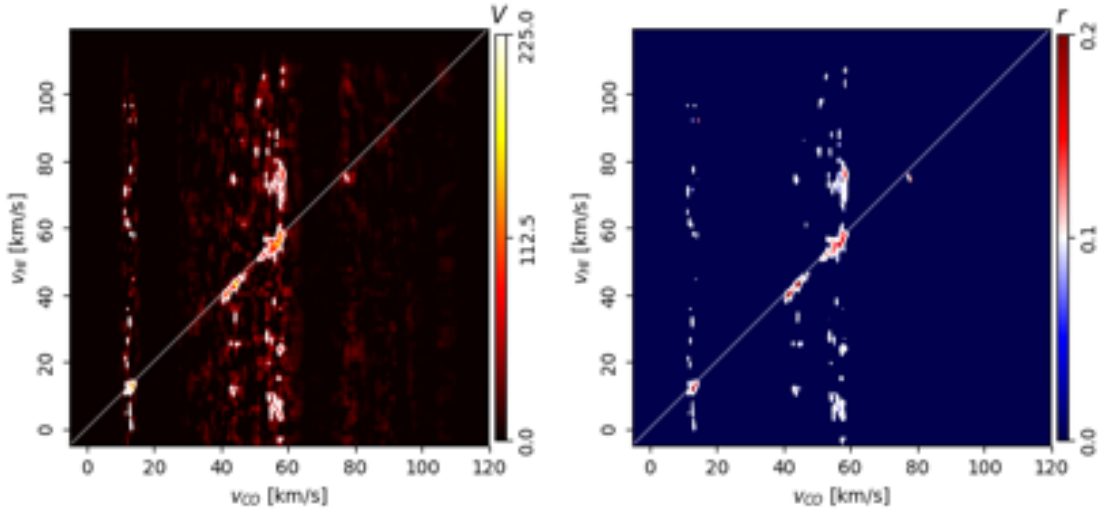

Fig. 9: Results of the HOG analysis of the THOR HI and GRS ^{13}CO observations. *Left.* Projected Rayleigh statistic, $V(v_{13CO}, v_{HI})$, the HOG statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps, defined in Eq. (3). The contours indicate the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ levels in the corresponding velocity range. *Right.* Mean resultant vector length, $r(v_{13CO}, v_{HI})$, with the $3\varsigma_V$ confidence interval, a HOG metric that is roughly equivalent to the percentage of gradient pairs that imply the spatial correlation between the velocity-channel maps, defined in Eq. (2).

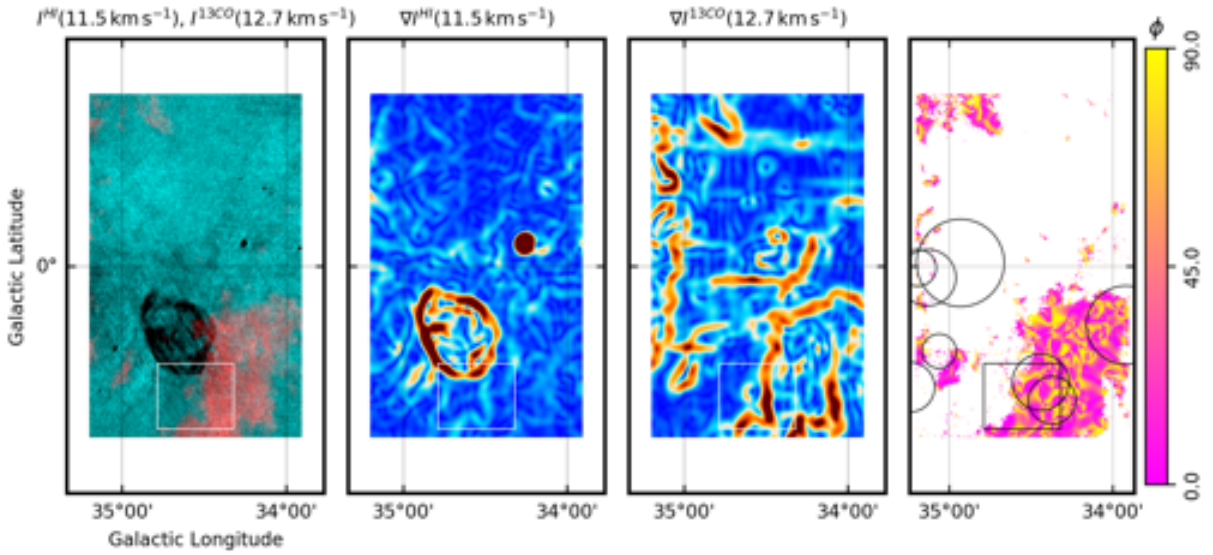

Fig. 10: Intensity, intensity gradients, and relative orientation angle maps from the THOR HI and GRS ^{13}CO observations presented in Fig, 7. *Left.* HI (teal) and ^{13}CO emission (red) in the velocity channels with the largest spatial correlation in the velocity range $-5 \leq v_{LSR} \leq 30\,km\,s^{-1}$, as inferred from the V values shown in Fig. 9. *Middle left.* Norm of the gradient of the HI intensity map in the indicated velocity channel. *Middle right.* Norm of the gradient of the ^{13}CO intensity map in the indicated velocity channel. *Right.* Relative orientation angle ϕ, Eq. (1), between the gradients of the HI and ^{13}CO intensity maps in the indicated velocity channels. The white color in the ϕ map corresponds to areas where the gradient is not significant in either tracer, as estimated using the rejection criteria described in App. A.2. The rectangle shows the block, selected from a 7×3 spatial grid, with the largest values of V. The black circles in the rightmost panel correspond to the positions and effective sizes of the MC candidates from the Rathborne et al. (2009) catalog in the aforementioned velocity range.

use of the objects identified in the Rathborne et al. (2009) MC catalog. Note that Rathborne et al. (2009) employs just one of the multiple methods for producing MC catalogs from emission observations and the MC candidates identified there are not indisputable. Here we use it just as a guide for our analysis of different portions of the studied area.

The rightmost panel of Figure 14 reveals that in the velocity channel maps with the largest V values, the spatial correlation between the HI and the ^{13}CO emission is located in extended patches. We further study these regions by estimating the V values in the block with the highest V values and in the effective area covered by four of the Rathborne et al. (2009) MC candidates; namely, GRS34.19+0.05, GRS34.47−0.67,

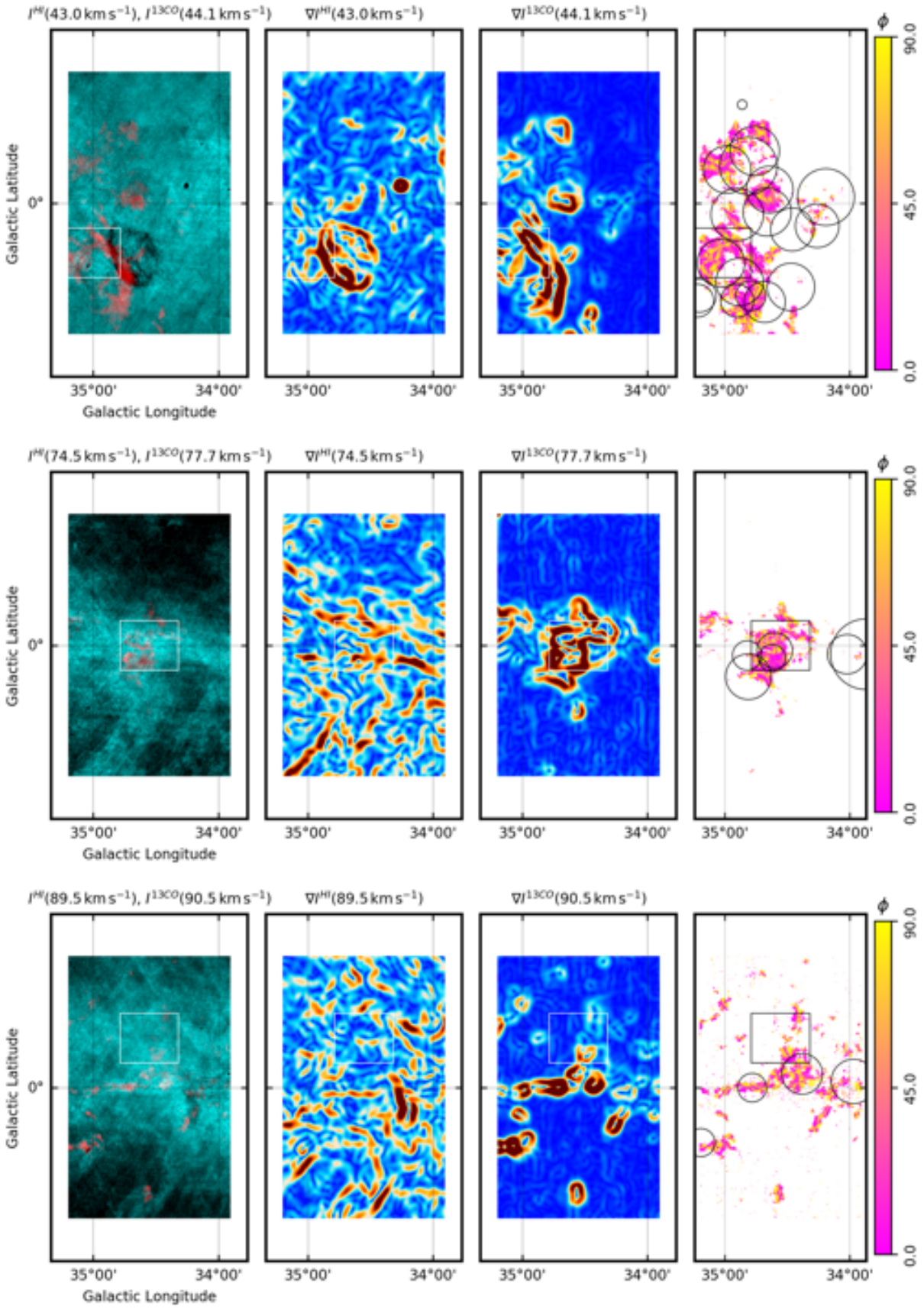

Fig. 11: Same as Fig. 10 for $30 \leq v_{\rm LSR} \leq 60\,{\rm km\,s^{-1}}$ (top), $60 \leq v_{\rm LSR} \leq 90\,{\rm km\,s^{-1}}$ (middle), and $90 \leq v_{\rm LSR} \leq 120\,{\rm km\,s^{-1}}$ (bottom).

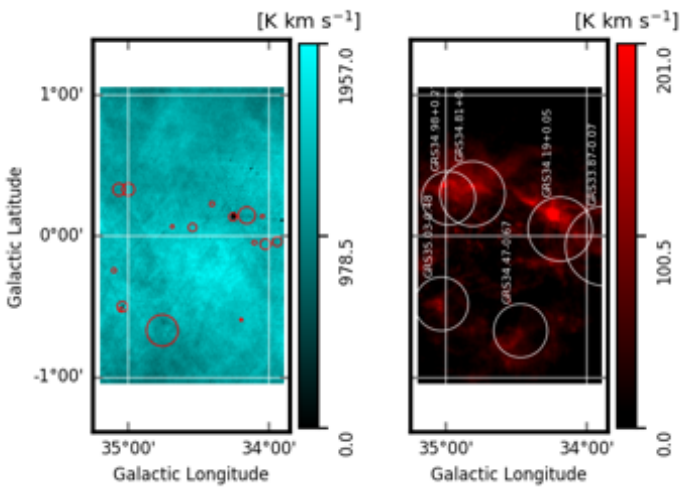

Fig. 12: Integrated HI (left) and ^{13}CO (right) emission in the range $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$. The red circles in the left-hand-side panel correspond to the HII regions from the Anderson et al. (2014) catalog in the aforementioned velocity range. The white circles in the right-hand-side panel correspond to the positions and effective sizes of the MC candidates from the Rathborne et al. (2009) catalog in the aforementioned velocity range.

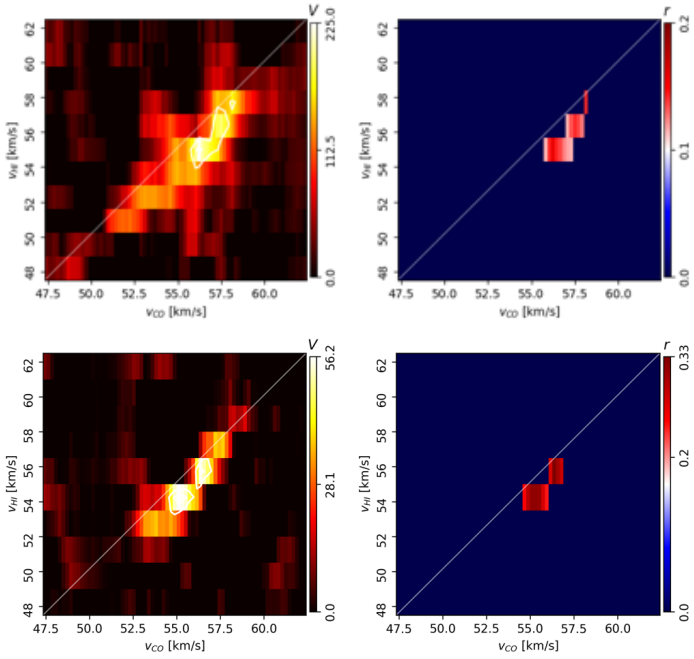

Fig. 13: *Top.* Results of the HOG analysis of the THOR HI and GRS ^{13}CO observations in the velocity range $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$ over the region presented in Fig. 12. *Left.* Projected Rayleigh statistic, $V(v_{\rm 13CO}, v_{\rm HI})$, the HOG statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps, defined in Eq. (3). The contours indicate the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ levels in the corresponding velocity range. *Right.* Mean resultant vector length, $r(v_{\rm 13CO}, v_{\rm HI})$, with the $3\varsigma_V$ confidence interval, a HOG metric that is roughly equivalent to the percentage of gradient pairs that imply the spatial correlation between the velocity-channel maps, defined in Eq. (2). *Bottom.* Same for the maximum-V block shown in Fig. 14

GRS34.81+0.3, GRS34.98+0.27. We exclude from this analysis MC candidates GRS33.87−0.07 and GRS35.03−0.48, also found in the selected velocity range, given the partial coverage of GRS33.87−0.07 and the low V values found toward GRS35.03−0.48.

Figure 13 shows the correlation plane toward the block with the highest V values, indicated by the box in Fig. 14. Toward that portion of the map, the V values around $v_{\rm LSR} \approx 55\,{\rm km\,s^{-1}}$ are maximum for $v_{\rm HI} \approx v_{\rm 13CO}$ in the range $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$. This behavior is similar to that observed in the synthetic observations presented in Sec. 3. However, it does not necessarily imply the presence of colliding clouds toward this region. Note that the block with the largest V value does not contain any identified HII regions.

5.2.1. V values toward MC candidates

The correlation plane corresponding to the MC candidates GRS34.81+0.3, GRS34.98+0.27, and GRS34.19+0.05 show that the concentration of significantly high V values along $v_{\rm HI} \approx v_{\rm 13CO}$ is not a general trend. For example, toward G34.81+0.3 and G34.98+0.27 the V values, presented in Fig. 15, are large around $v_{\rm HI} \approx v_{\rm 13CO}$, but also around $v_{\rm 13CO} \approx 52.5$ and $v_{\rm HI} \approx 57\,{\rm km\,s^{-1}}$. The latter implies morphological correlation in the distribution of the emission in channels maps with a velocity offset of few kilometers per second. This velocity offset does not necessarily imply the flow of one tracer with respect to the other, as discussed in Sec. 3, but it does suggest a dynamic behavior beyond that described by the colliding clouds.

Even more interestingly, the HOG analysis toward GRS34.19+0.05 presented in Fig. 15 shows large V values distributed across a broad range of velocities, thus implying morphological correlations in velocity channels separated by up to a few kilometers per second. This behavior is not entirely unexpected if we consider that GRS34.19+0.05 contains the G34.256+0.136 HII region at $v_{\rm LSR} \approx 54\,{\rm km\,s^{-1}}$ that extends across an area of approximately 3.4′ in diameter (Kuchar & Clark 1997; Kolpak et al. 2003; Anderson et al. 2014). At glance, one could explain it by considering the HI absorption toward the HII region that is present over a range of velocities, but this would only produce a vertical stripe in the distribution of V, that is high V values for a broad range of $v_{\rm HI}$ and a narrow range of $v_{\rm 13CO}$.

It is plausible that the energy injection from the HII region into the surrounding ^{13}CO and HI can produce the high V values in a broad range of $v_{\rm HI}$ and $v_{\rm 13CO}$, by contrast, a region like GRS34.47−0.67 lacks an embedded energy source and shows high V values only around $v_{\rm HI} \approx v_{\rm 13CO}$. Molecular candidates GRS34.81+0.30 and GRS34.90+0.28 are also in the vicinity of HII regions in the right velocity range, in this case G035.0528−00.5180 and G035.1992-01.7424 (Lumsden et al. 2013), yet their distribution of V values across $v_{\rm HI}$ and $v_{\rm 13CO}$ is not as broad as in GRS34.19+0.05. The study of dedicated MHD simulations of the impact of HII regions in a MC (see for example, Geen et al. 2017; Kim et al. 2018) is necessary to unambiguously describe the imprint of this kind of feedback in the HOG correlation.

5.2.2. Is the spatial correlation between HI and CO related to HI self-absorption?

The distribution of HI and ^{13}CO intensities shown in Fig. 12 and Fig. 14 suggests that the high V values mostly correspond to the spatial correlation between ^{13}CO emission and the contours

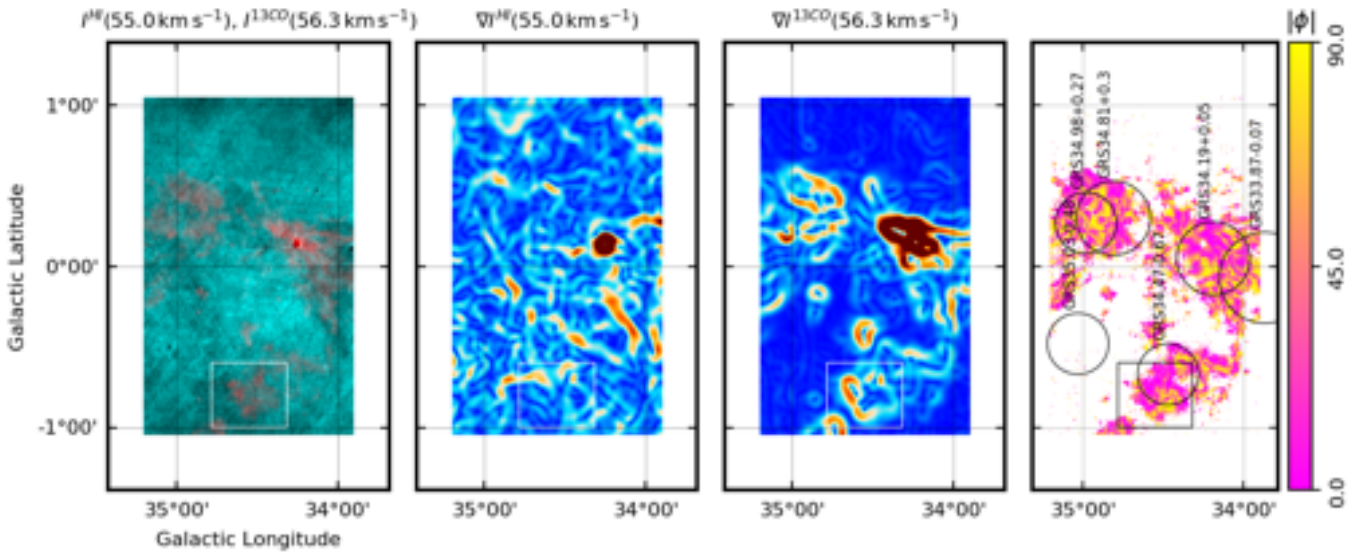

Fig. 14: Same as Fig. 10 for $47.5 \leq v_{\rm LSR} \leq 62.5\,{\rm km\,s^{-1}}$. The labels in the rightmost panel correspond to the MC candidates from the Rathborne et al. (2009) catalog in this velocity range.

of regions with a relative decrease in the HI intensity, which would be produced by HISA. To further explore this possibility, we consider the HI and ^{13}CO spectra toward the MC candidates GRS34.47−0.67, GRS34.19+0.05, GRS34.81+0.3, and GRS34.98+0.27.

These spectra, presented in Fig. 16, suggest that toward the GRS34.19+0.05 and GRS34.98+0.27 MC candidates there are dips in the HI emission around $50\,{\rm km\,s^{-1}}$ that can be associated with the ^{13}CO emission. Closer evaluation of the spectra toward these regions indicates that they correspond to HISA (Bihr 2016; Wang et al. In prep.). However, the same is not true for GRS34.47−0.67 and GRS34.81+0.3, where the peaks in ^{13}CO spectra do not seem associated with a decrease in the HI that can be readily identified as HISA. It is possible that the CNM, which can be spatially correlated with the ^{13}CO towards those two regions, does not have enough contrast with the hotter HI background to produce a clearly identifiable HISA feature in the spectra. But it is also possible that there is a spatial correlation between the ^{13}CO and the thermally unstable HI, which does not produce HISA features, as it is shown in the synthetic observations presented in App. C.1.

5.3. HOG correlation at large separations between $v_{\rm HI}$ and $v_{\rm 13CO}$

Fig. 9 shows that the most significant spatial correlation revealed by the HOG technique appears at $v_{\rm HI} \approx v_{\rm 13CO}$. However, there is a substantial signal in both V and r in velocity channels separated by tens of kilometers per second, for example, around $v_{\rm 13CO} \approx 10\,{\rm km\,s^{-1}}$ and $60 \lesssim v_{\rm HI} \lesssim 100\,{\rm km\,s^{-1}}$ and $v_{\rm 13CO} \approx 50\,{\rm km\,s^{-1}}$ and $0 \lesssim v_{\rm HI} \lesssim 40\,{\rm km\,s^{-1}}$. To explore the origin of these features, we consider the distribution of the gradients and relative orientation angles in HI and ^{13}CO velocity-channel pairs with high V that are separated by a few tens of kilometers per second.

The gradients in the velocity-channel maps corresponding to $v_{\rm HI} = 71.5$ and $v_{\rm 13CO} = 11.4\,{\rm km\,s^{-1}}$, presented in the top panel in Fig. 17, indicate that there is indeed some extended correlation in the spatial distribution of both tracers around $l \approx 34\overset{\circ}{.}5$ and $b \approx -1\overset{\circ}{.}0$. In this particular case, the ^{13}CO distribution seems to be associated with some elongated HI features oriented at roughly 45° with respect to the vertical direction. Similarly, the velocity-channel maps corresponding to $v_{\rm HI} = 5.5$ and $v_{\rm 13CO} = 55.2\,{\rm km\,s^{-1}}$, presented in the bottom panel in Fig. 17, also indicate some extended correlation around $l \approx 35\overset{\circ}{.}0$ and $b \approx 1\overset{\circ}{.}0$. What distinguishes this correlation from that found around $v_{\rm HI} \approx v_{\rm 13CO}$ is that in the former the high V values, $V > 5\varsigma_V$, appear just in a few scattered pairs of velocity channels. In contrast the high V values around $v_{\rm HI} \approx v_{\rm 13CO}$ appear distributed in several pairs of consecutive velocity channels.

The presence of the vertical stripes in the distribution of V indicates that there is some degree of chance correlation wherever there is significant ^{13}CO emission, although in most cases it is below the $5\varsigma_V$ confidence level. This correlation is distributed over a broad range of HI velocity channels due to the fact that there is HI extended structure in all of them, thus increasing the amount of chance correlation with the ^{13}CO emission. This conclusion is confirmed by the presence of similar vertical stripes in the null tests introduced in App. A.3, where the values of V can only be the result of chance correlation.

6. Discussion

The analysis of the HI and ^{13}CO observations using the histogram of oriented gradients (HOG) technique produces three main results that we discuss here.

1. There is a significant spatial correlation between the two tracers in extended portions of the region studied.
2. When considering the spatial correlation revealed by the HOG technique toward particular MC candidates, we find that different clouds present substantial differences in the velocity ranges over which the HOG correlation is distributed.
3. Toward some of the MC candidates the HOG results imply a morphological correlation in the emission of the two tracers in velocity channels separated by up to a few kilometers per second.

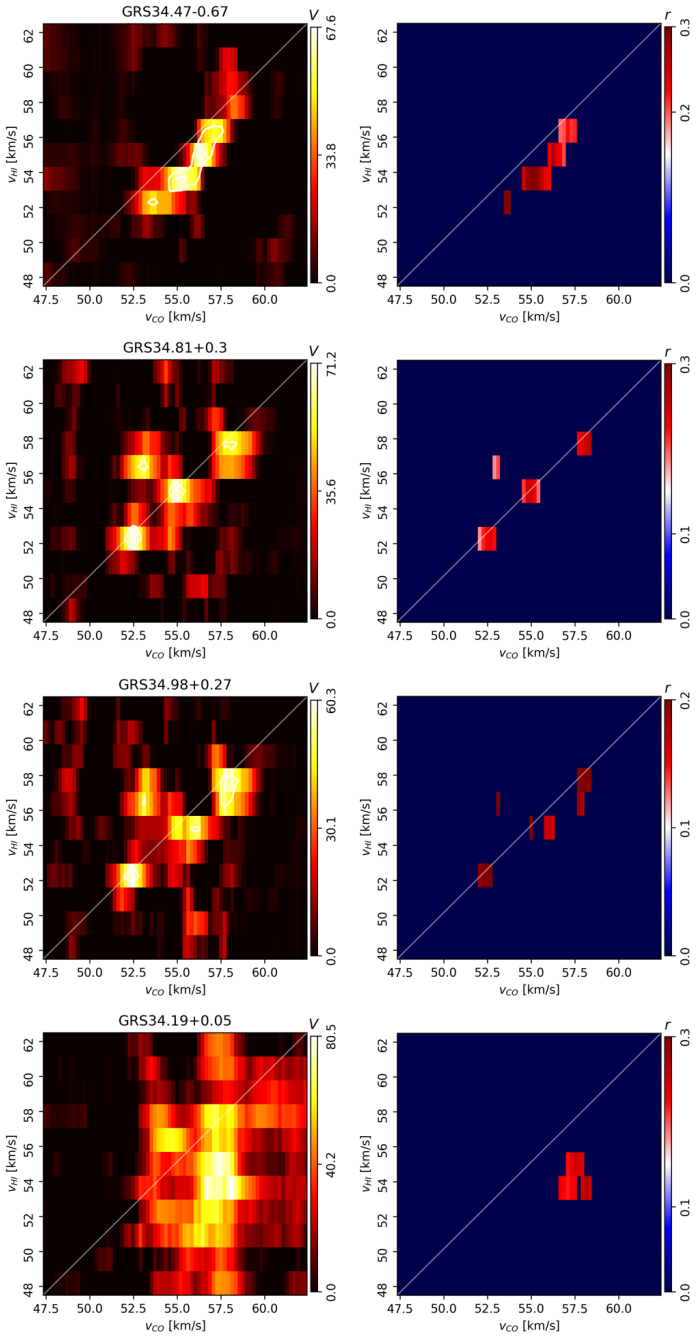

Fig. 15: Same as Fig. 13 for MC candidates GRS34.47−0.67, GRS34.81+0.3, GRS34.98+0.27, and GRS34.19+0.05.

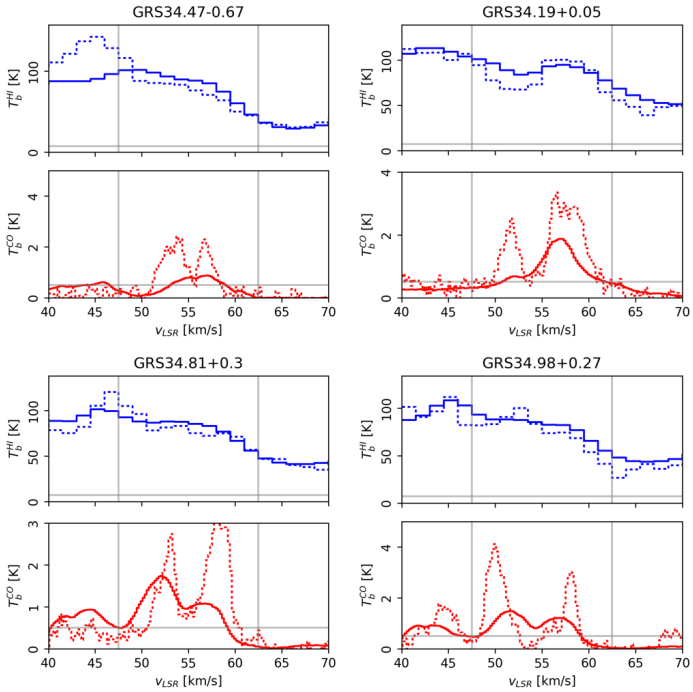

Fig. 16: HI and ^{13}CO spectra toward Rathborne et al. (2009) MC candidates GRS34.47−0.67, GRS34.19+0.05, GRS34.81+0.3, and GRS34.98+0.27. The solid and dotted lines represent the mean spectra over their effective area and toward the central position of the MC candidate, respectively.

6.1. Spatial correlation of HI and ^{13}CO

Using HOG, we find evidences of the spatial correlation of HI and ^{13}CO, or more explicitly, we find that the two tracers have coincident intensity contours traced by the orientation of their gradients. We quantify this spatial correlation using the tools of circular statistics, namely, the projected Rayleigh statistic, V. Previous studies of the association between HI absorption features and molecular gas have been based on the agreement between the velocities, the close agreement of non-thermal line widths, and the matching of the inferred temperatures (e.g., Kavars et al. 2003; Li & Goldsmith 2003; Barriault et al. 2010).

In an overly simplistic model of the ISM, a spherical cloud of diffuse gas and dust in axisymmetric collapse immersed in a bath of isotropic interstellar radiation begins to form a MC when the column density gets sufficiently high that the gas/dust can self-shield, the HI converts to H_2, and the ^{13}CO appears toward the center. In this toy model, the HOG correlation indicates that some of the contours of the HI emission match with the contours of the ^{13}CO emission, even if they do not share a boundary in 3D. Given that we are comparing the gradients, the HOG correlation is not directly related to the correlation or anti-correlation between the amount of atomic and molecular gas, but rather to their spatial distributions. For this toy model cloud it is expected that the gradients of the HI and ^{13}CO emission match, but this is not necessarily the case for a real MC, where the density structure is much more complex and the spectra of both tracers are affected by optical depth and self-absorption, such that even a perfect correlation between atomic and molecular hydrogen would not necessarily result in a good correlation of the HI and ^{13}CO maps. However, the results of the HOG analysis reveal that this spatial correlation is present in the observations.

6.1.1. HI self-absorption and ^{13}CO

The observation of spatial correlation between HI and ^{13}CO has been reported in previous studies of the association of molecular gas and HI self-absorption (HISA) features (Gibson et al. 2005) and narrow HI self-absorption (HINSA) features (Goldsmith & Li 2005; Krčo et al. 2008). However, it was limited by the process of identification and extraction of HISA features, which entails a particular level of complexity. In our blind approach, the HI contours are not particularly associated to the cold gas producing the HISAs, but are rather any contour features that characterize the map. Then, it is convenient to discuss how an object that in principle has no defined edges, such as a cloud of gas

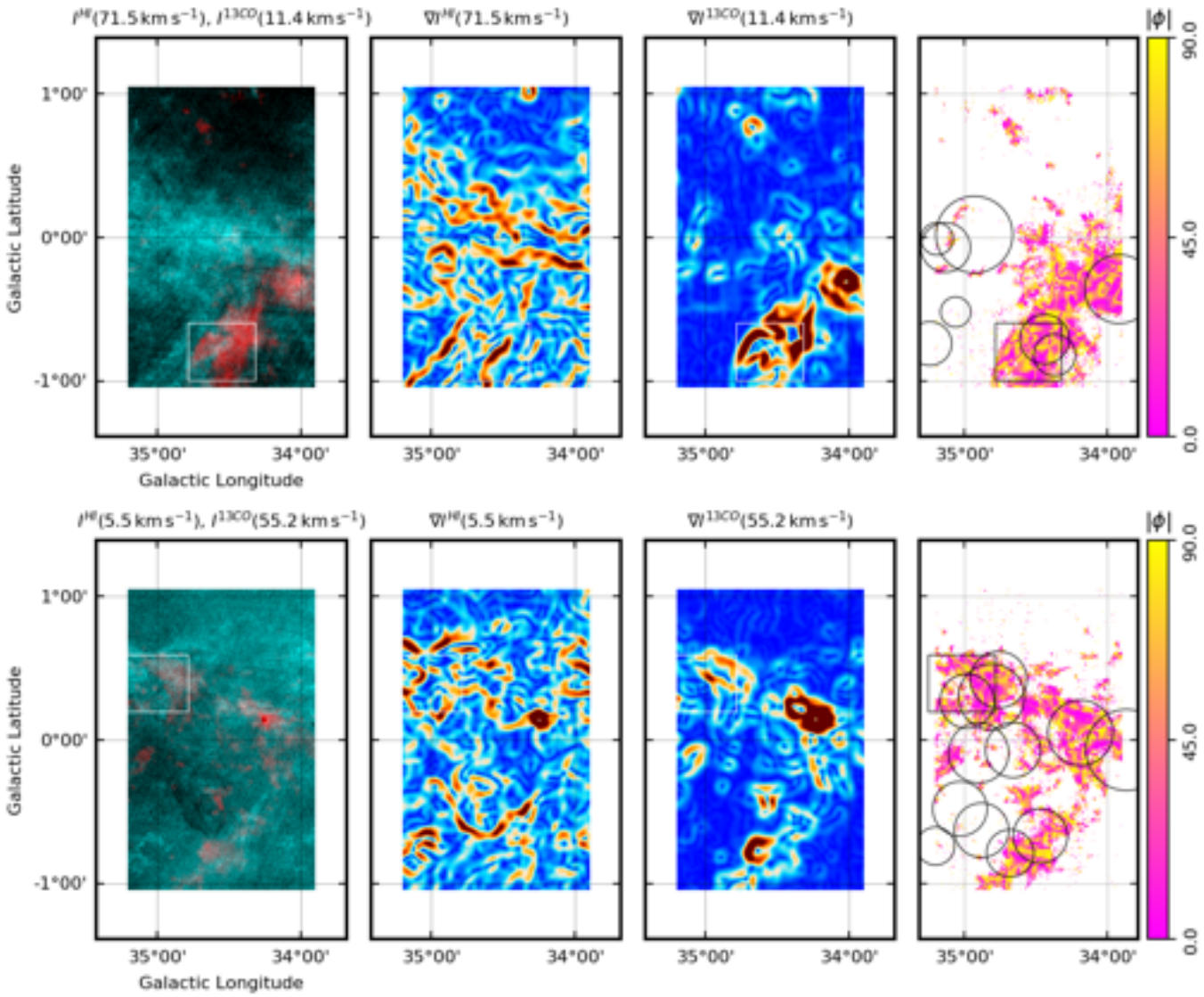

Fig. 17: Same as Fig. 10 for two pairs of HI and ^{13}CO velocity channels maps with high spatial correlation, as inferred from the V values in Fig. 9, but large difference between the velocities $v_{\rm HI}$ and $v_{\rm 13CO}$.

in the ISM, can produce structures that can be identified in two different tracers.

Heiles & Troland (2003) indicate that a model of CNM cores contained in WNM envelopes, as suggested in McKee & Ostriker (1977), provides a good description of the data toward many sources. Additionally, some of these HI envelopes are identified around MCs (e.g., Wannier et al. 1983; Stanimirović et al. 2014). In the turbulent ISM these different phases are not contained within each other like a matryoshka doll; there are no clearly defined boundaries but rather gradients that depend on the distribution of column density structure, radiation field, and spin temperature. Those are the gradients that we consider as potentially responsible for the signal that is found using the HOG technique.

For the particular case of the comparison of HI and ^{13}CO, the conditions of the transition between HI and H_2 and the relation between H_2 and ^{13}CO that are ultimately responsible for the observed emission gradients are very hard to determine for a random MC candidate. The HOG technique does not address the physical and chemical phenomena that produce those gradients, but rather embraces their complexity following a phenomenological and statistical approach to find out where are they coincident and what do they reveal about the MC formation process.

It is unexpected that the HI and ^{13}CO have a tendency to have coincident intensity contours unless these arise from regions of HI self-absorption, as supported by the simulation analysis presented in Sec. 3. However, the observed spatial correlation is not exclusively related to HISA features, as shown in Sec. 5.2.2. This indicates two possibilities: either the spatial correlations are related to self-absorption that is not evident in the central and average spectra presented in Fig. 16, or the spatial correlation is produced by the general HI emission. The first possibility calls for the combination of HOG and the dedicated identification of HISA, which we will address in a subsequent publication (Wang et al. In prep.). The second possibility implies that the interpretation of the HOG results is less simple than what is inferred from the study of the atomic-cloud-collision MHD simulations presented in Sec. 3. For a given velocity channel, the HI signal is contributed from gas parcels both within the cloud/cloud envelope and material not physically associated with the cloud

but with one with broad velocity dispersion that leaks into the cloud velocity interval. Although the study of the MHD simulations without HI emission background, presented in App. C, shows that there is a significant level of spatial correlation between HI and ^{13}CO even without the explicit presence of HI self-absorption, the general interpretation of the spatial correlation between the two tracers will have to be supported by further study of MHD simulations and synthetic observations that reproduce the HISAs better.

6.1.2. Emission background

In contrast to the continuum emission maps in the application of the HOG to the *Planck* data (Soler et al. 2013; Planck Collaboration Int. XXXV 2016), the velocity channels in this analysis include a background component that is not simply the result of the integration of the emission along the line of sight. A particular velocity channel map potentially includes contributions from structures that are not physically connected but produce emission at the same velocity, for example, emission from locations of the Galaxy that have the same $v_{\rm LSR}$ or from portions of unconnected expanding shells, spiral shocks, or non-circular motions near the Galactic bar. The HOG technique evaluates the morphological correlation between the intensity maps of two tracers, independent of the physical conditions producing the observed intensity distribution in a particular velocity channel map. In principle, it is sensitive to the chance correlation introduced by this emission background. However, it is unlikely that this background emission from disconnected regions has a similar structure and would produce singularly high spatial correlation between the considered tracers.

In the Reid et al. (2014) spiral arm model A5 around Galactic longitude $l \approx 34.^{\circ}5$, the velocities $v_{\rm LSR} \approx 12$, 42, and 54 km s^{-1} correspond to kinematic distances of roughly 0.78 ± 0.45, 2.57 ± 0.37, and 3.21 ± 0.36 kpc in the near side of the Galaxy and approximately 12.67 ± 0.46, 10.97 ± 0.37 and 10.36 ± 0.35 kpc in the far side[3]. These large differences between the near and far distances make it unlikely that the morphological correlations between HI and ^{13}CO structures identified in the HOG analysis around those velocities are significantly affected by emission from the other side of the Galaxy. For the same Galactic longitude, the estimated gap between the near and far kinematic distances is lower for larger $v_{\rm LSR}$, for example, it is around 1.45 kpc for $v_{\rm LSR} \approx 100$ km s^{-1} and close to zero close to the tangent point, at roughly $v_{\rm LSR} \approx 120$ km s^{-1}. However, it is difficult to assess if the lack of HOG correlation at $v_{\rm LSR} > 90$ km s^{-1} can be entirely attributed to the blending of density structures into the same velocity range.

If we consider a CO cloud with line-of-sight velocity (LOS) v_0 located directly in front of an expanding HI shell with mean LOS velocity v_1 and expansion velocity v_e, the value of V corresponding to the spatial correlation between the emission of the two tracers at v_0 would not exclusively be that of the CO cloud and its atomic envelope, but would also include the emission from the portion of the shell moving at $v_0 = v_1 + v_e$. If the HI shell is spatially disconnected from the CO cloud, there is no reason why the spatial distribution of the its HI emission at v_0 should be correlated with the CO emission and its contribution to the estimated values of V is that of chance correlation. This chance correlation is well exemplified toward W44, where the expansion of the supernova remnant potentially contributes to the HI emission over a broad range of velocity channels, approximately $10 < v_{\rm LSR} < 45$ km s^{-1} as inferred from Fig. 8, but there is not an exceptionally high spatial correlation with the ^{13}CO emission in that velocity range, as shown in Fig. 9.

[3] BeSSeL survey revised kinematic distance calculator http://bessel.vlbi-astrometry.org

6.2. The HI and ^{13}CO correlation in different environments

When separating the region in individual MC candidates we find three interesting cases in terms of the spatial correlation inferred from V, all illustrated in Fig. 15. First, MCs where the HI and ^{13}CO emission appear correlated at roughly the same velocities. Second, clouds that show correlation around $v_{\rm HI} \approx v_{\rm 13CO}$ and also correlation in some HI and ^{13}CO velocity channels separated by a few kilometers per second. Third, clouds that show correlation between HI and ^{13}CO in many velocity channels distributed on a broad velocity range. Only the first case is arguably consistent with the synthetic observations of the Clark et al. (2018) colliding flows simulation.

The GRS34.19+0.05 MC candidate presents high V values close to $v_{\rm HI} \approx v_{\rm 13CO}$. This trend is very similar to that found in the synthetic observations presented in Sec. 3, however, it does not necessarily imply that this specific configuration corresponds to the physics responsible for the observed values. In principle, this spatial correlation is expected if the atomic and molecular gases are both cospatial and comoving. The spatial correlation, illustrated in Fig. 14, corresponds to the ^{13}CO emission associated with a relative decrease in the HI intensity, which is most likely produced by HISA (Bihr 2016; Wang et al. In prep.), thus suggesting that the observed correlation corresponds to that between the molecular gas sampled by ^{13}CO and the CNM.

The MC candidates GRS34.81+0.3 and GRS34.98+0.27 present high V values close to $v_{\rm HI} \approx v_{\rm 13CO}$, but they also show significant V values around $v_{\rm 13CO} = 52.5$ and $v_{\rm HI} = 56$ km s^{-1}. This significant velocity offset is not reproduced by the synthetic observations presented in Sec. 3, although velocity offsets have been traditionally associated with relative motions between the tracers (e.g., Motte et al. 2014). One possible explanation to this observation is the potential superposition of clouds along the line of sight (Beaumont et al. 2013). However, it is unlikely that spatially separated parcels of HI can have such a high spatial correlation with the same ^{13}CO cloud. Another possibility is that the HII regions introduce a velocity offset between the dense molecular gas and the less dense atomic gas. A final possibility is that more general conditions than those in the Sec. 3 synthetic observations can produce this trend. We discuss the latter two possibilities in more detail in Sec. 6.3.

The deviation from the clustering of high V values around $v_{\rm HI} \approx v_{\rm 13CO}$ is more evident toward GRS34.19+0.05. There, the broad range of velocities with large V can in principle be related to the effects of the HII regions G34.256+0.136 and G34.172+0.175 and the presence of infrared bubbles (Churchwell et al. 2006; Xu et al. 2016). It is worth noting that observationally, the MCs are arbitrarily defined identities, the spatial and velocity associations of ^{13}CO that we call MCs may not correspond to an individual objects with well defined boundaries. So, in general terms, what we are finding with the HOG is that the proximity of HII regions or the relative isolation is related to different behaviors of the spatial correlation sampled by V, and not necessarily that there are two types of MCs in the catalog.

6.3. Potential causes of the HI and ^{13}CO velocity offsets

6.3.1. Cloud evolution

Ionizing radiation from high-mass stars creates HII regions, while stellar winds and supernovae drive the matter in star-forming MCs into thin shells. These shells are accelerated by the combined effect of winds, radiation pressure, and supernova explosions (see Rahner et al. 2017, and references there in). Under the influence of the wind responsible for the shell expansion, the surrounding gas is accelerated, but the less dense atomic gas is accelerated more so that over time, a velocity difference is accumulated between it and the molecular gas (Pound & Goodman 1997; Pellegrini et al. 2007).

We considered this scenario of cloud evolution in the study of the THOR data towards the W49A region, where we found that cloud structure and dynamics of the region are in agreement with a feedback-driven shell that is re-collapsing due to the gravitational attraction Rugel et al. (2018). However, this is the first study where we include the atomic gas that is associated to the star-forming cloud. Potentially, the velocity separation between spatially-correlated HI and ^{13}CO channel maps can be used to study the energy input from HII, but fully exploring that possibility requires additional analysis of models and MHD simulations that are beyond the scope of this work.

6.3.2. Cloud formation

Our analysis of MHD simulations, presented in Sec. 3, suggests that the ideal head-on collision of atomic clouds does not reproduce the velocity offset between the velocity channels with high spatial correlation revealed by the HOG analysis. However, it is expected that more general MC-formation conditions; such as not-head-on collisions, Galactic shear, and different mean magnetic field orientation with respect to the collision axis; could produce different correlations between the atomic and the molecular emission. Indeed, numerical studies of the thermally bistable and turbulent atomic gas show that once formed, the CNM gas is dynamically stable and individual CNM structures have supersonic relative motions that are related to the dynamic of the WNM (Heitsch et al. 2006; Hennebelle & Audit 2007; Saury et al. 2014). For example, the presence of the magnetic field would impose an anisotropy in the flows and if two fronts of atomic gas were not directed parallel to the magnetic field lines, they would have to re-orient themselves and the accumulation of dense gas can appear at a different velocity with respect to the flow of gas that is producing it (Hennebelle & Pérault 2000; Hartmann et al. 2001; Soler & Hennebelle 2017). In a similar way, the Galactic shear, the spiral-arm gravitational potential, or simply the angle between the shock fronts of gas pushed by the ram pressure of supernovae can produce anisotropies that could potentially lead to velocity offsets observed between the atomic and molecular tracers.

In order to test the aforementioned hypothesis, we performed a quick experiment in a segment of one of the stratified, supernova-regulated, 1 kpc-scale, magnetized ISM magneto-hydrodynamical simulations presented in Hennebelle (2018). These simulations trace the evolution of the supernova-regulated multi-phase ISM and, although they do not explicitly estimate the formation of molecular gas, they provide self-consistent initial conditions for the dynamics of the bistable atomic gas. In this simulation, the presence of multiple shock fronts produced by the supernovae explosions makes it extremely unlikely that the accumulation of the dense gas, which can potentially become a MC, is the result of just one collision of atomic flows or the isotropic collapse into one gravitational potential well. In that sense, the accumulation of dense gas in this simulated volume represents a MC formation scenario that is less dependent on the initial conditions of the simulation.

We selected a $(20\,\mathrm{pc})^3$ volume around a density structure identified using a friend-of-friends (FoFs) algorithm with a threshold density $n_C = 10^3\,\mathrm{cm}^{-3}$. Although the FoFs algorithm is not optimal for the general selection of connected structures, in this case we simply use it to identify a reference parcel of gas. We applied the HOG technique to synthetic observations of HI and ^{13}CO emission produced using simple density and temperature thresholds, which is not an optimal approach but it is sufficient for our quick experiment. We refer to App. C.2 for further details on these synthetic observations.

The results of the HOG analysis, shown in Fig. 18, indicate not only the spatial correlation in velocity channels $v_{\mathrm{HI}} \approx v_{13\mathrm{CO}}$, but also a significant correlation in HI and ^{13}CO velocity channels separated by a few kilometers per second. These offsets are persistent for roughly 10^5 years in the simulation and change throughout the evolution of the region. Their presence alone does not clarify the origin of the offsets seen in the analysis of the observations, but suggests that HOG can potentially constitute a good metric for the study of the cloud evolution and formation in numerical simulations. A detailed study of the prevalence of these trends and the physical conditions that produce it in this particular set of MHD simulations is beyond the scope of this work, but constitutes an obvious step to follow in a forthcoming analysis. The main goal of such a study is to identify if the spatial correlation obtained with HOG can be related to the gas motions in MCs that have formed self-consistently within the kilo-parsec numerical simulation and compare its results with other complementary techniques used to characterize the MC kinematics (e.g., Lazarian & Pogosyan 2000; Henshaw et al. 2016; Chira et al. Submitted.).

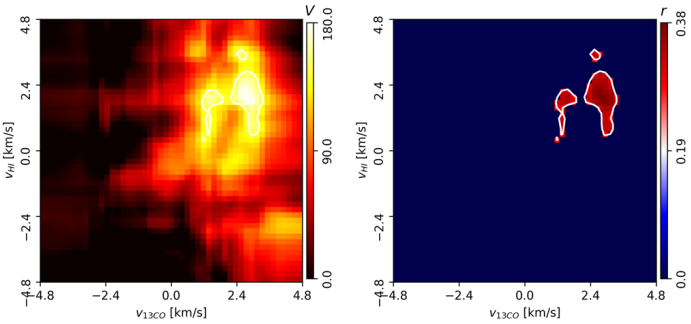

Fig. 18: Results of the HOG analysis of the HI (left) and the ^{13}CO synthetic observations of a segment of the 1-kpc stratified box MHD simulations presented in Hennebelle (2018) and detailed in Appendix C.2. *Left.* Projected Rayleigh statistic, $V(v_{13\mathrm{CO}}, v_{\mathrm{HI}})$, the HOG statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps, defined in Eq. (3). The contours indicate the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ levels in the corresponding velocity range. *Right.* Mean resultant vector length, $r(v_{13\mathrm{CO}}, v_{\mathrm{HI}})$, with the $3\varsigma_V$ confidence interval, a HOG metric that is roughly equivalent to the percentage of gradient pairs that imply the spatial correlation between the velocity-channel maps, defined in Eq. (2).

7. Conclusions and perspectives

We characterize the histogram of oriented gradients (HOG), a tool developed for machine vision that we employ in the study of spectral line observations of atomic and molecular gas. This technique does not assume the organization of the atomic or molecular gas in clouds or complexes. In that sense, it constitutes a "blind" estimator of the coincidence in the spatial distribution of the two tracers.

We applied HOG to a set of synthetic HI and ^{13}CO observations from a MHD simulation of MC formation in the collision of two atomic clouds. There we find significant spatial correlation between the synthetic HI and ^{13}CO emission contours across a broad range of velocity channels. The highest spatial correlation appears around velocity channel pairs with $v_{\rm HI} \approx v_{\rm 13CO}$ independently of the cloud-collision direction with respect to the line of sight.

Using HOG, we studied the spatial correlation of HI and ^{13}CO emission observations toward a portion of the Galactic plane. We significant spatial correlation between the HI and ^{13}CO emission. The highest spatial correlation appears around velocity channel pairs with $v_{\rm HI} \approx v_{\rm 13CO}$, although in some regions there is significant correlation in HI and ^{13}CO velocity channels separated by a few kilometers per second.

We used the catalog of MC candidates derived from the ^{13}CO observations (Rathborne et al. 2009) to analyze the spatial correlation toward particular objects. Part of the spatial correlation identified with the HOG technique appears to be associated with these MC candidates, however, there are extended portions of the maps that are spatially correlated and do not correspond to any of them. The HOG results indicate a different spatial correlation across velocity channels between the two tracers towards MC candidates in the proximity of HII regions. This observation can be interpreted in two ways: either the HII regions are producing this particular dynamical behavior, by their input of energy that potentially affects the atomic and the molecular medium in different ways, or the regions with this dynamical behavior are the ones producing HII regions, by resulting from efficient accumulation of gas. Either scenario is worth exploring in the future using dedicated synthetic observations of MHD simulations.

We showed that the significant correlation in HI and ^{13}CO velocity channels separated by a few km s^{-1} is also found in the synthetic observations of a portion of an MHD simulation with multiple supernovae explosions in a multiphase magnetized medium. But the identification of the physical conditions that produce this velocity offset and its importance for identifying a particular mechanism of MC formation will be the subject of future work based on MHD simulations. In the observational front, we will also continue this work by extending the HOG analysis to the full extent of the THOR observations, using of improved MC catalogs to evaluate HOG toward individual objects, and the combining HOG with the identification of the physical properties of the HI gas.

We conclude that the HOG is a useful tool to evaluate the spatial correlation between tracers of different regimes of the ISM. In this particular case, we used the extended HI and ^{13}CO emission to characterize MCs, but HOG can be used for the systematic comparison of extended observations of other tracers in Galactic and extragalactic targets. The broad range of scales, the diversity of physical conditions, and the large volumes of observed and simulated data make understanding of the dynamical behavior of the ISM a big-data problem. Hiding within those mounds of data are the trends that reveal what determines where and when stars form. HOG constitutes just one of the multiple data-driven tools that in the future should pave the way to a more comprehensive picture of the ISM.

Acknowledgements. JDS, HB, MR, YW, and JCM acknowledge funding from the European Research Council under the Horizon 2020 Framework Program via the ERC Consolidator Grant CSF-648505. SCOG and RK acknowledge support from the Deutsche Forschungsgemeinschaft via SFB 881, "The Milky Way System" (sub-projects B1, B2 and B8), and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) via the ERC Advanced Grant STARLIGHT (project number 339177). FB acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 726384. JK has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 639459 (PROMISE). SER acknowledges support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 706390. NR acknowledges support from the Infosys Foundation through the Infosys Young Investigator grant. RJS acknowledges support from an STFC ERF. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The Galactic Ring Survey is a joint project of Boston University and Five College Radio Astronomy Observatory, funded by the National Science Foundation. This research was carried out in part at the Jet Propulsion Laboratory, operated for NASA by the California Institute of Technology. Part of the crucial discussions that lead to this work took part under the program Milky-Way-Gaia of the PSI2 project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

We thank the anonymous referee for his/her thorough review and highly appreciate the comments, which significantly contributed to improving the quality of this paper. JDS thanks the following people who helped with their encouragement and conversation: Peter G. Martin, Marc-Antoine Miville-Deschênes, Norm Murray, Edith Falgarone, Hans-Walter Rix, Jonathan Henshaw, Shu-ichiro Inutsuka, and Eric Pellegrini.

References

Anderson, L. D., Bania, T. M., Balser, D. S., et al. 2014, ApJS, 212, 1
Anderson, L. D., Wang, Y., Bihr, S., et al. 2017, A&A, 605, A58
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Barriault, L., Joncas, G., Falgarone, E., et al. 2010, MNRAS, 406, 2713
Batschelet, E. 1981, Circular Statistics in Biology, Mathematics in biology (Academic Press)
Beaumont, C. N., Offner, S. S. R., Shetty, R., Glover, S. C. O., & Goodman, A. A. 2013, ApJ, 777, 173
Bergin, E. A. & Tafalla, M. 2007, ARA&A, 45, 339
Beuther, H., Bihr, S., Rugel, M., et al. 2016, A&A, 595, A32
Bihr, S. 2016, PhD thesis, Heidelberg University
Bihr, S., Johnston, K. G., Beuther, H., et al. 2016, A&A, 588, A97
Brunt, C. M., Heyer, M. H., Vázquez-Semadeni, E., & Pichardo, B. 2003, ApJ, 595, 824
Chira, R. A., Ibáñez-Mejía, J. C., MacLow, M. M., & Henning, T. Submitted., A&A
Churchwell, E., Povich, M. S., Allen, D., et al. 2006, ApJ, 649, 759
Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2012, MNRAS, 420, 745
Clark, P. C., Glover, S. C. O., Ragan, S. E., & Duarte-Cabral, A. 2018, ArXiv: 1809.00489
Dawson, J. R., McClure-Griffiths, N. M., Wong, T., et al. 2013, ApJ, 763, 56
Dickey, J. M. & Lockman, F. J. 1990, ARA&A, 28, 215
Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, ApJ, 585, 801
Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, Protostars and Planets VI, 3
Draine, B. T. & Bertoldi, F. 1996, ApJ, 468, 269
Durand, D. & Greenwood, J. A. 1958, The Journal of Geology, 66, 229
Ferrière, K. M. 2001, Reviews of Modern Physics, 73, 1031
Freeman, W. T. & Roth, M. 1994, Orientation Histograms for Hand Gesture Recognition, Tech. Rep. TR94-03, MERL - Mitsubishi Electric Research Laboratories, Cambridge, MA 02139
Geen, S., Soler, J. D., & Hennebelle, P. 2017, MNRAS, 471, 4844
Giacani, E. B., Dubner, G. M., Kassim, N. E., et al. 1997, AJ, 113, 1379
Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 2005, ApJ, 626, 195
Glover, S. C. O. & Clark, P. C. 2012, MNRAS, 421, 116
Glover, S. C. O., Clark, P. C., Micic, M., & Molina, F. 2015, MNRAS, 448, 1607
Glover, S. C. O., Federrath, C., Mac Low, M.-M., & Klessen, R. S. 2010, MNRAS, 404, 2
Glover, S. C. O. & Mac Low, M.-M. 2007, ApJS, 169, 239

Glover, S. C. O. & Mac Low, M.-M. 2011, MNRAS, 412, 337
Goldsmith, P. F. & Li, D. 2005, ApJ, 622, 938
Green, D. A. 2014, Bulletin of the Astronomical Society of India, 42, 47
Hartmann, L., Ballesteros-Paredes, J., & Bergin, E. A. 2001, ApJ, 562, 852
Heiles, C. & Troland, T. H. 2003, ApJ, 586, 1067
Heitsch, F., Slyz, A. D., Devriendt, J. E. G., Hartmann, L. W., & Burkert, A. 2006, ApJ, 648, 1052
Hennebelle, P. 2018, A&A, 611, A24
Hennebelle, P. & Audit, E. 2007, A&A, 465, 431
Hennebelle, P. & Falgarone, E. 2012, A&A Rev., 20, 55
Hennebelle, P. & Pérault, M. 2000, A&A, 359, 1124
Henshaw, J. D., Longmore, S. N., Kruijssen, J. M. D., et al. 2016, MNRAS, 457, 2675
Hu, R., Barnard, M., & Collomosse, J. 2010, in 2010 IEEE International Conference on Image Processing, 1025–1028
Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, ApJS, 163, 145
Jow, D. L., Hill, R., Scott, D., et al. 2018, MNRAS, 474, 1018
Kalberla, P. M. W. & Kerp, J. 2009, ARA&A, 47, 27
Kavars, D. W., Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, ApJ, 598, 1048
Kavars, D. W., Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2005, ApJ, 626, 887
Kim, C.-G., Ostriker, E. C., & Kim, W.-T. 2014, ApJ, 786, 64
Kim, J.-G., Kim, W.-T., & Ostriker, E. C. 2018, ApJ, 859, 68
Klessen, R. S. & Glover, S. C. O. 2016, Star Formation in Galaxy Evolution: Connecting Numerical Models to Reality, Saas-Fee Advanced Course, Volume 43. ISBN 978-3-662-47889-9. Springer-Verlag Berlin Heidelberg, 2016, p. 85, 43, 85
Kolpak, M. A., Jackson, J. M., Bania, T. M., Clemens, D. P., & Dickey, J. M. 2003, ApJ, 582, 756
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216
Krčo, M., Goldsmith, P. F., Brown, R. L., & Li, D. 2008, ApJ, 689, 276
Kuchar, T. A. & Clark, F. O. 1997, ApJ, 488, 224
Kulkarni, S. R. & Heiles, C. 1987, in Astrophysics and Space Science Library, Vol. 134, Interstellar Processes, ed. D. J. Hollenbach & H. A. Thronson, Jr., 87–122
Lazarian, A. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 215, Cosmic Evolution and Galaxy Formation: Structure, Interactions, and Feedback, ed. J. Franco, L. Terlevich, O. López-Cruz, & I. Aretxaga, 69
Lazarian, A. & Hoang, T. 2007, MNRAS, 378, 910
Lazarian, A. & Pogosyan, D. 2000, ApJ, 537, 720
Lazarian, A. & Yuen, K. H. 2018, ApJ, 853, 96
Leonardis, A., Bischof, H., & Pinz, A., eds. 2006, Lecture Notes in Computer Science, Vol. 3951, Histograms of Oriented Gradients for Human Detection, ed. A. Leonardis, H. Bischof, & A. Pinz (Springer)
Levrier, F., Le Petit, F., Hennebelle, P., et al. 2012, A&A, 544, A22
Li, D. & Goldsmith, P. F. 2003, ApJ, 585, 823
Liszt, H. 2001, A&A, 371, 698
Lumsden, S. L., Hoare, M. G., Urquhart, J. S., et al. 2013, ApJS, 208, 11
Mardia, K. 1972, Statistics of Directional Data, Probability and Mathematical Statistics a Series of Monographs and Textbooks (Academic Press)
McConnell, R. 1986, Method of and apparatus for pattern recognition, uS Patent 4,567,610
McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148
Miville-Deschênes, M.-A. & Martin, P. G. 2007, A&A, 469, 189
Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, ApJ, 834, 57
Molinari, S., Bally, J., Glover, S., et al. 2014, Protostars and Planets VI, 125
Motte, F., Nguyên Luong, Q., Schneider, N., et al. 2014, A&A, 571, A32
Nelson, R. P. & Langer, W. D. 1999, ApJ, 524, 923
Pellegrini, E. W., Baldwin, J. A., Brogan, C. L., et al. 2007, ApJ, 658, 1119
Planck Collaboration Int. XXXV. 2016, A&A, 586, A138
Pound, M. W. & Goodman, A. A. 1997, ApJ, 482, 334
Ragan, S. E., Henning, T., Tackenberg, J., et al. 2014, A&A, 568, A73
Rahner, D., Pellegrini, E. W., Glover, S. C. O., & Klessen, R. S. 2017, MNRAS, 470, 4453
Rathborne, J. M., Johnson, A. M., Jackson, J. M., Shah, R. Y., & Simon, R. 2009, ApJS, 182, 131
Rayleigh, L. 1879, Proceedings of the London Mathematical Society, s1-11, 57
Reach, W. T., Koo, B.-C., & Heiles, C. 1994, ApJ, 429, 672
Reach, W. T., Rho, J., Tappe, A., et al. 2006, AJ, 131, 1479
Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130
Roman-Duval, J., Heyer, M., Brunt, C. M., et al. 2016, ApJ, 818, 144
Rugel, M. R., Rahner, D., Beuther, H., et al. 2018, arXiv e-prints
Saury, E., Miville-Deschênes, M.-A., Hennebelle, P., Audit, E., & Schmidt, W. 2014, A&A, 567, A16
Sembach, K. R., Howk, J. C., Ryans, R. S. I., & Keenan, F. P. 2000, ApJ, 528, 310
Shetty, R., Glover, S. C., Dullemond, C. P., & Klessen, R. S. 2011, MNRAS, 412, 1686
Soler, J. D., Ade, P. A. R., Angilè, F. E., et al. 2017, A&A, 603, A64
Soler, J. D., Beuther, H., Rugel, M., Wang, Y., & THOR collaboration. In prep., A&A
Soler, J. D. & Hennebelle, P. 2017, A&A, 607, A2
Soler, J. D., Hennebelle, P., Martin, P. G., et al. 2013, ApJ, 774, 128
Spitzer, L. 1978, Physical processes in the interstellar medium
Springel, V. 2010, MNRAS, 401, 791
Stanimirović, S., Murray, C. E., Lee, M.-Y., Heiles, C., & Miller, J. 2014, ApJ, 793, 132
Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10
Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, AJ, 132, 1158
Szűcs, L., Glover, S. C. O., & Klessen, R. S. 2014, MNRAS, 445, 4055
Wang, Y., Beuther, H., Rugel, M., Soler, J. D., & THOR collaboration. In prep., A&A
Wannier, P. G., Lichten, S. M., & Morris, M. 1983, ApJ, 268, 727
Westerhout, G. 1958, Bull. Astron. Inst. Netherlands, 14, 215
Wilson, T. L. & Rood, R. 1994, ARA&A, 32, 191
Xu, J.-L., Li, D., Zhang, C.-P., et al. 2016, ApJ, 819, 117
Zhu, Q., Yeh, M.-C., Cheng, K.-T., & Avidan, S. 2006, in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol. 2, 1491–1498

Appendix A: Histograms of oriented gradients

A.1. Computation of the gradient

In the HOG, the computation of the gradient is performed by convolving the individual velocity-channel maps with the derivative of a two-dimensional Gaussian, what is known as a Gaussian derivative (Soler et al. 2013). The size of the Gaussian determines the area of the vicinity over which the gradient will be calculated. Varying the size of the Gaussian kernel enables the sampling of different scales and reduces the effect of noise in the pixels.

In algebraic terms, we estimate the gradient of the k-th velocity channel of the PPV cube $I_{i,j,k}$, where the indexes i and j run over the spatial coordinates x and y, by computing

$$\nabla I_{i,j,k} = I_{i,j,k} \otimes \nabla \left(A \exp \left[-\frac{x^2}{2\sigma_g^2} - \frac{y^2}{2\sigma_g^2} \right] \right), \qquad \text{(A.1)}$$

where $\otimes$ is the convolution operator, ∇ represents the standard gradient calculated using forward differences of adjacent pixels, and A is a normalization factor, such that the integral of the Gaussian function equals unity. For the sake of simplicity, we choose the same variance σ_g of the Gaussian function in the x- and y-direction. In practice, we apply the Gaussian filter routine in the multi-dimensional image processing `ndimage` package of `scipy` with reflecting boundary conditions.

Figure A.1 illustrates the effect of the derivative kernel size in the HOG analysis. The gradient diameter sets the spatial scale of the intensity contours that we compare using the gradient and consequently, it corresponds to the size of the correlated patches in the maps of ϕ, shown in the rightmost panel of Fig. A.1. However, we note that the regions of the map with $\phi \approx 0°$ are persistent across derivative kernels with FWHM = 46″ (GRS data resolution), 60″, 90″ (shown in Fig. 14), and 105″. Additionally, the shape of the HOGs is not significantly changed by the selection of the kernel size in the aforementioned range, as shown in Fig. A.2. This is not entirely unexpected given that the structures in the velocity-channel maps show spatial correlations across multiple scales (see for example Lazarian 2000; Brunt et al. 2003).

Despite the fact that the HOG results are persistent for the selected derivative kernel sizes, this may not always be the case.

On the one hand, the selection of a very large derivative gradient would wash out the signal and will not profit from the angular resolution of the observations. On the other hand, using a very small derivative gradient will make more evident the features produce by noise and non-ideal telescope beams. Given that to first order the selection of the kernel size between 46″ and 105″ does not significantly change the correlation that we report in this paper, as we further discuss in Sec. A.4, we have chosen to report the results of the analysis using the 90″ kernel, which is roughly twice the angular resolution of the THOR and GRS data (40 and 46″, respectively) and reduces most of the interferometer features in the THOR HI observations. In the future we will explore in detail further improvements that could be obtained with the selection of the derivative kernel sizes.

A.2. Gradient selection

The observed signal in the k-th velocity channel can be interpreted as

$$I_{i,j,k} = I^0_{i,j,k} + \delta_{i,j,k}, \tag{A.2}$$

where $I^0_{i,j,k}$ is the emission and $\delta_{i,j,k}$ the noise in the map. The gradient of the observed velocity-channel maps can be written as

$$\nabla I_{i,j,k} = \nabla I^0_{i,j,k} + \nabla(\delta_{i,j,k}). \tag{A.3}$$

We quantify the contribution of the noise to the gradient by evaluating $\nabla I_{i,j,k}$ in velocity-channel maps without signal. For that purpose, we identify the velocity-channel map, $I_{i,j,k'}$, with the lowest average emission and compute the reference noise intensity

$$I^{\rm N} \equiv \sqrt{\left\langle I^2_{i,j,k'} \right\rangle_{ij}}, \tag{A.4}$$

where $\langle \cdots \rangle_{ij}$ denotes the average over the spatial coordinates. Then, we compute the average intensity gradient in that velocity-channel map,

$$|\nabla I^{\rm N}| \equiv \left\langle |\nabla(I_{i,j,k'})| \right\rangle_{ij}, \tag{A.5}$$

where $|\cdots|$ denotes the norm of the vector.

Assuming that $I^{\rm N}$ is a good approximation for $\delta_{i,j,k}$ and $|\nabla I^{\rm N}|$ is representative of $\nabla(\delta_{i,j,k})$, we compute the HOG using only the gradients in regions of each velocity map where $I_{i,j,k} \geq 5I^{\rm N}$ and $|\nabla I_{i,j,k}| \geq 5|\nabla I^{\rm N}|$. The first criterion guarantees that the gradients are not coming from a region of the map with low signal-to-noise ratio (S/N). The second criterion guarantees that the gradients are larger than those produced by the noise in the velocity-channel maps. Additionally, we exclude the gradients that are within a distance $\sigma_{\rm g}$, as defined in Eq. (A.1), from the map edge.

A potential source of noise in the gradients is the presence of features from the image reconstruction of interferometric data, as is the case in the THOR HI observations. The noise in the THOR maps is highly non-uniform and non-Gaussian, so using the simple spatial average of the gradient of the noise to judge the significance of the intensity gradients may be insufficient. However, given that we are comparing the interferometric observations with single-dish observations, these spatial features would only be present in the HI velocity-channel maps and they would only contribute to the chance correlation between them. Additionally, the smoothing implied in the Gaussian derivative operation mitigates the effect of non-uniform noise. Further studies of the effect of noise on the intensity gradients in the specific context of the GRS ^{13}CO and THOR HI analysis are presented in Sec. B.1 and Sec. B.2.

A.3. Statistical evaluation of the HOG results.

In the histograms of relative orientation (HOG) method we have a set of orientation angles ϕ_k in the range $[-\pi/2, \pi/2]$, estimated from Eq. (1). To test for uniformity of these data we map each angle into twice itself, i.e. $\theta_k \rightarrow \theta = 2\phi_k$. This method of angle doubling is a common technique for converting axial data, which carries information about the orientation and not the direction, to circular data, in order to utilize the tools of circular statistics (see Jow et al. 2018, and references therein).

We present the results of the HOG analysis using the resultant vector length, r, which we define in Eq. (A.8). This is a normalized quantity that, to zeroth order, can be interpreted as the fraction of parallel gradient vectors and encapsulates the information in the HOGs. However, the value of r is purely descriptive. To quantify the statistical significance of r values we apply the projected Rayleigh test: a test that the angle distribution is peaked at 0°, that is, that it represents mostly parallel gradient vectors.

A.3.1. The mean resultant vector length

The first step in the analysis of relative orientations is the definition of the resultant vector length (Batschelet 1981). Given a set θ_k of N angles, one can associate them with a set of unitary vectors. In a rectangular x, y-coordinate system, the components of the unit vectors are

$$x_k = \cos\theta_k \text{ and } y_k = \sin\theta_k. \tag{A.6}$$

The sum of vectors, also called the resultant vector, has components

$$X = \sum_k^N w_k \cos\theta_k \text{ and } Y = \sum_k^N w_k \sin\theta_k, \tag{A.7}$$

where w_k is the statistical weight associated to the angle ϕ_k.

The length r of the mean vector is simply

$$r = \frac{1}{\sum_k^N w_k}(X^2 + Y^2)^{1/2}. \tag{A.8}$$

If the resultant vector length is close to zero, then no single preferred direction exists. This may be the case where all directions are equally likely, that is, a uniform distribution of angles. This may also be the case with certain multi-modal distributions, for example, when two opposite directions are equally probable. The statistical significance of a preferential relative orientation is evaluated by testing whether r differs from zero significantly.

The mean resultant vector length r is the normalized quantity that we use to systematically characterize the histograms of oriented gradients (HOGs), which we show for reference in Fig. A.2. If all the angles θ_k are identical, $r = 1$. If the angles θ_k are uniformly distributed, r is close to zero. If the angles θ_k are not uniformly distributed, r is larger than zero and roughly corresponds to the percentage of angles that represent a preferential orientation. Figure A.3 shows the values of r obtained from the relative orientation angles estimated using Eq. (1) in the HI and ^{13}CO emission maps at the indicated velocities.

A.3.2. The projected Rayleigh statistic

The Rayleigh test is used to determine whether or not a set of angles are uniformly distributed (Rayleigh 1879). The Rayleigh

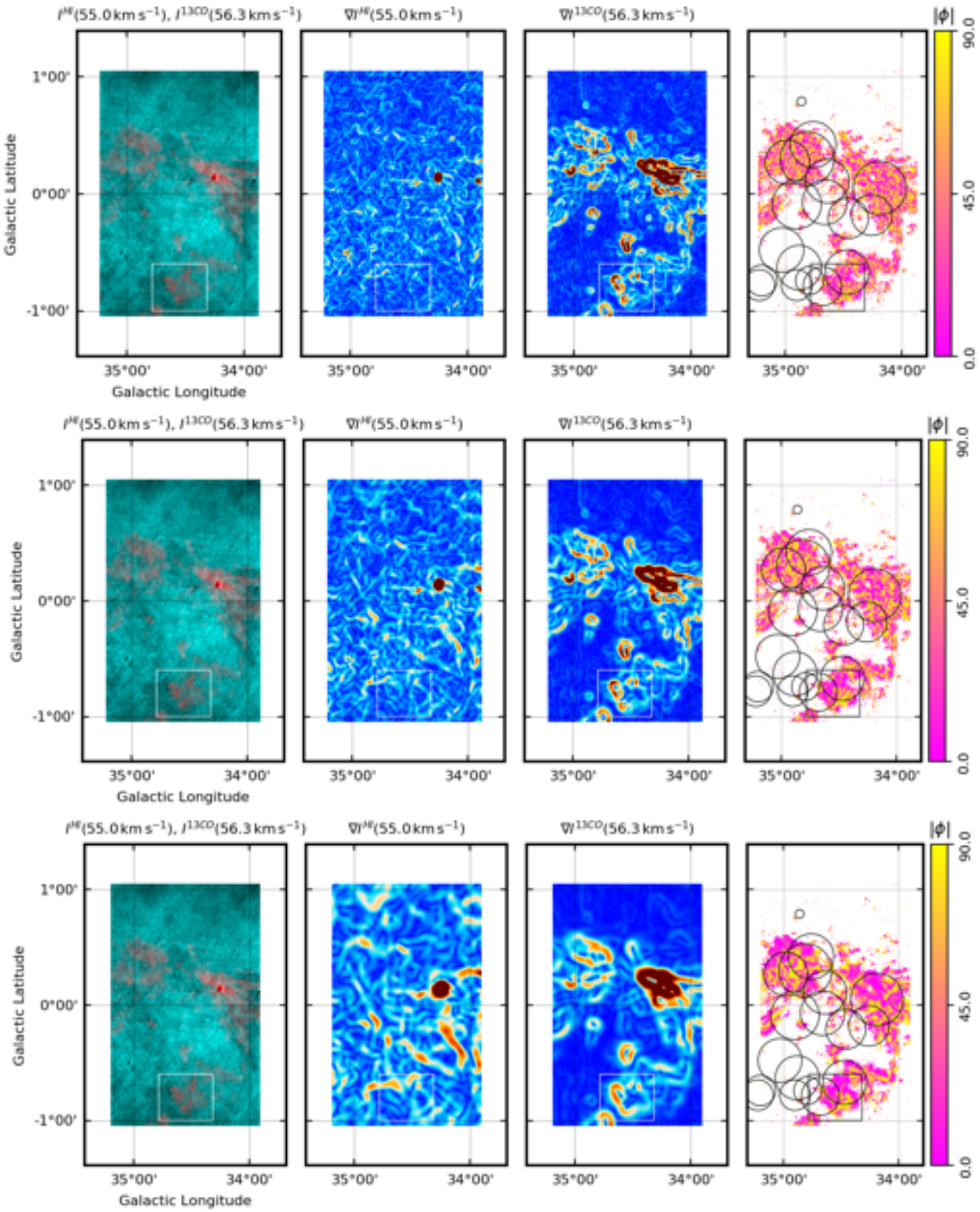

Fig. A.1: Same as Fig. 14 for derivative kernels with 46″ (top), 60″ (middle), and 105″ FWHM (bottom).

test assumes that the sample is generated from a von Mises distribution, that is, a continuous probability distribution on the circle, a close approximation to the wrapped normal distribution, which is the circular analogue of the normal distribution. That means

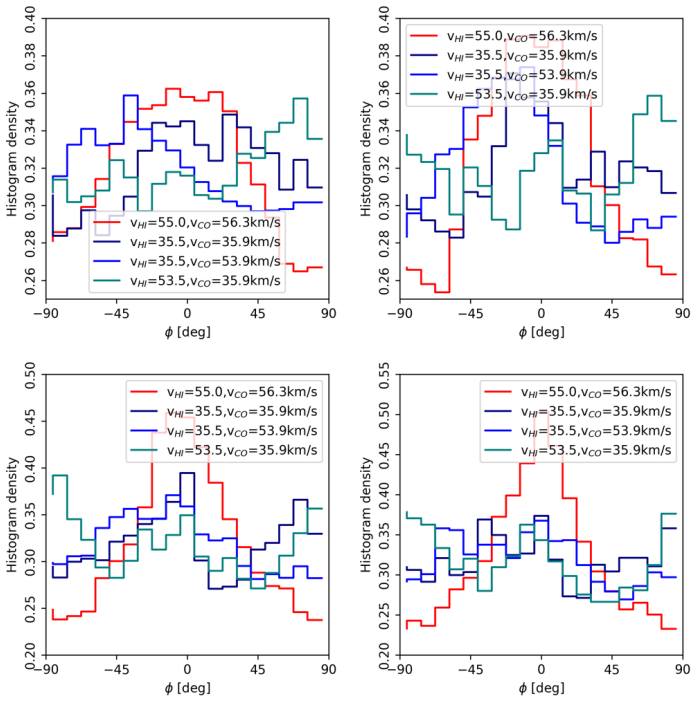

Fig. A.2: Histograms of oriented gradients (HOGs) corresponding to analysis presented in Fig. A.1. They correspond to derivative kernels with 46″ (top left), 60″ (top right), 90″ (bottom left) and 105″ (bottom right) FWHM.

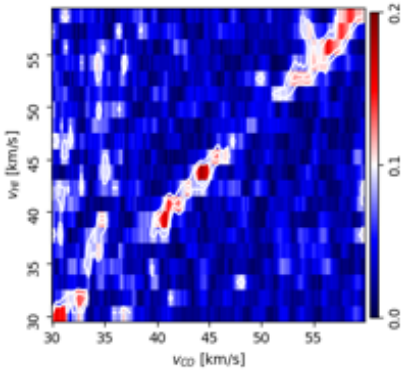

Fig. A.3: Mean resultant vector length, $r(v_{13\mathrm{CO}}, v_{\mathrm{HI}})$, as defined in Eq. (A.8), for the relative orientation angles ϕ, Eq. (1), calculated from the HI and ^{13}CO velocity-channel maps in the range $30 < v_{\mathrm{LSR}} < 60\,\mathrm{km\,s^{-1}}$. The contours are at the $3\varsigma_V$ and $4\sigma_r$ levels, where σ_r represents the standard deviation of $r(v_{13\mathrm{CO}}, v_{\mathrm{HI}})$ in the indicated velocity range.

that it should be used only when the distribution is unimodal. In the Rayleigh test, under the null hypothesis H_0 the population is uniformly distributed around the circle. Under the alternative hypothesis H_A the population is not distributed uniformly around the circle.

In the Rayleigh test, if the magnitude of the mean resultant length in the most common instance of unspecified mean direction, defined as

$$Z \equiv (X^2 + Y^2)^{1/2} = \left[r^2 \sum_k^N w_k \right]^{1/2}, \qquad \text{(A.9)}$$

is large, the null hypothesis H_0 is rejected. This can be interpreted as the net displacement of a random walk with N steps, each with a corresponding length w_k. If the angles θ_k are uniformly distributed, the net displacement is close to zero. If the angles θ_k are not uniformly distributed, the net displacement is larger than zero.

Figure A.4 shows the results of the Rayleigh test applied to the same observations presented in Fig. A.3. The values of Z clearly show that many of the points where the values of r were significant do not pass the non-uniformity test. Most of the high-Z values are located along the diagonal of the plot, that is, in emission maps corresponding to velocity channels where $v_{\mathrm{HI}} \approx v_{13\mathrm{CO}}$. To account for the correlation by random chance, we consider only the values which are above three times the standard deviation of Z within the selected velocity range.

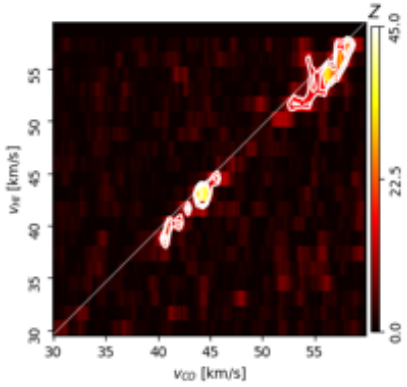

Fig. A.4: Rayleigh statistic Z, as defined in Eq. (A.9), for the relative orientation angles ϕ calculated between HI and ^{13}CO velocity-channel maps in the range $30 < v_{\mathrm{LSR}} < 60\,\mathrm{km\,s^{-1}}$. The contours represent the $3\sigma_Z$ and $4\sigma_Z$ contours, where σ_Z represents the standard deviation of Z within the indicated velocity range.

In the HOG application, we are interested in testing whether the relative orientation is preferentially parallel (corresponding to $\phi_i = 0$) and what is the statistical strength of that trend. To perform that test we use the projected Rayleigh statistic (PRS, Jow et al. 2018), also known in its general form as the V statistic (Durand & Greenwood 1958; Mardia 1972).

The null hypothesis H_0 that we test is randomness, which means that the angles of the sample are independent observations from a uniform circular distribution. The V test for circular uniformity is similar to the Rayleigh test with the difference that under the alternative hypothesis H_A is assumed to have a known mean direction, which in our case corresponds to $\theta_A = 0°$. In that particular case, the PRS is

$$V = \frac{\sum_k^N w_k \cos 2\phi_k}{\sqrt{\sum_k^N (w_k)^2/2}}, \qquad \text{(A.10)}$$

where w_k is the statistical weight assigned to the angle ϕ_k.

In our application, $V > 0$ indicates mostly parallel relative orientation between the gradient vectors. The values $V < 0$ correspond to the mostly perpendicular relative orientation between the gradient vectors, which does not carry any particular significance when comparing two images, but it is important when this test is used in the study of the relative orientation between column density structures and the magnetic field (Soler et al. 2017; Jow et al. 2018).

The top panel of Fig. A.5 shows the results of the project Rayleigh test applied to the same observations considered in Fig. A.4. As in the case of Z, most of the high-V values are located along the diagonal of the plot, that is, in emission maps corresponding to velocity channels with $v_{\rm HI} \approx v_{\rm ^{13}CO}$. The asymptotic limit of the V distribution, in the large N limit, is the standard normal distribution (Jow et al. 2018). Thus, for a general distribution of angles, the variance in V is simply the variance of each $\cos(2\phi)/\sqrt{1/2}$ and can be estimated as

$$\sigma_V = \frac{2\sum_k^N (\cos 2\phi_k)^2 - V^2}{N}. \quad \text{(A.11)}$$

Note that this value corresponds to the variance in the distribution of V for a particular pair of velocity channels maps, shown in the top-middle panel of Fig. A.5. In general, the values derived from Eq. A.11 are smaller than those that arise from the random correlation between velocity channels, ς_V, which is estimated using Eq. 5, hence we use the latter to report the statistical significance of the V values. For the sake of completeness, we also report in Fig. A.5 the values of $\sigma_V^{\rm MC}$ derived from the Monte Carlo sampling introduced in Sec. B.1. The large difference between the σ_V values derived from Eq. (A.11) and $\sigma_V^{\rm MC}$ confirms the inadequacy of the assumption of statistically independent angles in the analysis of the HOG results. Once pre-selected using the values of ς_V, the values of r roughly corresponds to the fraction of parallel gradient vectors. This selection leads to the values presented in the bottom panel of Fig. A.5.

We also calculated the alignment measurement (AM), an alternative method for estimating the degree of alignment between vectors. The alignment measurement is widely used in the study of dust grain alignment with magnetic fields (see for example, Lazarian & Hoang 2007) and also in the family of papers represented by Lazarian & Yuen (2018). The calculation of this quantity is made using

$$\mathrm{AM} = \left\langle 2\cos\phi_{ij} - 1 \right\rangle_{ij}, \quad \text{(A.12)}$$

where $\langle\cdots\rangle_{ij}$ represents the average over the relative orientation angles. The values of AM, shown in Fig. A.6, present a very similar trend to that found with V, although the AM has a shorter dynamic range than V, that is, the contrast between regions with low and high correlation is much lower.

A.4. Derivative kernel size

We discussed the effect of the derivative kernel size in the relative orientation angle, ϕ, maps and the HOGs in Sec. A.1. Here we consider the effect of the derivative kernel size in the distribution of V and r. Figure A.7 show the values of both quantities for derivative kernels with 46″, 60″, 75″, and 105″ FWHM. It is clear from the distribution of these two quantities that the kernel size does not significantly affect the positions of the high-V velocity ranges or their corresponding range of r values. However, the level of significance of the off-diagonal high-V regions around $v_{\rm HI} \approx v_{\rm ^{13}CO} \approx 55\,\mathrm{km\,s^{-1}}$ changes depending on the size of the derivative kernel.

Appendix B: Statistical significance of the HOG method

B.1. Impact of noise in the HOG results

Throughout this paper, we have reported the results of the HOG analysis based on the selection of gradients in pixels with intensities a number of times above the noise level. Here we detail the

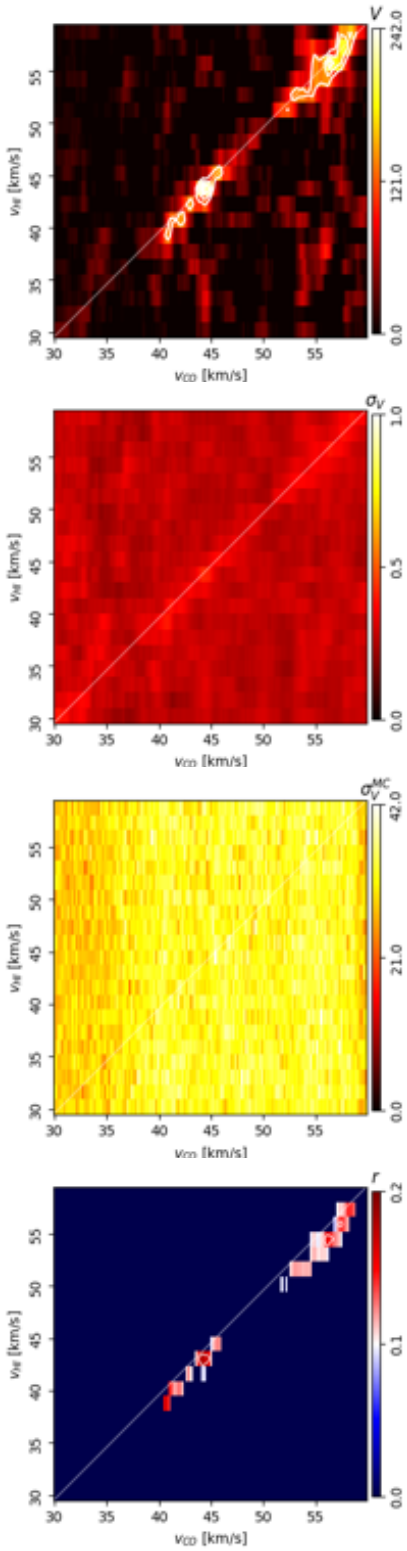

Fig. A.5: From top to bottom: projected Rayleigh statistic (V, Eq. A.10), its variance (σ_V, Eq. A.11), its variance estimated using the Monte Carlo sampling introduced in Sec. B.1, and r values from channel pairs with $V > 3\varsigma_V$, where ς_V represents the population variance of V in the indicated velocity range, as estimated from Eq. 5. These values correspond to the relative orientation angles ϕ calculated between HI and ^{13}CO velocity-channel maps in the range $30 < v_{\rm LSR} < 60\,\mathrm{km\,s^{-1}}$.

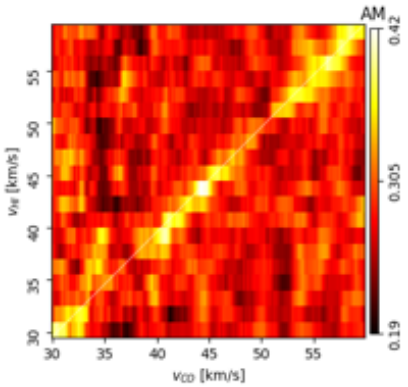

Fig. A.6: Alignment measurement (AM), the method for estimating the degree of alignment between vectors introduced in (Lazarian & Hoang 2007), for the relative orientation angles ϕ calculated between HI and ^{13}CO velocity-channel maps in the range $30 < v_{\rm LSR} < 60\,{\rm km\,s^{-1}}$.

effect of this gradient selection threshold and present an alternative method based on Monte Carlo sampling to quantify the effect of noise and propagate the observation errors into the HOG results.

B.1.1. Gradient selection threshold

One approach to the estimation of the V values is the selection of the gradient vectors based on the intensity signal-to-noise ratio (S/N), as described in Sec. A.2. Given that the gradients are independent of the intensity, this selection aims to guarantee that the relative orientation angles used in the calculation of V come from regions that are not dominated by noise. Although this approach is very practical, it reduces the number of samples on which to derive the projected Rayleigh statistic, as illustrated in Fig. B.1.

To illustrate the effect of the gradient selection on the values of V, we present in Fig. B.2 the results of the HOG analysis applied to the HI and ^{13}CO PPV cubes in the velocity range $30 \le v_{\rm LSR} \le 60\,{\rm km\,s^{-1}}$ after four different selections of the intensity S/N threshold. It is clear from the general distribution of V in this particular velocity range that the results of the HOG analysis do not critically depend on the selection of the threshold. There are two main reasons that potentially explain this result. First, the regions that dominate the signal in V are those with high intensity S/N and consequently, the V values are unaffected by the intensity threshold. Second, if the noise were uniform across the velocity-channel maps, the regions with low signal to noise would have randomly-oriented gradients that are uncorrelated in both tracers and thus would not significantly affect the distribution of V across velocity channels.

B.1.2. Error propagation in HOG

An alternative to the selection of gradients based on the intensity signal-to-noise ratio (S/N) is the use of Monte Carlo sampling to propagate the uncertainties in the observations into the HOG results. To do this, we generate draws $I^{\rm mc}_{i,j,l}$ from a Gaussian probability distribution function described by the mean value $I_{i,j,l}$ and the variance $(\sigma_I)_{i,j,l}$, where $I_{i,j,l}$ are the observed intensities in the PPV cubes and $(\sigma_I)_{i,j,l}$ is estimated from the low intensity S/N channels. As a zeroth-order approximation, we assume that $(\sigma_I)_{i,j,l} = \sigma_I$, that is, that the noise is constant in each velocity-channel map and across the velocity channels. The impact of the non-uniform noise distribution is further explored in Sec. B.2.

We report in Fig. B.3 the signal-to-noise ratios, V/σ_V, obtained from a Monte Carlo sampling in the velocity ranges $-5, \le v_{\rm LSR} \le 30$, $30 \le v_{\rm LSR} \le 60$, $60 \le v_{\rm LSR} \le 90$, and $90 \le v_{\rm LSR} \le 120\,{\rm km\,s^{-1}}$. These were obtained using 100 realizations of each $I^{\rm mc}_{i,j,l}$ value in the HI and ^{13}CO PPV cubes. It is clear

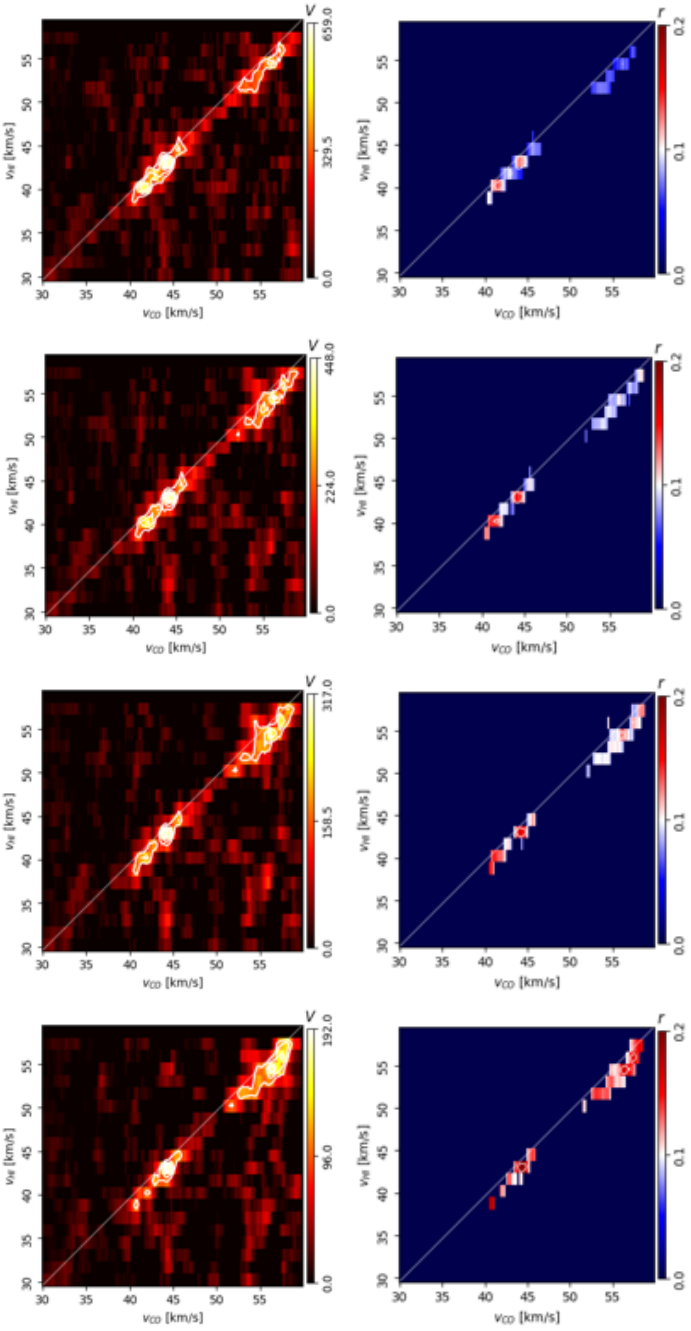

Fig. A.7: Results of the HOG analysis of the THOR HI and GRS ^{13}CO observations between $30 \le v_{\rm LSR} \le 60\,{\rm km\,s^{-1}}$ using derivative kernels with 46″ (top), 60″ (middle top), 75″ (middle bottom), and 105″ (bottom) FWMH. *Left*. Projected Rayleigh statistic, V. *Right*. Mean resultant vector length, r. The contours correspond to the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ values in the corresponding velocity ranges.

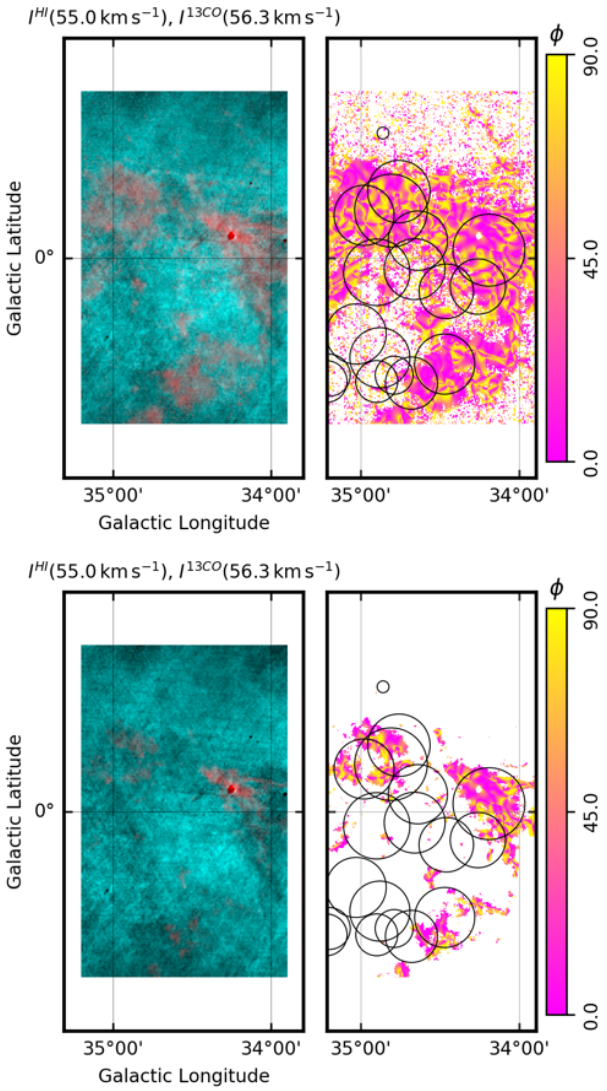

Fig. B.1: Intensity and relative orientation angle maps from the THOR HI and GRS ^{13}CO observations presented in Fig, 7. *Left.* HI (teal) and ^{13}CO emission (red) in the velocity channels with the largest spatial correlation in the velocity range $30 \leq v_{\rm LSR} \leq 60\,{\rm km\,s^{-1}}$, as inferred from the V values shown in Fig. 9. *Right.* Relative orientation angle ϕ, Eq. (1), between the gradients of the HI and ^{13}CO intensity maps in the indicated velocity channels. The white portions of the ϕ map correspond to areas where the gradient is not significant in either tracer, as estimated using the rejection criteria $I \geq \sigma_I$ (top) and $I \geq 7\sigma_I$ (bottom). The black circles in the right panels correspond to the positions and effective sizes of the MC candidates from the Rathborne et al. (2009) catalog in the aforementioned velocity range.

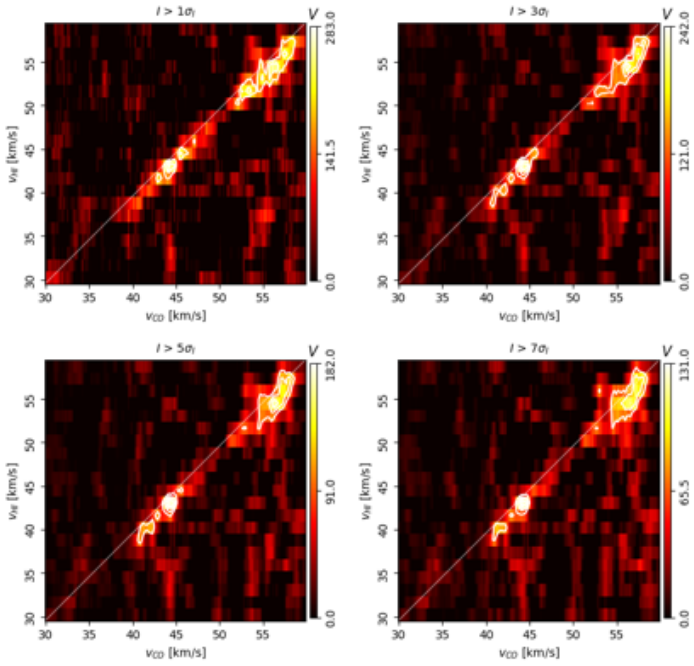

Fig. B.2: Projected Rayleigh statistic, V, the statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps in the HOG method. The panels correspond to the rejection criteria $I \geq \sigma_I$, $I \geq 3\sigma_I$, $I \geq 5\sigma_I$, and $I \geq 7\sigma_I$. The contours correspond to 3, 4, and $5\varsigma_V$, where ς_V is the population variance defined in Eq. (5).

from the distribution of V/σ_V that the confidence intervals are very similar to those obtained using the population variance ς_V introduced in Eq (5). This result reassures our assumption that velocity-channel maps in a broad range of velocities provide a good set of independent samples to determine the statistical significance of the observed V values.

In contrast to the intensity S/N selection, the Monte Carlo sampling does not reduce the number of samples from which to derive the projected Rayleigh statistic and fully propagates the measurement errors in the values of V. Additionally, it does not depend on the velocity range selected for the estimation of ς_V. However, it is computationally costly as it requires the HOG calculations of multiple realizations of each pair of velocity channels maps; the estimation of the values presented in Fig. B.3 requires a considerably larger amount of time and computational resources than those reported in Fig. 9. For the sake of simplicity and given the proximity in the results of both methods, we have chosen to report the statistical significance of the HOG results in terms of ς_V in the main body of this paper.

B.2. Impact of non-uniform noise in the HOG results

In radio interferometry, the imaging of a source requires the measurement of the two-dimensional spatial frequency spectrum. This spectrum is not fully sampled, as there are gaps between the antennas. Hence, the Fourier transformation of the spatial frequency spectrum coverage, which is the spatial resolution element of the array, is not a simple two-dimensional Gaussian function but a rather complicated structure commonly called the "dirty beam". During the imaging process, the "dirty beam" produces artifacts in the intensity distribution around a strong source that are commonly called side lobes. There are several different methods to remove the side lobes, the most common of them is the `clean` algorithm. One of the products of the `clean` algorithm is the residual image, which can be used as a good estimate for the level and spatial distribution of the noise in the observation (Bihr et al. 2016). Note that there is an additional contribution to the noise from the single-dish data used for the construction of the THOR HI products, but that contribution is expected to be uniform across the observed region.

Although in principle uncorrelated, the spatial distribution of the THOR HI noise can have some fortuitous correlation with the

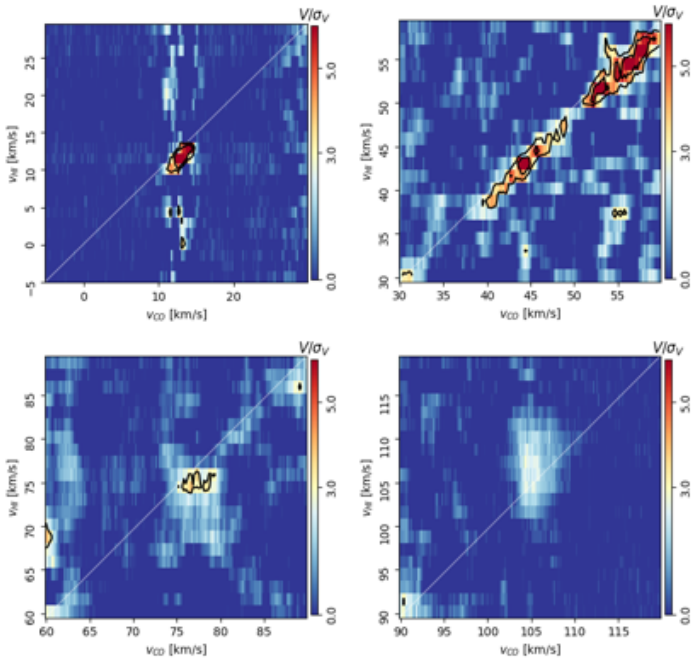

Fig. B.3: Signal-to-noise ratio, V/σ_V, of the projected Rayleigh statistic in four different velocity ranges. The contours correspond to $V/\sigma_V \geq 3$ and $V/\sigma_V \geq 5$, where σ_V is estimated using the Monte Carlo realizations of the data introduced in Sec. B.1.

distribution of GRS ^{13}CO. In order to test if such a correlation has a critical impact in the distribution of V across velocity channels, we calculated the HOG correlation between the GRS ^{13}CO observations and the HI noise cube constructed with the residual images of each velocity channel map. The results of this test, presented in Fig. B.4, indicate that there is some chance correlation between both data sets, although the spatial window used to determine the noise maps from the `clean` process introduces spatial correlations (see Bihr et al. (2016) for details) that increase the values of V and they cannot be directly compared with those in Fig. B.2. However, the distribution of V values across velocity channels is very different from the distribution of high V values reported in Fig. B.2, which is sufficient to demonstrate that the spatial distribution of the noise has no significant impact in the results presented in the body of this paper.

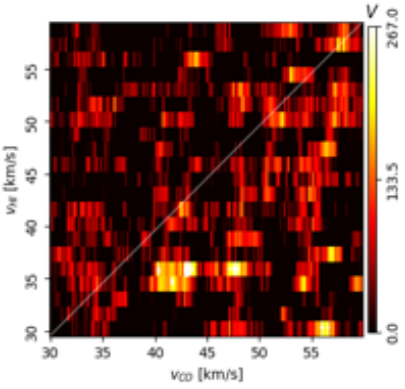

Fig. B.4: Projected Rayleigh statistic, V, resulting from the HOG analysis of the HI noise cube and the ^{13}CO PPV cube in the indicated velocity range.

B.3. Impact of chance correlation in the HOG results

Given that reproducing the statistical properties of each studied PPV cube is almost impossible, we test for chance correlation by using the same PPV cubes but flipping its spatial coordinates. Explicitly, we repeat the analysis using three test configurations:

- Keeping the original HI PPV cube but flipping the ^{13}CO PPV cube in the vertical (galactic latitude) direction.
- Keeping the original HI PPV cube but flipping the ^{13}CO PPV cube in the horizontal (galactic longitude) direction.
- Keeping the original HI PPV cube but flipping the ^{13}CO PPV cube in both the vertical and horizontal directions.

The results of these "flipping" tests, presented in Fig. B.5, show that the high-V regions around $v_{\rm HI} \approx v_{\rm 13CO}$ are not present when the spatial distribution of the ^{13}CO emission is not the observed one. This indicates that the high V values around $v_{\rm HI} \approx v_{\rm 13CO}$ are not the product of the concentration of the emission in particular velocity channels, but rather a significant correlation between the contours of both tracers in the corresponding velocity ranges.

In addition to the "flipping" tests we compute HOG for the HI PPV cube centered on $l = 34.^{\circ}55$, which is the subject of the analysis presented in this paper, with ^{13}CO PPV cubes with the same size but centered on $l = 32.^{\circ}05$, $33.^{\circ}3$, $35.^{\circ}8$, and $37.^{\circ}05$. The results of these tests, presented in Fig. B.6, indicate that the high-V regions around $v_{\rm HI} \approx v_{\rm 13CO}$ are not present when comparing these PPV cubes. These reinforces the conclusion that the high V values around $v_{\rm HI} \approx v_{\rm 13CO}$ are not the product of the concentration of the emission in particular velocity channels. The maximum values from these "offset" tests are significantly lower than those obtained in the original HOG analysis. Furthermore, the distribution of the V values in either the "flipping" or the "offset" tests is not related with those found in the original HOG analysis.

Appendix C: HOG in synthetic observations of MHD sims

C.1. Further analysis of the cloud collision simulations

We complement the analysis presented in Sec. 3 by considering the correlation between ^{13}CO synthetic observations and the HI emission, this time without the 100 K HI background emission. The results of the calculation of the oriented gradients in a pair of velocity-channel maps with high V values are presented in Fig. C.1. The spatial distribution of gradients with relative orientation angles $\phi \approx 0^{\circ}$ indicates that the correlation between the HI and ^{13}CO emission is not exclusively found in regions with HI self-absorption.

The HOGs, presented in Fig. C.2, also confirm that there is a significant population of parallel gradients, $\phi \approx 0^{\circ}$ in some channels, despite the fact that these synthetic observations do not include the HI self-absorption features produced with the 100 K HI background emission. These results confirm that the HI and ^{13}CO spatial correlation is not exclusively related to the HI self-absorption, where it is mostly expected, but that also the general distribution of the HI is related to that of the molecular gas. This does not contradict the conclusion that the HI and ^{13}CO correlation is mostly associated to the CNM, but indicates that the HI structure is correlated with the molecular gas in emission and not only in absorption.

There is, however, one significant difference between the spatial correlation of ^{13}CO and HI with and without the 100 K

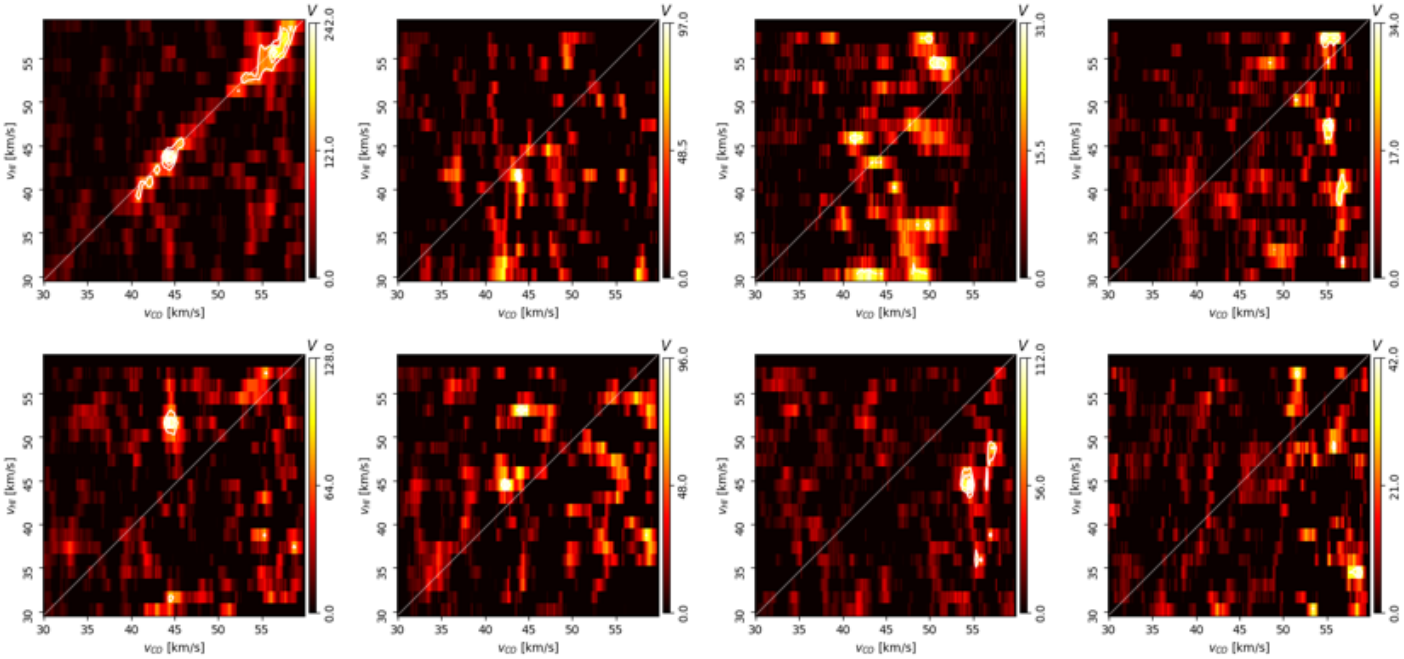

Fig. B.5: Projected Rayleigh statistic, V, the statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps in the HOG method, as defined in Eq. (3), after flipping the velocity channels maps in the following directions. *Top left*. Original THOR HI and GRS ^{13}CO tiles. *Top right*. Original THOR HI tile and GRS ^{13}CO flipped in the vertical (galactic latitude) direction *(*Bottom left). Original THOR HI tile and GRS ^{13}CO flipped in the horizontal (galactic longitude) direction. *Bottom right*. Original THOR HI tile and GRS ^{13}CO flipped in both the vertical and horizontal directions. The contours correspond to the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ values in the indicated velocity range. The large differences between the top-left and the other panels indicates that the effect of chance correlation in the results of the HOG method is small.

Fig. B.6: Same as Fig. B.5 for comparisons between the $\Delta l = 1.^{\circ}25$ THOR HI observations centered on $l = 34.^{\circ}55$ and the $\Delta l = 1.^{\circ}25$ GRS ^{13}CO tiles centered on $l = 32.^{\circ}05$ (top left), $33.^{\circ}3$ (top right), $35.^{\circ}8$ (bottom left), and $37.^{\circ}05$ (bottom right). The large differences between the top-left in Fig. B.5 and these panels indicates that the effect of chance correlation in the results of the HOG method is small.

HI background emission that is evident in the distribution of the V and r values across multiple velocities, as shown in Fig. C.3. While the V and r corresponding to the face-on synthetic observations seem unchanged and reveal high values centered exclusively around $v_{\rm HI} \approx v_{\rm 13CO} \approx 0\,{\rm km\,s^{-1}}$, the edge-on synthetic observations present correlation around two pairs of velocities: at $v_{\rm HI} \approx v_{\rm 13CO} \approx 0\,{\rm km\,s^{-1}}$ and around $v_{\rm HI} \approx 2\,{\rm km\,s^{-1}}$ and $v_{\rm 13CO} \approx 0\,{\rm km\,s^{-1}}$. Most likely, this correlation is a consequence of the HI edge of the shocked interface appearing in multiple velocity channels and being correlated with the ^{13}CO clouds at $v_{\rm 13CO} \approx 0\,{\rm km\,s^{-1}}$, an observation that is also consistent with the less-significant vertical stripe in the V values seen in Fig. C.3. Unlike the background-emission examples discussed in Sec. 6.1.2, the edge-on synthetic observations constitute a very rare scenario where the common boundary is maintained by the motions of the colliding clouds and highlighted by the particular line of sight orientation. It is implausible that this singular configuration provides an explanation for the spatial correlations at $v_{\rm HI} \neq v_{\rm 13CO}$ reported in Fig. 15.

C.2. Analysis of FRIGG simulations

In addition to the numerical simulations presented in Sec. 3 and Sec. C.1, we applied the HOG method to a set of synthetic observations from the stratified, 1-kpc scale, magneto-hydrodynamical simulations part of the FRIGG project (Hennebelle 2018). The ISM in these simulations is regulated by supernovae and include self-gravity, magnetic fields, cooling and heating processes, and a gravity profile that accounts for the distribution of stars and dark matter. The FRIGG simulations were designed to cover spatial scales between the intermediate galactic scales and the self-gravitating prestellar cores, ranging between the 1-kpc side simulation domain down to a maximum resolution of 3.8×10^{-3} pc. Further details on the initial conditions, zoom-in resolution strategies, and included physics are presented in Hennebelle (2018).

We extracted segments of the simulation around a set of density structures identified using a friend-of-friends (FoF) algorithm with a threshold density $n_0 = 10^3\,{\rm cm^{-3}}$ in a snapshot of the simulation taken at $t = 9.11$ Myr.x The size of the extracted volumes was set to be twice the effective diameter of the identified structures. Explicitly, the extraction yields to regular density, temperature, velocity, and magnetic field cubes with a common grid resolution, which we choose to be 0.12 pc. For the sake of simplicity, we focused in a structure with an effective radius of 10 pc, which was the fourth largest linked density structure identified by the FoF algorithm. Using the estimated proper motion of this structure, we traced back its evolution and extracted its properties in snapshots taken at $t = 9.06$, 9.01, 8.96 and 8.92 Myr.

C.2.1. Synthetic HI observations

We produce synthetic HI observations by applying the procedure described in Miville-Deschênes & Martin (2007), which is itself based on the general radiative transfer equations presented in

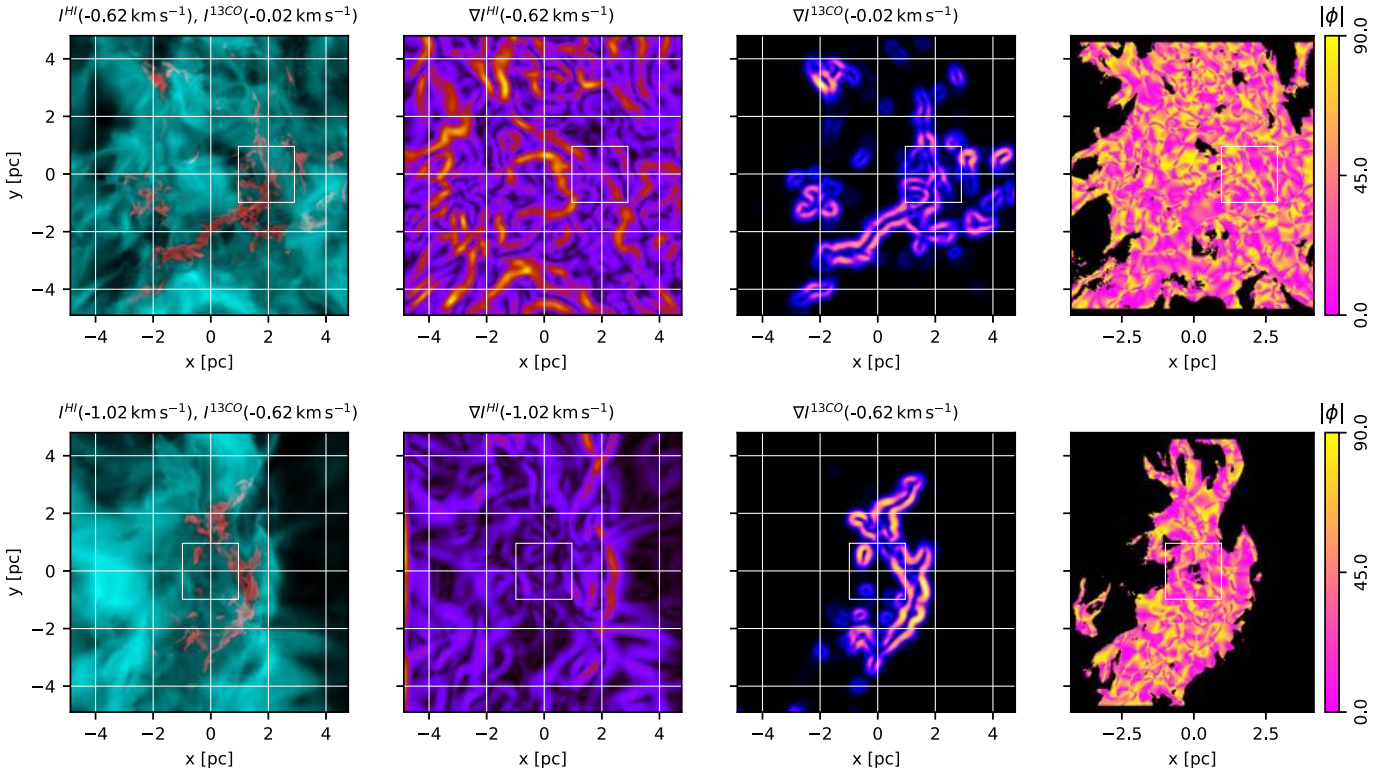

Fig. C.1: Intensity, intensity gradients, and relative orientation angle maps from the synthetic observations of the Clark et al. (2018) simulations. In contrast with the simulations presented in Sec. 3, the synthetic observation do not include the 100-K HI background emission. *Left.* HI (teal) and ^{13}CO emission (red) in the velocity channels with the largest spatial correlation, as inferred from the V values shown in Fig. 6. *Middle left.* Norm of the gradient of the HI intensity map in the indicated velocity channel. *Middle right.* Norm of the gradient of the ^{13}CO intensity map in the indicated velocity channel. *Right.* Relative orientation angle ϕ, Eq. (1), between the gradients of the HI and ^{13}CO intensity maps in the indicated velocity channels. The white color in the ϕ map corresponds to areas with no significant gradient in either tracer. The square indicates the block, selected from a 7×7 spatial grid, with the largest values of V. The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

Spitzer (1978) in the optically-thin regime. Given a density cube n_{ijk}; where the indexes i, j, and k run over the spatial coordinates x, y, and z, respectively; the brightness temperature map in the xy-plane at velocity u_l for a distant cloud with respect to its size is

$$[T_B]_{ij,l} = \frac{1}{C\sqrt{2\pi}} \sum_k \frac{n_{ijk}}{\sigma} \times \exp\left(-\frac{(u_l - [v_z]_{ijk})^2}{2\sigma^2}\right) \Delta z, \quad \text{(C.1)}$$

where $[v_z]_{ijk}$ is the velocity along the line of sight, which we assumed to be the z axis. The term $\sigma^2 \equiv k_B T_{ijk}/m$ represents the thermal broadening of the 21-cm line; T_{ijk} is the temperature cube, m is the hydrogen atom mass, and k_B is the Boltzmann constant. Finally, the constant $C = 1.813 \times 10^{18}$ cm^{-2} (Spitzer 1978).

Besides the assumption that the emission is optically thin, this treatment assumes that the excitation temperature (spin temperature) of the 21-cm line is the same as the kinetic temperature of the gas. The latter assumption is not representative of the general conditions in the ISM (see for example Kim et al. 2014, and the references therein), but it is sufficient for the our current exploration of the simulations. We present some selected spectra in the left-hand-side panel of Fig. C.4 and integrated emission maps in the left column of Fig. C.5.

C.2.2. Synthetic ^{13}CO observations

We produced synthetic ^{13}CO($J = 1 \rightarrow 0$) emission observations using an approach similar to the one described for HI, changing the corresponding coefficients and masses. To account for the ^{13}CO distribution in the simulated volume, we considered a critical density $n_c = 500$ cm^{-3} and a critical temperature $T_c = 100$ K. That is, we considered that only gas at $n > n_c$ and $T < T_c$ produces ^{13}CO emission. This strategy is evidently an over-simplification of the model and does not replace the proper treatment of a chemical network in the simulation, as in the Clark et al. (2018) simulations. However, this simple experiment can shed light on the spatial correlation between the dense and the diffuse gas across a range of velocities without intending to fully reproduce the complexity of the observations. We present some selected spectra in the right-hand-side panel of Fig. C.4 and integrated emission maps in the center column of Fig. C.5.

C.2.3. HOG analysis

The results of the HOG analysis of the synthetic observations of the selected portion of the FRIGG simulations in different time steps is presented on the right-hand-side column of Fig. C.5. It is clear from the distribution of values of the projected Rayleigh statistic, V, that the velocity channels with the highest correlation are not concentrated around $v_{HI} \approx v_{13CO}$, but there is a sig-

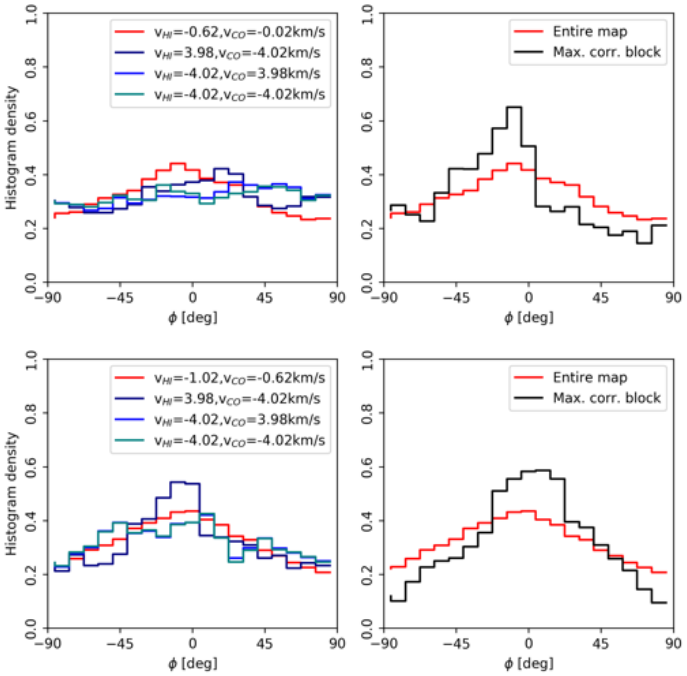

Fig. C.2: *Left*. Histograms of oriented gradients (HOGs) corresponding to the pair of velocity-channel maps with the largest spatial correlation, as inferred from the V values shown in Fig. C.3, and three pairs of arbitrarily selected velocity channels in the synthetic observations of the Clark et al. (2018) simulations without the 100-K HI background emission considered in Sec. 3. *Right*. For the pair of velocity-channel maps with the largest spatial correlation, HOGs corresponding to the entire map and just the block with the largest V indicated in Fig. C.1. The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

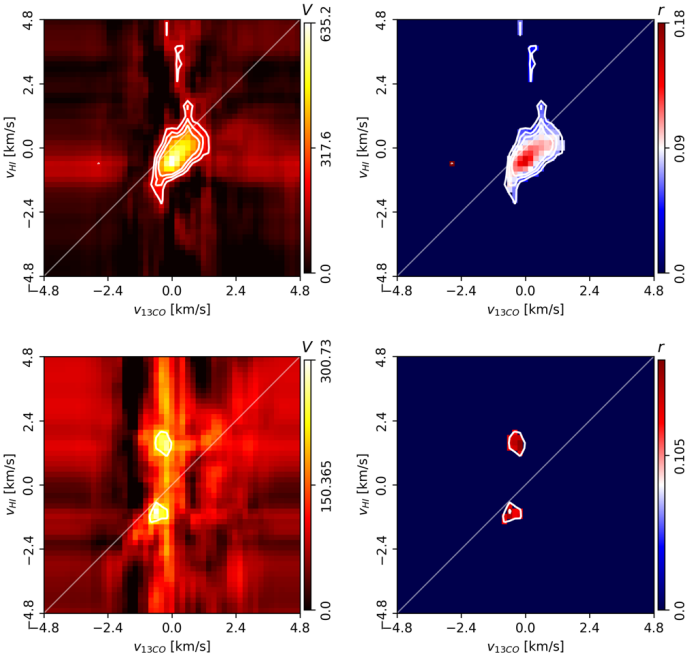

Fig. C.3: Results of the HOG analysis of the HI (without 100-K HI background emission) and ^{13}CO synthetic observations of the Clark et al. (2018) simulations. *Left.* Projected Rayleigh statistic, $V(v_{13CO}, v_{HI})$, the HOG statistical test of spatial correlation between HI and ^{13}CO velocity-channel maps, defined in Eq. (3). The contours indicate the $3\varsigma_V$, $4\varsigma_V$, and $5\varsigma_V$ levels in the corresponding velocity range. *Right.* Mean resultant vector length, $r(v_{13CO}, v_{HI})$, within the $3\varsigma_V$ confidence interval, a HOG metric that is roughly equivalent to the percentage of gradient pairs that imply the spatial correlation between the velocity-channel maps, defined in Eq. (2). The top and bottom panels correspond to the face-on and edge-on synthetic observations, respectively.

nificant level of correlation between velocity channel maps separated by approximately 1 km s^{-1}. This is evident in Fig. C.6, where we present the comparison between the velocity channels centered on $v_{HI} \approx 2.0$ and $v_{13CO} \approx 2.8$ km s^{-1}, which show the largest V values. Given the assumptions that we made to produce the synthetic HI and ^{13}CO synthetic observations, it is apparent that the spatial correlations in the HOG analysis of observations can be reproduced with the physical conditions included in the FRIGG simulations. However, the particular origin of the spatial correlation in HI and ^{13}CO velocity channels separated by a few kilometers per second is yet to be determined.

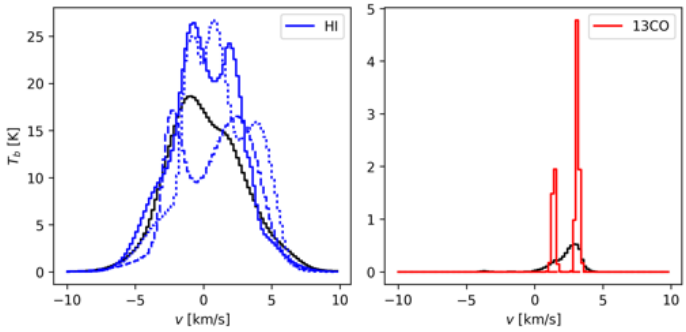

Fig. C.4: Spectra from the synthetic observations of HI and ^{13}CO emission from a segment of the FRIGG simulations in a snapshot taken at $t = 9.11$ Myr, also shown in the top panels of Fig. C.5. The black lines correspond to the average spectra over the whole map. The solid, dashed, and segmented colored lines correspond to the spectra toward the positions $[x, y] = [0,0]$, $[5,0]$, and $[0,5]$ pc, respectively.

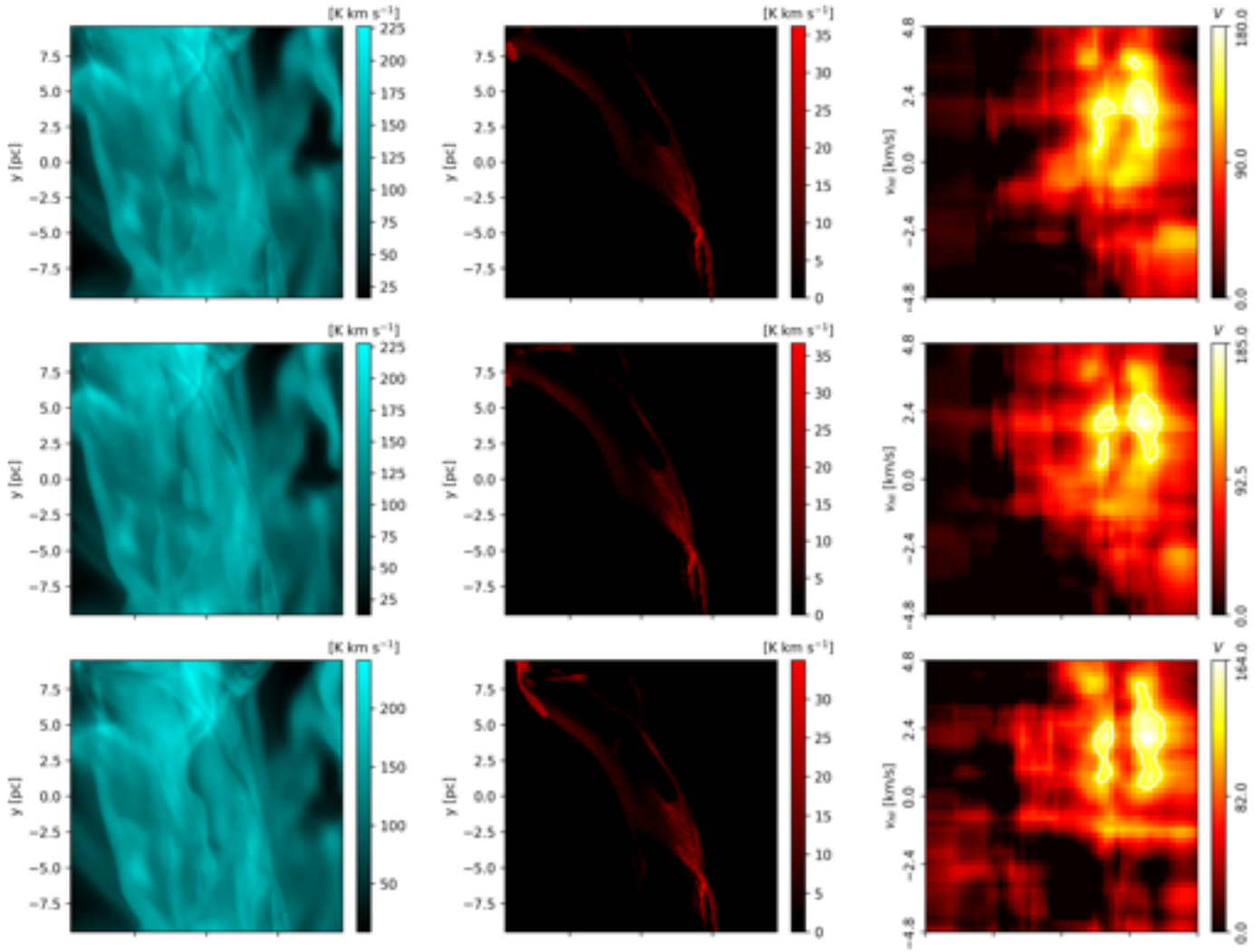

Fig. C.5: Integrated HI (left), ^{13}CO emission (center), and projected Rayleigh statistic (V, right) corresponding to the synthetic observations of a segment of the MHD simulations presented in Hennebelle (2018). From top to bottom, the panels represent snapshots taken at $t = 9.11$, 9.06, and 9.01 Myr, respectively.

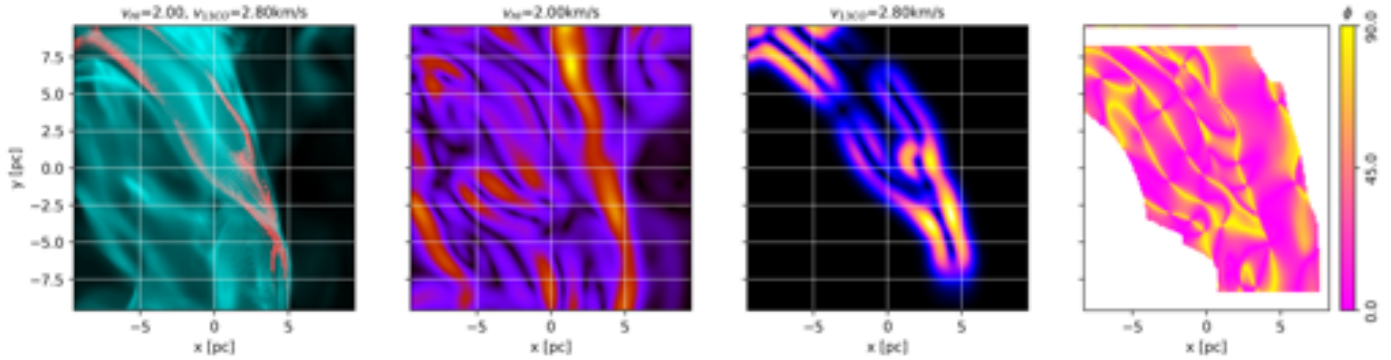

Fig. C.6: Intensity, intensity gradients, and relative orientation angle maps from the synthetic observations of a segment of the FRIGG simulation in the snapshot taken at $t = 9.11$ Myr. *Left.* Synthetic HI (teal) and ^{13}CO emission (red) in the velocity channels with the largest HOG correlation, as inferred from the values shown in Fig. C.5. *Middle left.* Norm of the gradient of the HI intensity map in the indicated velocity channel. *Middle right.* Norm of the gradient of the ^{13}CO intensity map in the indicated velocity channel. *Right.* Relative orientation angle ϕ, Eq. (1), between the gradients of the intensity maps shown in the left-hand-side panel.