

The integrated properties of the molecular clouds from the JCMT CO(3–2) High-Resolution Survey

D. Colombo,^{1★} E. Rosolowsky,² A. Duarte-Cabral,³ A. Ginsburg,⁴ J. Glenn,⁵
E. Zetterlund,⁵ A. K. Hernandez,⁶ J. Dempsey⁷ and M. J. Currie^{7,8}

¹Max Planck Institute for Radioastronomy, Auf dem Hügel 69, D-53121 Bonn, Germany

²Department of Physics, University of Alberta, 4-181 CCIS, Edmonton, AB T6G 2E1, Canada

³School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK

⁴National Radio Astronomy Observatory, 1003 Lopezville Rd, Socorro, NM 87801, USA

⁵CASA, Department of Astrophysical and Planetary Sciences, University of Colorado 389-UCB, Boulder, CO 80309, USA

⁶Department of Astronomy, University of Wisconsin, 475 North Charter Street, Madison, WI 53706, USA

⁷East Asian Observatory, 660 North A'ohoku Place, Hilo, Hawaii 96720, USA

⁸RAL Space, Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxfordshire OX11 0QX, UK

Accepted 2018 November 27. Received 2018 November 16; in original form 2018 October 12

ABSTRACT

We define the molecular cloud properties of the Milky Way first quadrant using data from the JCMT CO(3–2) High-Resolution Survey. We apply the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) algorithm to extract objects from the full-resolution data set, creating the first catalogue of molecular clouds with a large dynamic range in spatial scale. We identify more than 85000 clouds with two clear sub-samples: ~ 35500 well-resolved objects and ~ 540 clouds with well-defined distance estimations. Only 35 per cent of the catalogued clouds (as well as the total flux encompassed by them) appear enclosed within the Milky Way spiral arms. The scaling relationships between clouds with known distances are comparable to the characteristics of the clouds identified in previous surveys. However, these relations between integrated properties, especially from the full catalogue, show a large intrinsic scatter (~ 0.5 dex), comparable to other cloud catalogues of the Milky Way and nearby galaxies. The mass distribution of molecular clouds follows a truncated-power-law relationship over three orders of magnitude in mass with a form $dN/dM \propto M^{-1.7}$ with a clearly defined truncation at an upper mass of $M_0 \sim 3 \times 10^6 M_\odot$, consistent with theoretical models of cloud formation controlled by stellar feedback and shear. Similarly, the cloud population shows a power-law distribution of size with $dN/dR \propto R^{-2.8}$ with a truncation at $R_0 = 70$ pc.

Key words: methods: analytical – techniques: image processing – ISM: clouds – ISM: structure.

1 INTRODUCTION

Molecular clouds are the raw material for all star formation (SF) in the local Universe. These clouds are the initial seeds for the SF process, and their internal conditions dictate the relevant physics for SF and therefore the evolution of galaxies. The star-forming molecular interstellar medium (ISM) is cold ($T = 10$ – 30 K), relatively dense ($n_H \gtrsim 10^2 \text{ cm}^{-3}$), dominated by supersonic turbulence (Mach number $\mathcal{M} > 10$; Reynolds numbers $> 10^8$), and moderate magnetic fields (Alfvén speed comparable to flow speed $v_A \lesssim v$). Despite a rich interplay of physical processes, SF yields a surprisingly uniform initial mass function for stars. Given the wealth

of physical processes at work, the SF field has relied upon statistical characterization of the molecular gas to understand the range of conditions necessary for SF and the processes regulating gas evolution. In particular, molecular gas has been observed to organize into discrete molecular clouds (MCs) bounded by the photodissociation-regulated transition from atomic to molecular gas and the chemical creation of molecular gas tracers (most commonly CO). This chemical boundary provides a means by which molecular clouds can be inventoried and characterized as a population of discrete objects.

Historically, the molecular cloud ‘paradigm’ has provided a way by which the complexity of the molecular ISM could be simplified into the form mimicking that of the population studies that successfully drove the understanding of stellar populations and their evolution. In particular, wide area mapping of the Galactic

* E-mail: dcolombo@mpifr-bonn.mpg.de

plane in CO emission presented an opportunity to describe the star-forming molecular clouds across the Galactic disc. Given these survey data, several groups provided statistical descriptions of the Milky Way molecular ISM (e.g. Scoville & Solomon 1975; Solomon, Sanders & Scoville 1979; Sanders et al. 1986; Scoville et al. 1987; Solomon et al. 1987). Such cataloguing approaches encountered complications when applied to observational data sets where nominally discrete clouds appeared blended. These seminal studies adopted several approaches for describing both the blending of emission and the resolution of the distance ambiguity that affects kinematic distance determinations in the inner Milky Way. While some of these approaches relied on by-eye assignment of molecular emission into clouds, later work used contour-based methods. These contours considered the survey volume as a position–position–velocity (PPV) data cube of brightness temperature $T(l, b, v)$, where l , b , and v indicate, respectively, the Galactic longitude, latitude, and line-of-sight velocity. Clouds and their substructure were identified as discrete features of emission above fixed brightness temperature thresholds (Scoville et al. 1987; Solomon et al. 1987).

With these cloud definitions, the resulting analysis of molecular emission established the canonical scalings between the discrete features in the molecular ISM. In general, cloud properties were determined by measuring the (emission-weighted) size of the features in the survey space and resolving the distances to these clouds. These works showed that Milky Way molecular clouds followed a size–linewidth relationship suggestive of supersonic turbulence: $\sigma_v \propto R^\beta$; $\beta \in [0.4, 0.7]$, virial parameters $\alpha \sim 2$, and a top-heavy mass distribution with $dN/dM \propto M^\gamma$; $\gamma \in [-1.8, -1.5]$.

Interferometers showed that extragalactic studies of molecular clouds, when analysed using similar techniques, followed similar relationships between their bulk properties (Bolatto et al. 2008a; Fukui & Kawamura 2010), but high density systems (Oka et al. 2001; Wilson et al. 2003; Rosolowsky & Blitz 2005) showed significant departures from the scaling relationships seen in the Milky Way. In particular, those clouds exhibit larger turbulent linewidths on a fixed physical scale. None the less, the cloud populations still showed $\alpha \sim 2$. A careful re-examination of the molecular cloud properties in the Milky Way by Heyer et al. (2009) also found variation of the turbulent linewidths within the Galaxy. This work proposed another fundamental relationship between the molecular gas surface density Σ_{mol} and the linewidth on a fixed physical scale: $\sigma_v/R^{1/2} \propto \Sigma_{\text{mol}}$ (see also Field, Blackman & Keto 2011).

At the heart of these analyses is the definition of a discrete molecular cloud that provides a suitable basis for cataloguing. Since the approaches forwarded in the early works, the quality of survey data has improved dramatically. These improved data reveal that objects are *blended* in both the inner Milky Way and in the relatively low physical resolution studies of nearby galaxies with interferometers (Bolatto et al. 2008b). The edge of the CO emission in PPV space, even at a specified threshold, no longer serves as a good boundary to define an object. Several strategies have been proposed to deal with object identification in blended emission, and the primary approaches used historically fall into two main categories: functional fitting (e.g. GAUSSCLUMPS; Stutzki & Guesten 1990) and watershed algorithms (e.g. CLUMPFIND by Williams, de Geus & Blitz 1994; SEXTRACTOR by Bertin & Arnouts 1996; and CPROPS by Rosolowsky & Leroy 2006). The shortcomings of these approaches emerge in their application to the emission from the molecular ISM: because molecular gas is permeated by turbulence, the emission in this medium has structures on a wide range of scales. The emission structure is further filtered by chemical

(e.g. CO destruction or depletion), opacity, and excitation effects. Furthermore, the shapes of large scale molecular ISM do not have specific functional forms, though objects smaller than the thermal scale for turbulence appear to be well represented by physically motivated models such as Bonnor–Ebert spheres (i.e. cores; di Francesco et al. 2007).

In contrast to explicitly modelling the structure of the ISM emission, the watershed approach has the advantage of being model free. However, the major problem with this approach is that the blind application of watershed algorithms to emission with structure on a range of scales finds objects with scales comparable to the resolution element (Pineda, Rosolowsky & Goodman 2009; Leroy et al. 2016). This shortcoming has been avoided with several strategies that rely on prior information about the expected scales to be recovered in the emission. Heyer et al. (2009) simply use the cloud definitions established using lower resolution data of previous studies. Rathborne et al. (2009) smooth their data to a resolution of ~ 10 pc scales before applying CLUMPFIND. Well-resolved studies of extragalactic clouds (Bolatto et al. 2008b) use the notion of a ‘physical prior’ where the watershed algorithm is seeded on 10 pc scales comparable to the expected sizes of clouds. These approaches facilitate comparing data sets of disparate qualities, and allow population-based approaches to studying the molecular ISM. However, these decomposition approaches necessarily ignore the full dynamic range of information in the observational data set.

With several new surveys of the Galactic plane in emission from the molecular ISM, the gap between the quality of the available data and the tools used to define molecular clouds has grown particularly large. Recent studies have revisited the definition of cloud identification in the Milky Way. The combined survey of CO(1–0) emission over the Galaxy by Dame, Hartmann & Thaddeus (2001) (hereafter the Dame survey) provides a uniform reference for the molecular ISM at $l \approx 125^\circ$. Rice et al. (2016) use a *dendrogram* approach to create a cloud catalogue of this emission. The dendrogram representation transforms PPV data cubes into a tree-like graph that is defined by the connectivity of their emission contours (Rosolowsky et al. 2008). Identification of cloud structures is done by breaking the graph into individual trees based on specified criteria. In this case, the authors fix the amount of substructure that can present in a tree and tune this parameter to match the structures seen in the Dame survey. They use the Reid et al. (2016) distance determination code, which is based on trigonometric parallaxes to map PPV space to the three-dimensional structure of the Galaxy. This mapping assumes that all emission is concentrated in the spiral structure of the Galaxy and that each location in PPV space can be assigned to a unique distance.

Miville-Deschênes, Murray & Lee (2017) also use the Dame survey but adopt a complementary approach that first decomposes individual spectra into a family of Gaussian line components. Clouds are identified by clustering these components together into groups using assignment guided by a watershed analysis of the original PPV data set. The clustering approach defines clusters based on the brightest emission components and associates other components with these peaks if their coordinates are within the scatter of the coordinates for the peaks. Distances are assigned by assuming a zero-intrinsic-scatter size–linewidth relationship that accounts for variations in cloud surface density following Heyer et al. (2009). Clouds are assigned to a kinematic distance that gives a size most consistent with the (distance-independent) linewidth.

Both of these recent works present new approaches for identifying objects on scales larger than the resolution element, avoiding the main problems of watershed-based decomposition algorithms.

Since the resolution of the underlying Dame survey data is only 0.125° , these studies necessarily find objects significantly larger than this scale with little information on smaller scales. Other data sets have much better angular resolution, though they do not cover the entire Galactic plane uniformly. These include the Galactic Ring Survey [GRS, with 45 arcsec resolution in $^{12}\text{CO}(1-0)$; Jackson et al. 2006]; the Three-mm Ultimate Mopra Milky Way Survey [ThRUMMS, with 66 arcsec resolution in $^{12}\text{CO}(1-0)$ and other species; Barnes et al. 2015]; the JCMT $^{12}\text{CO}(3-2)$ High-Resolution Survey [COHRS, with 16.6 arcsec resolution in $^{12}\text{CO}(3-2)$; Dempsey, Thomas & Currie 2013]; the $^{13}\text{CO}/\text{C}^{18}\text{O}(3-2)$ Heterodyne Inner Milky Way Plane Survey [CHIMPS, with 15 arcsec resolution in $^{13}\text{CO}(3-2)$ and $\text{C}^{18}\text{O}(3-2)$; Rigby et al. 2016]; the Structure, excitation, and dynamics of the inner Galactic interstellar medium [SEDIGISM, with 30 arcsec resolution in $^{13}\text{CO}(2-1)$ and $\text{C}^{18}\text{O}(2-1)$; Schuller et al. 2017] survey; and the FOREST Unbiased Galactic plane Imaging survey with the Nobeyama 45-m telescope [FUGIN, with 20 arcsec resolution in $^{12}\text{CO}(1-0)$, $^{13}\text{CO}(1-0)$, and $\text{C}^{18}\text{O}(1-0)$; Umemoto et al. 2017]. The high spatial resolution and sensitivity of these data sets make it now possible to obtain vastly improved information about the molecular cloud population not only by being able to detect smaller and lower mass objects than those seen in previous studies, but also by the ability to detect variations in the substructure of the clouds. It becomes, therefore, essential to be able to extract molecular clouds by retaining the maximum amount of information on the hierarchical structure of the gas. This is one of the key advantages of the cloud extraction algorithm we employ here.

In this study, we present a new catalogue of molecular clouds that emphasizes the spatial dynamic range within recovered objects. We use the COHRS data of Dempsey et al. (2013) as the underlying data set because of its excellent spatial resolution. We then decompose these data using the Spectral Clustering for Interstellar Molecular Emission Segmentation algorithm (SCIMES; Colombo et al. 2015), which uses graph-based image processing to decompose a dendrogram representation of the emission into individual structures. Finally, we combine this decomposition with the distance catalogue generated in Ellsworth-Bowers et al. (2013).

We present this approach in the following sections. In Section 2, we briefly describe the COHRS. We summarize the SCIMES decomposition approach in Section 3 and its particular application to the COHRS data in Section 4. Sections 5 and 6 detail the distance determination procedure, and the determination of the $^{12}\text{CO}(3-2)$ -to- H_2 conversion factor, respectively. We describe the cloud properties in Section 7. Section 8 shows the content of the catalogue, the analysis of the cloud property distributions with the comparison with Roman-Duval et al. (2010) catalogue (Section 8.1), and the fit to the mass and size cumulative spectra (Section 8.2). We study those properties in relation to the cloud location within the Milky Way in Section 9 and we describe the correlations among the properties in Section 10. In Section 10.2 we contextualize the results from our catalogue considering other molecular cloud surveys of the Milky Way and nearby galaxies. Our findings and perspective for the future research with SCIMES are summarized in Section 11.

2 THE JCMT $^{12}\text{CO}(3-2)$ HIGH-RESOLUTION SURVEY

The COHRS is a large-scale CO survey that observed the inner Galactic plane in $^{12}\text{CO}(J=3 \rightarrow 2)$ emission using the Heterodyne Array Receiver Programme B-band instrument on the James Clerk Maxwell Telescope (JCMT). The current data release mapped a strip

of the Milky Way $|b| \leq 0.5^\circ$ between $10^\circ 25' < l < 17^\circ 5'$ and $50^\circ 25' < l < 55^\circ 25'$, and $|b| \leq 0.25^\circ$ between $17^\circ 5' < l < 50^\circ 25'$. The survey covers a velocity range of $-30 \text{ km s}^{-1} < v_{\text{LSR}} < 155 \text{ km s}^{-1}$. The data have a spectral resolution of 1 km s^{-1} and an angular resolution of $\theta_{\text{FWHM}} = 16.6 \text{ arcsec}$, achieving a mean noise level of $\sigma_{\text{RMS}} \sim 1 \text{ K}$. The $J=3-2$ transition of the ^{12}CO molecule traces the warm molecular medium (10–50 K) around the active SF regions. For full details about COHRS, refer to Dempsey et al. (2013).

3 SPECTRAL CLUSTERING FOR INTERSTELLAR MOLECULAR EMISSION SEGMENTATION

To decompose clouds in the COHRS data we use the publicly available Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES) algorithm.¹ The method has been explained fully in Colombo et al. (2015), and here we give only a brief description and note some changes relative to the original work in Appendix A. In general, SCIMES finds relevant objects within a dendrogram of emission using spectral clustering. A *dendrogram* is a tree representation of image data that encodes the hierarchical structure emission (e.g. Rosolowsky et al. 2008). The dendrogram is composed of two types of structures: *branches*, which are structures which split into multiple substructures, and *leaves*, which are structures that have no substructure. Leaves are associated with the local maxima in the emission. We also consider the *trunk*, which is the superstructure that has no parent structure, and contains all branches and leaves.

SCIMES uses graph theory to analyse dendrograms. A *graph* is a collection of objects (*nodes*) that possess defined relationships (*edges*). In this case, the edges connect a given branch to the substructures of a branch and the structures containing that branch. Under this interpretation, each edge in the graph corresponds to an isosurface (i.e. contour) in the PPV data. We specifically use weighted graphs, where each graph edge carries a numerical value called the *affinity* where larger values of the affinity represent more similarity between two parts of the graph. In SCIMES, we consider two different affinities corresponding to the properties of the isosurface based on the PPV *volume* and *luminosity*. The volume is defined as $V = \sigma_v \pi R_{\text{eff}}^2$, where σ_v is the velocity dispersion and R_{eff} is the effective radius of the isosurfaces (see Section 7 for further details). The luminosity is calculated as $L_{\text{CO}} = F_{\text{CO}} d^2$, where F_{CO} is the integrated emission within the isosurface (the *flux*) and d is the distance to the structure. If the distance to the structure is unknown the flux is considered. The affinity between two parts of the graph is defined as the inverse of the volume or the luminosity for the bigger or more luminous object. SCIMES generates an *affinity matrix*, where the element A_{ij} is the affinity between leaf i and leaf j , which correspond to the graph nodes.

The final part of SCIMES is to use the affinity matrix to divide the graph into separate components using *spectral clustering*, corresponding to segmenting the emission into individual clouds. The SCIMES algorithm considers the eigenvalues of the affinity matrix and finds the k most significant vectors that represent clusters in the spectral decomposition of the affinity matrix. Selecting these k vectors divides the graph into k regions, corresponding to structures within the dendrogram, which in turn correspond to connected regions of emission in PPV space. We generically call these objects

¹<https://github.com/Astroua/SCIMES>

molecular gas clusters since the literature can describe them as clouds, clumps, or cores depending on the scale of the emission.

While the SCIMES method is abstract and complex, it features the major advantage of being able to utilize data with wide spatial dynamic range (i.e. many resolution elements across a cloud). SCIMES has been developed to mimic the action of by-eye decomposition, but it is automated and requires no manual tuning. It relies on natural transitions in the emission structure to define objects and is robust across scales. In particular, SCIMES is a multiscale decomposition approach that explicitly takes the hierarchical nature of the ISM into account.

4 APPLICATION OF SCIMES TO COHRS DATA

The full first data release of the COHRS data is publicly available.² The data are provided in tiles of 0.5° in longitude. Before proceeding with the cloud identification we built a single data cube of the full survey using the SPECTRAL-CUBE³ and MONTAGE software. Given the large data set, we first construct a *dilated* mask of the data using the technique discussed in Rosolowsky & Leroy (2006). The mask is built using a two-step process, first including all PPV pixels that have emission above $10\sigma_{\text{RMS}}$ where σ_{RMS} is the local noise level. This mask is then expanded to all connected pixels that have emission above $3\sigma_{\text{RMS}}$ in two consecutive channels. In this way small clumps with $\text{SNR} < 10$ are incorporated within a larger structure avoiding noisy regions. The noise (σ_{RMS}) is estimated by calculating the standard deviation along each line of sight from the first and the last 10 line-free channels of the data cube. Using this mask, we can then extract sub-cubes from the full data cube that contain connected regions for decomposition with SCIMES. Those cubes span 1200 pixels in longitude, i.e. $\sim 2^\circ$ given the pixel size of the COHRS data of 6 arcsec. In practice, we perform this extraction in two stages, pulling out sub-cubes, processing those cubes with SCIMES and identifying objects on the longitude edges. Longitude edge objects are then rejected from the first catalogue pass ('main' SCIMES run) since their contours are by definition not closed (see Fig. 1). We then extract sub-cubes around the rejected edge objects and carry out a SCIMES decomposition ('filler' SCIMES run). We then include all objects that do not overlap between the two decompositions and retain the larger of any overlapping objects that occur in both passes. These two passages are performed to overcome the difficulty to generate a single dendrogram from the full COHRS data set which is computationally expensive. The same is true for the affinity matrix analysis performed by SCIMES where each additional required cluster is equivalent to an additional dimension in the clustering space. Several objects are also found along the survey actual (upper and lower) latitudinal edges. Those clouds are instead retained within the catalogue and marked as 'edge' clouds (see below).

For each sub-cube, we generate a dendrogram of the emission using SCIMES parameters that (1) require each branch of the dendrogram to be defined by an intensity change of $> 3\sigma_{\text{RMS}}$ ($\text{min_delta} = 3(\sigma_{\text{RMS}})$), (2) contain all of the emission in the mask ($\text{min_value} = 0$ K), and (3) contain at least three resolution elements worth of pixels ($\text{min_npix} = 3\Omega_{\text{bm}}$, where Ω_{bm} is the solid angle of the beam expressed in pixels). We do not know the distance of the all dendrogram structures a priori, so the *volume* and *luminosity* affinity matrices are generated using the PPV

volumes and integrated intensity values instead of spatial volumes and intrinsic luminosities. The scaling parameter (see Colombo et al. 2015) is searched above $3\sigma_{\text{RMS}}$ (see Appendix A for further details). SCIMES searches for the dendrogram branches that can be considered as single independent 'molecular gas clusters'. As discussed in Colombo et al. (2015), leaves that do not form isolated clusters, are grouped all together in sparse clusters without any neighbours in PPV space between constituent objects. In the original implementation of SCIMES, a sparse cluster was pruned by those leaves and only the largest branch (i.e. the branch that contains the largest number of leaves) was retained as representative of the sparse cluster. The structures pruned from the sparse cluster can consist of isolated leaves or small branches. Elements that cannot be assigned to any cluster given the clustering criterion are called 'noise' in clustering theory (e.g. Ester et al. 1996). Here, however, we retain those branches and leaves since they are significant emission and well-resolved objects (given the dendrogram construction parameters) even if not assigned to another cluster.

5 DETERMINATION OF $^{12}\text{CO}(3-2)$ -TO- $^{12}\text{CO}(1-0)$ FLUX RATIO

To calculate the masses and column densities of molecular clouds, we use a CO-to- H_2 conversion factor X_{CO} , allowing us to scale the integrated intensities of the CO emission, W_{CO} , directly to H_2 column densities, N_{H_2} . In general, X_{CO} is a parameter that depends on environmental conditions (e.g. Bolatto, Wolfire & Leroy 2013; Barnes et al. 2018). Even so, for the purpose of this paper, we assume a constant X_{CO} for direct comparison to literature results, with $X_{^{12}\text{CO}(1-0)} \sim 2 \pm 1 \times 10^{20} \text{ cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$ (e.g. Dame et al. 2001; Bolatto et al. 2013; Duarte-Cabral et al. 2015). We express masses in terms of the molecular gas luminosity using the scaled conversion factor $M_{\text{lum}} = \alpha_{\text{CO}} L_{\text{CO}}$, with $\alpha_{^{12}\text{CO}(1-0)} \sim 4.35 M_{\odot} \text{ pc}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$, which assumes a mean molecular weight of $2.8m_{\text{H}}$ per hydrogen molecule (e.g. Bolatto et al. 2013). These conversion factors are calibrated using the $^{12}\text{CO}(1-0)$ transition, and therefore we must assume a line ratio $R_{31} \equiv ^{12}\text{CO}(3-2)/^{12}\text{CO}(1-0)$ to scale our calculated properties directly to physical properties.

To calculate R_{31} , we convolve the COHRS data to 45 arcsec and reproject it to the coordinate grid of the $^{12}\text{CO}(1-0)$ data from the University of Massachusetts Stony Brook Survey (Sanders et al. 1986). We show the pixel-by-pixel plot of the brightness in the two data cubes in Fig. 2. We use these relationships to estimate the typical line ratio across the survey. There is not a single, well-defined line ratio across the Galaxy reflecting the variation in excitation conditions (e.g. Barnes et al. 2015; Peñaloza et al. 2018). On average, we observe $R_{31} \equiv ^{12}\text{CO}(3-2)/^{12}\text{CO}(1-0) = 0.4$ for the faint emission, but $R_{31} = 0.6$ describes the line ratio of the brightest emission. As a global summary across the survey region, we adopt $R_{31} = 0.5$ and use this line ratio to establish the CO-to- H_2 conversion based on the calibration work done for $\alpha_{^{12}\text{CO}(1-0)}$. This line ratio value is consistent to the value measured in nearby galaxies by Warren et al. (2010). In this way we scale from the standard values for the CO(1-0) line (Bolatto et al. 2013) to get

$$X_{^{12}\text{CO}(3-2)} = \frac{X_{^{12}\text{CO}(1-0)}}{R_{31}} = 4 \times 10^{20} \frac{\text{cm}^{-2}}{\text{K km s}^{-1}}, \quad (1)$$

and

$$\alpha_{^{12}\text{CO}(3-2)} = \frac{\alpha_{^{12}\text{CO}(1-0)}}{R_{31}} = 8.7 \frac{M_{\odot}}{\text{K km s}^{-1} \text{ pc}^2}. \quad (2)$$

²<http://dx.doi.org/10.11570/13.0002>

³<https://doi.org/10.5281/zenodo.1213217>

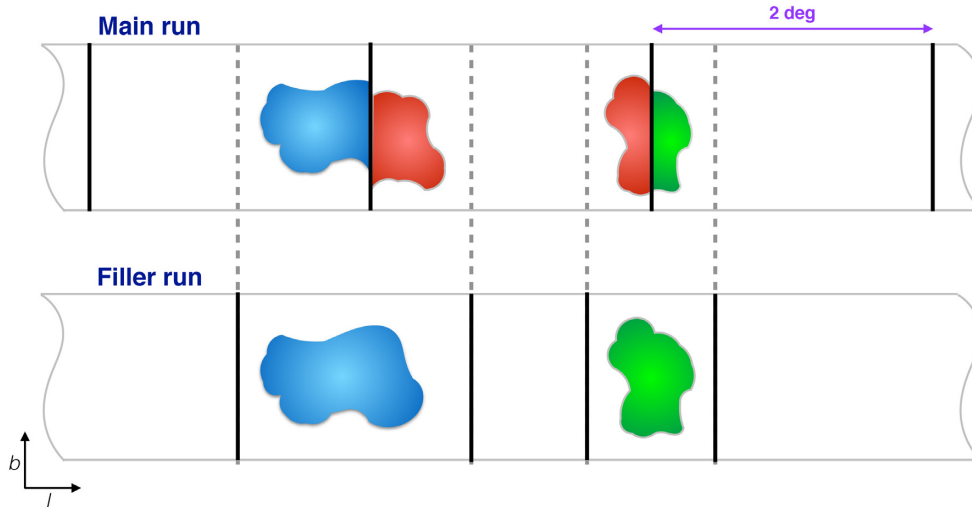


Figure 1. Scheme of the full COHRS data (grey, full lines) division into sub-cubes that define the ‘main’ (top) and ‘filler’ (bottom) runs. Sub-cubes of the ‘main’ run span all approximately 2° in longitude. In the figure, sub-cube edges are indicated with black vertical lines. Clouds on the longitudinal edges are removed from the ‘main’ run cubes since they do not have closed iso-contours. In the figure, clouds in the same sub-cube are indicated with the same colour. To recover the clouds on the longitudinal edges of the ‘main’ run cubes we build customized, ‘filler’ run, sub-cubes which span the full extent of the largest clouds on two consecutive edges of two adjacent ‘main’ run sub-cubes, allowing a padding of 20 pixels on both sides. In some cases, the clouds recovered in the ‘filler’ run do not always perfectly correspond to the two chunks of edge clouds of the ‘main’ run.

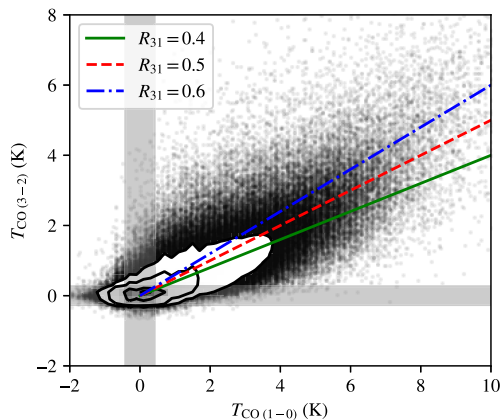


Figure 2. Pixel-by-pixel brightness comparison between $^{12}\text{CO}(1-0)$ from the Massachusetts-Stony Brook Galactic Plane CO Survey (Sanders et al. 1986) and the $^{12}\text{CO}(3-2)$ from COHRS. The grey-shaded regions indicate the $\pm 1\sigma$ levels of the noise. We obtain an average $R_{31} = ^{12}\text{CO}(3-2)/^{12}\text{CO}(1-0) = 0.6$ for the bright emission (blue dashed dot line), while a value $R_{31} = 0.4$ is appropriate for faint emission (green full line). For the whole sample we choose the average value of $R_{31} = 0.5$ (red dashed line).

The data in Fig. 2 show significant scatter in the low end and this simple conversion of CO luminosity to mass becomes even less certain for these low luminosity objects. Given this, the line ratio alone suggests an uncertainty of at least 40 per cent in referencing to $\alpha_{^{12}\text{CO}(1-0)}$.

6 DISTANCE TO THE MOLECULAR CLOUDS

We use the distances defined by Zetterlund et al. (2018; see also Ellsworth-Bowers et al. 2013, 2015) to establish the physical properties to the decomposed molecular gas clusters. These distance estimates are based on an analysis of the Bolocam Galactic Plane

Survey (BGPS) version 2 (Ginsburg et al. 2013) and a suite of multiwaveband data.

The BGPS is a 1.1 mm dust continuum survey that largely overlaps with COHRS, covering the Galactic plane region between $-10^\circ \leq l \leq 90^\circ$ and $|b| \leq 0.5$ with a spatial resolution of 33 arcsec, making it particularly useful for our needs. The distances are obtained using a Bayesian approach that provides a posterior probability density function of distances to an object through a suite of techniques including kinematic distances from matching to molecular gas emission, maser distances, and a model of the infrared emission and absorption of different features in the plane. In most cases, the method allows the definition of a single distance through the maximum-likelihood distance or the probability-weighted mean distance. Each object then has a distance assigned to it from one of these two methods according to its ability to resolve the kinematic distance ambiguity (see Ellsworth-Bowers et al. 2013 for further details). This method generates both estimates of the distances and known uncertainties. We focus on the 2202 objects with well-constrained distances in the BGPS distance sample, i.e. those with distance uncertainties of ± 0.5 kpc or better.

For BGPS objects with a single, well-constrained distance, this measurement corresponds to a single position (pixel) within the PPV data as their spectroscopic v_{LSR} are also defined (see Shirley et al. 2013). That pixel can be associated with the dendrogram structure that contains the known-distance pixel, which then inherits that distance measurement. Around 140 decomposed molecular gas clusters (~ 0.2 per cent of the total number of clouds) have substructures that have different distances from each other. For these objects, we assign the cluster to the near distance corresponding to the brightest spot within the object, assuming that that sub-structure carries the largest amount of cloud mass.

We call this distance attribution method *exact*. In contrast, a molecular gas cluster may not contain any pixels with a distance measurement. These objects may be the substructures of larger connected emission features with distance assignments. In this case, we assign the cluster the closest distance in PPV space contained within the larger structure and describe this assignment method as

a *broadcast*. For isolated objects without an assigned distance or a parent structure with a distance, we assign the object a distance to be the same as the *closest* PPV pixel with a known distance. In this way we are able to break the kinematic distance degeneracy at least for the objects with better defined distance (see Appendix B for further details). We also report a parameter, called *broadcast inaccuracy*, which indicates (in pixels) how far is a distance pixel to the outer edge of the cloud. By definition *exact* distance clouds have broadcast inaccuracy equal to zero.

7 MOLECULAR CLOUD INTEGRATED PROPERTIES

For most of the cloud decomposition in the literature, the size of clouds is governed by the resolution element of the survey. Therefore, the internal structure of these objects remains elusive and cataloguing focuses on the ‘integrated’ properties obtained by summing across the cloud pixels. While SCIMES is able to retain the internal structures, we focus this work on the integrated properties for comparison with the existing literature and we will present studies based on the well-resolved nature of COHRS clouds in future work.

7.1 Coordinates

For each cloud in our catalogue we provide four sets of coordinates in different projections: pixel, Galactic, heliocentric, and Galactocentric.

The pixel coordinates ($x_{\text{cen}}, y_{\text{cen}}, v_{\text{cen}}$) are the mean positions of the clouds related to the sub-cube they have been segmented from. Galactic coordinates are listed as longitude (l), latitude (b), and velocity with respect to the local standard of rest (v_{LSR}). Heliocentric coordinates are defined with the x -axis on the line that connects the Sun and the Galactic centre:

$$\begin{aligned} x_{\odot} &= d \cos l \cos b, \\ y_{\odot} &= d \sin l \cos b, \\ z_{\odot} &= d \sin b, \end{aligned} \quad (3)$$

where d is the cloud assigned distance.

Following Ellsworth-Bowers et al. (2013) and Rice et al. (2016) we define the Galactocentric coordinates as

$$\begin{aligned} x_{\text{Gal}} &= R_0 \cos \theta - d(\cos l \cos b \cos \theta + \sin b \sin \theta), \\ y_{\text{Gal}} &= -d \sin l \cos b, \\ z_{\text{Gal}} &= R_0 \sin \theta - d(\cos l \cos b \cos \theta - \sin b \cos \theta), \end{aligned} \quad (4)$$

where $\theta = \arcsin(z_0/R_0)$, and $z_0 = 25$ pc is the height of the Sun above the mid-plane of the Milky Way (Goodman et al. 2014), and $R_0 = 8.51$ kpc is the radius of the Solar circle, i.e. the distance of the Sun to the Galactic Centre (Ellsworth-Bowers et al. 2013).

7.2 Pixel-based properties

Many properties of the isosurfaces are already defined by the dendrogram implementation we used. Flux and Volume are employed to generate the affinity matrices necessary for the cluster analysis. The properties considered by the dendrogram are ‘pixel-based’ if distances are not provided a priori. The statistics offered by the ASTRODENDRO software have been defined by Rosolowsky & Leroy (2006).

The basic properties of the dendrogram structures are calculated through the moment technique. This technique has been shown to perform better than the area method to recover the actual effective

radius of the clouds (see Rosolowsky & Leroy 2006 for details). The centroid of the clouds is given by the intensity-weighted mean of the pixel coordinates along the two spatial and the velocity axes of the datacube. The principal axes of the emission map using intensity-weighted second moments over position defines the major (σ_{maj}) and minor axis (σ_{min}) sizes of the cloud as well as the position angle (ϕ). In a similar way, we calculate the velocity dispersion (σ_v) from the intensity-weighted second moment along the velocity axis. Through the major and minor semi-axis measurements, we also define an area measurement for clouds based on the elliptical area:

$$A_{\text{ellipse}} = (8\pi \ln 2) \sigma_{\text{maj}} \sigma_{\text{min}}. \quad (5)$$

We also measure the area of each cloud from the number of pixels that the cloud occupies in the spatial dimensions (A_{exact}). Finally, the flux (F_{CO}) is given by the sum of all the pixel values within the isosurface that contains the cloud.

We enrich the dendrogram-generated catalogue with several other properties related to the hierarchical structure itself (see Appendix A). Each final cloud has the number of pixels and leaves within the structure (N_{pixel} and N_{leaves} , respectively), the identifier for the parental structure that fully contains the structure (*parent*), the identifier for the structure at the bottom of the hierarchy that fully contains the structure under analysis (*ancestor*), and classification flag (*type*) that indicates which kind of structure we are dealing with (in dendrogram terminology): ‘L’ (leaf) a structure without children; ‘B’ (branch) a structure with children and parent; and ‘T’ (trunk) structure with children and without parent (bottom of the hierarchy).

7.3 Physical properties

By assigning a distance to each molecular gas cluster as per Section 6, we can convert the pixel-based properties into physical properties of the molecular structures. The semimajor and semiminor axes are converted into parsecs via the usual small angle formula and the world-coordinate-system information from the data file. The effective radius of the cloud (R_{eff}) is generated from quadrature sum of the semimajor and semiminor axes:

$$R_{\text{eff}} = \eta \sqrt{\sigma_{\text{maj}}^2 + \sigma_{\text{min}}^2}. \quad (6)$$

Here, $\eta = 1.91$ is assumed from Rosolowsky & Leroy (2006) (and Solomon et al. 1987) to relate the quadrature sum of the two semi-axes and the radius of a spherical cloud. The final velocity dispersion of the cloud is obtained by multiplying the channel width, Δv , by the velocity dispersion measured in pixels.

The CO luminosity of the cloud is obtained by

$$L_{\text{CO}} = \sum_i T_{\text{CO}}^i \Delta v \Omega_{\text{pix}} d^2, \quad (7)$$

where d is the distance to the cloud (in parsecs), Ω_{pix} is the solid angle subtended by a pixel, Δv is the channel width (in km s^{-1}), and T_{CO}^i is the brightness temperature of each voxel i within the cloud (in K).

The effective radius, velocity dispersion, and CO luminosity are the basis for all other properties presented in the catalogue. The mass is derived from the CO luminosity (or luminosity mass) by assuming a $^{12}\text{CO}(1-0)$ -to- H_2 conversion factor α_{CO} and R_{31} (see Section 5).

We also make a dynamical measurement of the cloud mass assuming the clouds are virialized, spherical objects, with an internal

density profile that scales with the cloud radius like $\rho(r) \propto r^{-1}$. We called this measurement virial mass:

$$M_{\text{vir}} = 1040\sigma_v^2 R_{\text{eff}}, \quad (8)$$

where we ignore the contribution from external pressure and magnetic fields (Rosolowsky & Leroy 2006).

The ratio between virial and luminous mass gives the virial parameter, α , which is often used to characterize the deviation of a cloud from virial equilibrium. Variations on the estimated α can be due to a true unbalance between physical effects such as gravity, pressure, and magnetic fields, as well as due to time evolution, or simply due to observational biases, such as the possible variations of the assumed α_{CO} (Bertoldi & McKee 1992; Houlahan & Scalo 1992):

$$\alpha = 1.12 \frac{M_{\text{vir}}}{M_{\text{lum}}}. \quad (9)$$

In general, $\alpha > 2$ would indicate that the cloud is stabilized against collapse, while finding $\alpha \ll 2$ might suggest a significant magnetic support (e.g. Kauffmann, Pillai & Goldsmith 2013). Generally $\alpha \approx 1$ means that the cloud is virialized.

Through the measurements of M_{lum} and R_{eff} we measure the mean molecular mass surface density and volume density by assuming clouds have an uniform density and a spherical shape with radius R_{eff} :

$$\Sigma_{\text{mol}} = \frac{M_{\text{lum}}}{\pi R_{\text{eff}}^2}, \quad (10)$$

$$\rho_{\text{mol}} = \frac{3M_{\text{lum}}}{4\pi R_{\text{eff}}^3}. \quad (11)$$

7.4 Extrapolation and deconvolution

The dendrogram implementation we choose assumes a bijection paradigm to calculate the properties of the structures (see Rosolowsky et al. 2008), i.e. there is a direct connection between pixel intensities in PPV space and the corresponding emission in real space. In this approach, a clump of emission is associated with a physical structure above a certain column density threshold. Rosolowsky & Leroy (2006) showed that cloud properties are strongly dependent on the brightness level at which they are identified above the surrounding emission. Following Rosolowsky & Leroy (2006), we consider an alternative measure of cloud properties that attempts to correct for the biases introduced by finite sensitivity and resolution.

The first step in this procedure is the *extrapolation* (indicated as ‘ex’ in the catalogue), which infers the moments of the cloud (σ_{maj} , σ_{min} , σ_v , and F_{CO}) that would be measured with infinite sensitivity. The extrapolation works by considering the scaling of cloud properties as a function of brightness threshold and fitting a linear relation between the measured moments (σ_{maj} , σ_{min} , and σ_v) and the brightness. This relationship is extrapolated to the 0 K contour. A similar procedure is used to derive the extrapolated flux, except a quadratic extrapolation is used. In the rest of the section, we indicate the extrapolated properties as $\sigma_{\text{maj}}(0 \text{ K})$, $\sigma_{\text{min}}(0 \text{ K})$, $\sigma_v(0 \text{ K})$, and $F_{\text{CO}}(0 \text{ K})$.

The second step consists in the deconvolution of the survey beam and channel width from the extrapolated moments. The deconvolution is performed by subtracting the beam width in

quadrature from the measured radius:

$$\sigma_{\text{maj,dc}}^2(0 \text{ K}) = \sigma_{\text{maj}}^2(0 \text{ K}) - \left(\frac{\theta_{\text{FWHM}}}{\sqrt{8 \ln(2)}} \right)^2 \quad (12)$$

with a similar expression for the minor-axis. The channel width is also deconvolved from the linewidth:

$$\sigma_{v,dc}^2(0 \text{ K}) = \sigma_v^2(0 \text{ K}) - \frac{\Delta v^2}{2\pi}, \quad (13)$$

where the subscript ‘dc’ indicates deconvolved properties, θ_{FWHM} is the survey beam, and Δv is the channel width. All the other properties are then recalculated using these extrapolated, deconvolved measurements. We use these extrapolated, deconvolved properties as the basis for our analysis. In Appendix D, we explore how the extrapolation and deconvolution affects the inferred cloud properties. Generally, the deconvolution affects mostly low values of R_{eff} and σ_v . Extrapolation, instead, shifts some of the velocity dispersion values by up to ~ 0.5 dex and L_{CO} by to ~ 1 dex, while it does not change measured effective radii significantly.

We choose to apply the extrapolation corrections so that all cloud properties are referenced to a common intensity threshold, which facilitates comparing the cloud properties to each other. Without a common reference threshold, each cloud would be subject to different biases in the measured properties (Rosolowsky & Leroy 2006). However, this application then engenders a specific scientific interpretation of the results, namely we are estimating cloud properties, defining such objects as bounded by a 0 K intensity isosurface. Since these emission structures are part of larger, hierarchical ISM, enforcing this interpretation can obscure the true complexity of the ISM. Indeed, as we note in Section 8, where the emission is heavily blended, SCIMES will segment structures high above the noise level of the data, and this extrapolation may effectively overcorrect for the amount of emission of each cloud and its extent. Nevertheless, this cloud-segmentation approach is selected so we can create a catalogued set of objects, which can then be compared to other work executing similar analyses. Our catalogue provides both corrected and uncorrected properties so that other work could use the same SCIMES decomposition without these corrections, provided such an interpretation suits the question being investigated.

7.5 Uncertainties on cloud properties

The uncertainties on the physical properties in our cloud catalogue are dominated by two sources of errors: the distance (d) and the CO-to-H₂ conversion (α_{CO}). Therefore, we use error propagation by taking into account the uncertainty on the distances provided by Zetterlund, Glenn & Rosolowsky (2018) and by assuming 40 per cent error on our calculated α_{CO} (see Section 5). Using the CO-to-H₂ conversion factor method also introduces systematic uncertainties at the factor of ~ 2 level (Bolatto et al. 2013). For the ‘closest’ distance objects we use the near–far distance ambiguity as distance uncertainty given by

$$\delta d = |R_0 \cos l|, \quad (14)$$

where $R_0 = 8.51 \text{ kpc}$ (Ellsworth-Bowers et al. 2013) and l is the Galactic longitude of the cloud centroid.

For purely pixel quantities (σ_{maj} , σ_{min} , σ_v , and F_{CO}) we use the bootstrap approach described in Rosolowsky & Leroy (2006). This method generates several synthetic clouds by considering a cloud as a set of N volumetric pixels with coordinates x_i , y_i , v_i , and T_i ; i.e. two spatial coordinates, one velocity coordinate, and the brightness value, respectively. At each iteration, N sets of the cloud

data are sampled randomly from the observed values allowing for repeated draws. The sets of bootstrapped σ_{maj} , σ_{min} , σ_v , and F_{CO} are measured at each iteration. The uncertainty is given by the standard deviation of the bootstrapped quantities. We also rescaled each uncertainty by an oversampling rate, given by the square root of the pixels in the beam. The oversampling rate accounts for not all pixels in each cloud being independent. These bootstrap uncertainties are summed in quadrature with the uncertainties induced by the distance and conversion factors. While the distance and conversion factors are both typically 40 per cent, the uncertainties in the sizes ($\delta\sigma_{\text{maj}}$, $\delta\sigma_{\text{min}}$) are typically 15 per cent and the flux uncertainty (δF_{CO}) is typically 6 per cent.

8 MOLECULAR CLOUD CATALOGUE

Fig. 3 shows an example of a cloud segmented by SCIMES. The figure illustrates how SCIMES naturally works on a multiscale data. In Appendix E the longitude–latitude and longitude–velocity masks for the clouds in the full survey are given. Those figures show that the SCIMES approach identifies a variety of cloud morphologies, complexity, and sizes. The segmented structures are mostly coherent, however multiple velocity components are sometimes merged in the same object. This behaviour is especially associated with clouds on the border of the data cubes that do not have closed contours. While the objects on the right and left edges of the sub-cubes are removed by construction, the clouds on the lower and upper edges are retained in the catalogue since the span in latitude is an intrinsic limitation of the survey rather than the algorithm. The same is true for the clouds on the left edge of the first sub-cube and on the right edge of the last sub-cube which constitute the outermost sections of the survey.

The catalogue we produced contains the data listed in Table 1. The whole catalogue is made of 85 020 objects: 73 140 (86 per cent) are leaves and 11 880 (12 per cent) are branches. Dendrogram leaves dominate the catalogue. These leaves are generally small, isolated structures with sizes comparable to the imposed minimum size limit for the inclusion in the dendrogram (i.e. only a few resolution elements) that cannot be uniquely associated with any other cluster in the catalogue, and are therefore retained as independent entities. While these features are not consistent with the definition of *molecular gas cluster* proposed in Colombo et al. (2015) (since they do not have substructures within them), they can correspond to clumps collected in the BGPS sample.

This cloud segmentation contains 36 per cent of the total flux of the survey. This percentage is slightly higher than the flux attributed to GMCs of the full Milky Way catalogue designed by Rice et al. (2016) (25 per cent). We attribute the higher fraction to CO(3–2) tracing higher densities than CO(1–0) and is more likely to be associated with compact objects. This measurement represents just those pixels that are identified with catalogued objects. The extrapolated flux is 94 per cent of the total flux in the survey. Note that the extrapolated flux is not bounded to be less than 100 per cent and our recovery of a fraction near 100 per cent does not mean we are characterizing all the flux in the COHRS data. In heavily blended, bright emission where the SCIMES decomposition segments structures high above the noise level of the data, this extrapolation may effectively overcorrect for the amount of emission and its extent. Most of the flux that is missed by algorithm is the low-brightness emission near catalogued objects.

In our catalogue, 406 objects are ‘exact’ distance clouds, 41 896 are ‘broadcasted’, while 42 718 have a ‘closest’ distance association. Given this, across the analysis we will distinguish between the *full*

sample (all clouds), and a *fiducial sample*, consisting of those objects that have a *broadcast inaccuracy* below 5, i.e. the distance pixel is fewer than 5 pixels away from the cloud surface. The latter criterion should compensate for possible mismatches between COHRS and BGPS astrometries. The fiducial sample consists of 597 clouds for which we have an accurate measurement of their distances. Clouds in the full sample have a median peak SNR ~ 10 , while for the objects in the fiducial sample the median peak SNR ~ 50 .

For the dendrogram construction, we required that a local maximum had to be separated from other local maxima in space by least $3\theta_{\text{FWHM}}$ to be considered independent. Nevertheless, the effective radius of the clouds can be formally smaller than this limit since it is intensity-weighted. The same is true for the velocity dispersion, which is derived following the same philosophy. For the analyses of the paper we consider only objects with extrapolated σ_{maj} , $\sigma_{\text{min}} > \theta_{\text{FWHM}}/\sqrt{8\ln(2)}$ and $\sigma_v > \Delta v/\sqrt{2\pi}$, where for COHRS $\theta_{\text{FWHM}} = 17$ arcsec and $\Delta v = 1$ km s $^{-1}$. This restricts the full and the fiducial samples to 35 446 and 542 well-resolved entries, respectively. For approximately 75 per cent of the excluded objects the extrapolation failed to derive proper semimajor, semiminor, and/or velocity dispersion; since those structures have generally low signal-to-noise (typically peak SNR < 4).

8.1 Ensemble properties

Here we compare the properties of the objects identified in the COHRS data to the cloud catalogue presented in Rathborne et al. (2009) and Roman-Duval et al. (2010). That catalogue was obtained using CLUMPFIND (Williams et al. 1994) on the GRS (Jackson et al. 2006) data. The GRS observed $^{13}\text{CO}(1-0)$ emission over a large part of the first Galactic quadrant: $18^\circ < l < 55.7^\circ$ and $|b| \leq 1^\circ$, comparable to but larger than the COHRS. The spatial resolution the GRS is 46 arcsec and the channel width is 0.21 km s $^{-1}$.

In Fig. 4, we compare spatial/velocity distribution the fiducial and full sample of our catalogue to the GRS catalogue. The distributions of GRS clouds and our fiducial sample appear very similar. Nevertheless, we identify three orders of magnitude more objects in the full catalogue of the COHRS data. This large discrepancy is because Rathborne et al. (2009) smooth the GRS data with a Gaussian kernel of 6 arcmin and 0.6 km s $^{-1}$ meaning that one spatial resolution element in the GRS catalogue contains 440 resolution elements of the COHRS data. The smoothing increases the signal-to-noise ratio of the GRS data but was primarily done to enable the identification of large, GMC-scale objects using the CLUMPFIND algorithm. Since CLUMPFIND typically recovers objects a few resolution elements across (Pineda et al. 2009), it is necessary to suppress the small-scale local maxima with the 6 arcmin smoothing beam. SCIMES, instead, finds naturally clusters of emission across a wide range of scales and provides large complexes comparable to the GRS catalogue without the need of data smoothing.

Clouds in the full sample are uniformly distributed by number along all Galactic longitudes surveyed by COHRS. At large longitudes ($l > 40^\circ$), the number of the fiducial sample clouds drops by 30 per cent due to the fact that BGPS distances are less available there. The GRS catalogue follows a similar trend, but the decrease of sources at increasing longitudes is less prominent. The median latitude of the three samples peak at latitudes slightly lower than $b = 0^\circ$ because of the offset of the Sun above the Galactic plane (Goodman et al. 2014). Since the latitude range of the GRS data is wider than that of COHRS over a wide longitude range, the latitude distribution of extracted sources is larger in the GRS catalogue. In contrast, our fiducial sample has more clouds than the GRS

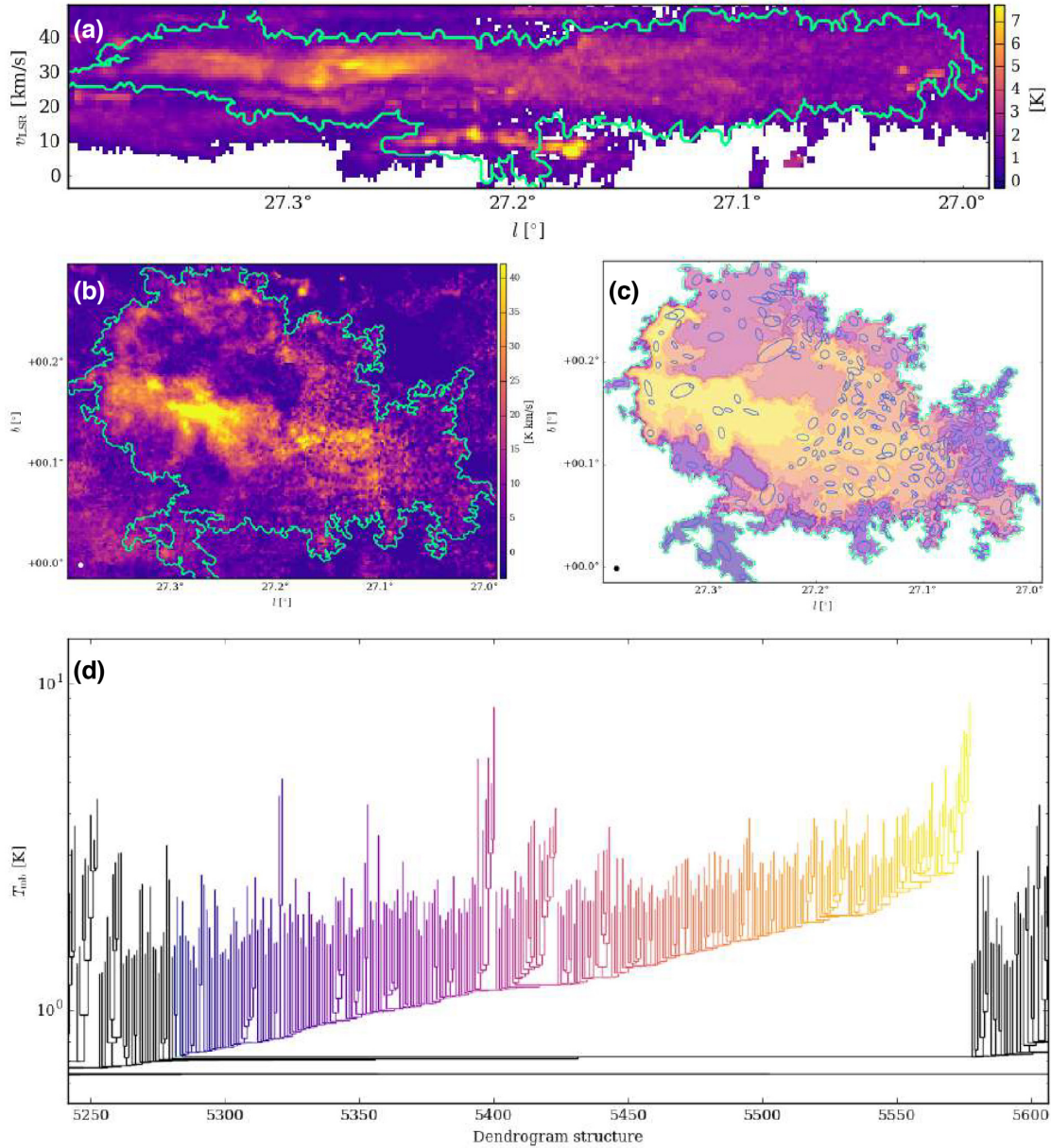


Figure 3. Peak temperature longitude–velocity map (panel *a*) and integrated intensity map (panel *b*) of the cloud ID 3076 from the cohrrs.19_22800_24000 sub-cube. Data are masked as explained in Section 4. Green contours indicate the full extend of the cloud. Cloud ID 3076 hierarchical structure rebuilt through the dendrogram (panel *c*). In the figure every second dendrogram level sub-structure is shown. Blue ellipses indicate the leaves within the cloud. The ellipses parametrize the intensity-weighted major and minor axis of the leaves and their orientation with respect to the longitudinal direction. The section of the dendrogram corresponding to Cloud ID 3076 is highlighted in colour in panel *d*, where each dendrogram structure is colour-encoded as in panel *c*. The circle in the bottom left corner of panels *b* and *c* show the beam of the COHRS data. SCIMES identifies clouds considering and preserving the multiscale and hierarchical nature of the ISM.

catalogue in longitudes between $|b| \leq 0.3^\circ$. The cloud samples we are comparing span similar velocity ranges. The GRS catalogue and our fiducial sub-sample do not contain exactly the same clouds; however, the distribution of sources and number of clouds in our fiducial sample are directly comparable to the GRS. Both the full sample of COHRS clouds and the fiducial sub-sample are peaked around 5 kpc from the Galactic centre.

Fig. 5 shows that the effective radius of our fiducial sample has its median around $R_{\text{eff}} \sim 9$ pc, similar to the GRS clouds which have median sizes $R_{\text{eff}} \sim 7.4$ pc. Nevertheless, our full sample has a median radius of ~ 1 pc reflecting our ability to recover

smaller-sized clouds. In terms of velocity dispersion, the COHRS fiducial sample has a median value of $\sigma_v \sim 2.6 \text{ km s}^{-1}$, while the GRS catalogue is smaller: $\sigma_v \sim 1.3 \text{ km s}^{-1}$. The COHRS object distributions for the fiducial sample are skewed towards larger σ_v values than those of Rathborne et al. (2009), and while this difference could be partially attributed to the relatively coarse spectral resolution of the COHRS data (1 km s^{-1}) which is a factor 5 worse than that of the GRS (0.2 km s^{-1}), the full catalogue for COHRS is still able to recover a median $\sigma_v \sim 1.4 \text{ km s}^{-1}$. The main limitation, may in fact come from the fact that the GRS uses the $^{13}\text{CO}(1-0)$ to observe the molecular gas, which is an optically

Table 1. Contents of the COHRS cloud catalogue. In the following analysis we consider also the clouds on the edge of the data cubes. Removing them does not significantly alter the results. The catalogue includes uncorrected (without suffix, e.g. `radius_pc`), extrapolated ('ex' suffix, e.g. `radius_ex_pc`), deconvolved ('dc' suffix, e.g. `radius_dc_pc`), and extrapolated and deconvolved ('ex_dc' suffix, e.g. `radius_ex_dc_pc`) properties. For properties that depended only on the flux (e.g. L_{CO} , M_{lum} , Σ_{mol}) only uncorrected and extrapolated properties are defined. In the catalogue, uncertainties on the properties are specified with the prefix 'err', e.g. `err_radius_pc`. The electronic version of the catalogue is available online.

Quantity	Unit	Description	Catalogue entry
(1)	(2)	(3)	(4)
ID		Identification number	<code>cloud_id</code>
DS		Structure number within the dendrogram	<code>dendro_structure</code>
File name		Sub-cube assignment file name	<code>orig_file</code>
x_{cen}	pixel	Centroid position along the sub-cube x -axis	<code>xcen_pix</code>
y_{cen}	pixel	Centroid position along the sub-cube y -axis	<code>ycen_pix</code>
v_{cen}	pixel	Centroid position along the sub-cube velocity	<code>vcen_pix</code>
l	degree	Mean Galactic longitude	<code>glon_deg</code>
b	degree	Mean Galactic latitude	<code>glat_deg</code>
v_{LSR}	km s^{-1}	Mean velocity w.r.t. the local standard of rest	<code>vlsr_kms</code>
x_{\odot}	pc	Heliocentric coordinate X	<code>xsun_pc</code>
y_{\odot}	pc	Heliocentric coordinate Y	<code>ysun_pc</code>
z_{\odot}	pc	Heliocentric coordinate Z	<code>zsun_pc</code>
x_{Gal}	pc	Galactocentric coordinate X	<code>xgal_pc</code>
y_{Gal}	pc	Galactocentric coordinate Y	<code>ygal_pc</code>
z_{Gal}	pc	Galactocentric coordinate Z	<code>zgal_pc</code>
σ_{maj}	pixel arcsec $^{-1}$	Major semi-axis size	<code>major_sigma</code>
σ_{min}	pixel arcsec $^{-1}$	Minor semi-axis size	<code>minor_sigma</code>
ϕ	deg	Position angle w.r.t. the cube x -axis	<code>pa_deg</code>
F_{CO}	K	Integrated flux	<code>flux_K</code>
A_{exact}	pixel arcsec $^{-2}$	Area defined as projected total number of pixel	<code>area_exact</code>
A_{ellipse}	pixel arcsec $^{-2}$	Area of the ellipse from $\sigma_{\text{maj,min}}$	<code>area_ellipse</code>
T_{peak}	K	Peak T_{mb} within the cloud	<code>t_peak_K</code>
T_{mean}	K	Mean T_{mb} within the cloud	<code>t_mean_K</code>
SNR_{peak}		Peak signal-to-noise within the cloud	<code>peak_snr</code>
SNR_{mean}		Mean signal-to-noise within the cloud	<code>mean_snr</code>
d	pc	Object distance	<code>distance_pc</code>
Broadcast type		Distance quality (0 = exact, 1 = broadcasted, 2 = closest)	<code>broadcast_type</code>
Broad. inaccuracy	pixel	Broadcast inaccuracy	<code>broadcast_inaccuracy_pix</code>
R_{eff}	pc	Effective radius	<code>radius_pc</code>
σ_v	km s^{-1}	Velocity dispersion	<code>sigv_kms</code>
L_{CO}	$\text{K km s}^{-1} \text{pc}^2$	CO luminosity	<code>lco_kkms_pc2</code>
M_{lum}	M_{\odot}	Mass from the CO luminosity	<code>mlum_msun</code>
M_{vir}	M_{\odot}	Mass from the virial theorem	<code>mvir_msun</code>
σ_0^2	$(\text{km s}^{-1})^2 \text{pc}^{-1}$	Scaling parameter	<code>scalpar_kms2_pc</code>
I_{CO}	K km s^{-1}	Integrated CO luminosity	<code>surf_bright_k_kms</code>
N_{H_2}	cm^{-2}	H_2 column density	<code>col_dens_cm2</code>
Σ_{mol}	$M_{\odot} \text{pc}^{-2}$	Surface density	<code>surf_dens_msun_pc2</code>
ρ_{mol}	$M_{\odot} \text{pc}^{-3}$	Volumetric density	<code>dens_msun_pc3</code>
Volume	$\text{pc}^2 \text{km s}^{-1}$	Volume	<code>volume_pc2_kms</code>
α		Virial parameter	<code>alpha</code>
N_{pix}		Number of pixel within the cloud	<code>npixel</code>
N_{leaves}		Number of leaves within the cloud	<code>nleaves</code>
Edge		The cloud is on the cube lower or upper border	<code>edge</code>
Parent		Cloud parental structure ID	<code>parent</code>
Ancestor		Cloud parental structure ID at the bottom of the hierarchy	<code>ancestor</code>
Struct. type		Structure type (T = trunk, B = branch, L = leaf)	<code>structure_type</code>
Spiral arm		Spiral arm associated with the cloud: (Sa = Sagittarius, Sc = Scutum, Lo = Local, Pe = Perseus, No = Norma)	<code>assoc_sparm</code>
Dist. to arm	pc	Distance to the associated spiral arm	<code>dist_to_sparm</code>

thinner tracer than the $^{12}\text{CO}(3-2)$ used in our study. In practice, this means that the $^{13}\text{CO}(1-0)$ emission is able to trace higher density regions of the molecular clouds, that are not traceable with $^{12}\text{CO}(3-2)$. As a result, the linewidths for the clouds measured from ^{12}CO will be naturally larger than those of the GRS, particularly

in clouds that contain high-density regions within them. Hence this optical depth effect affects more the fiducial sample, which naturally contains the most massive/dense star forming regions since they have associated compact continuum emission as detected with the BGPS.

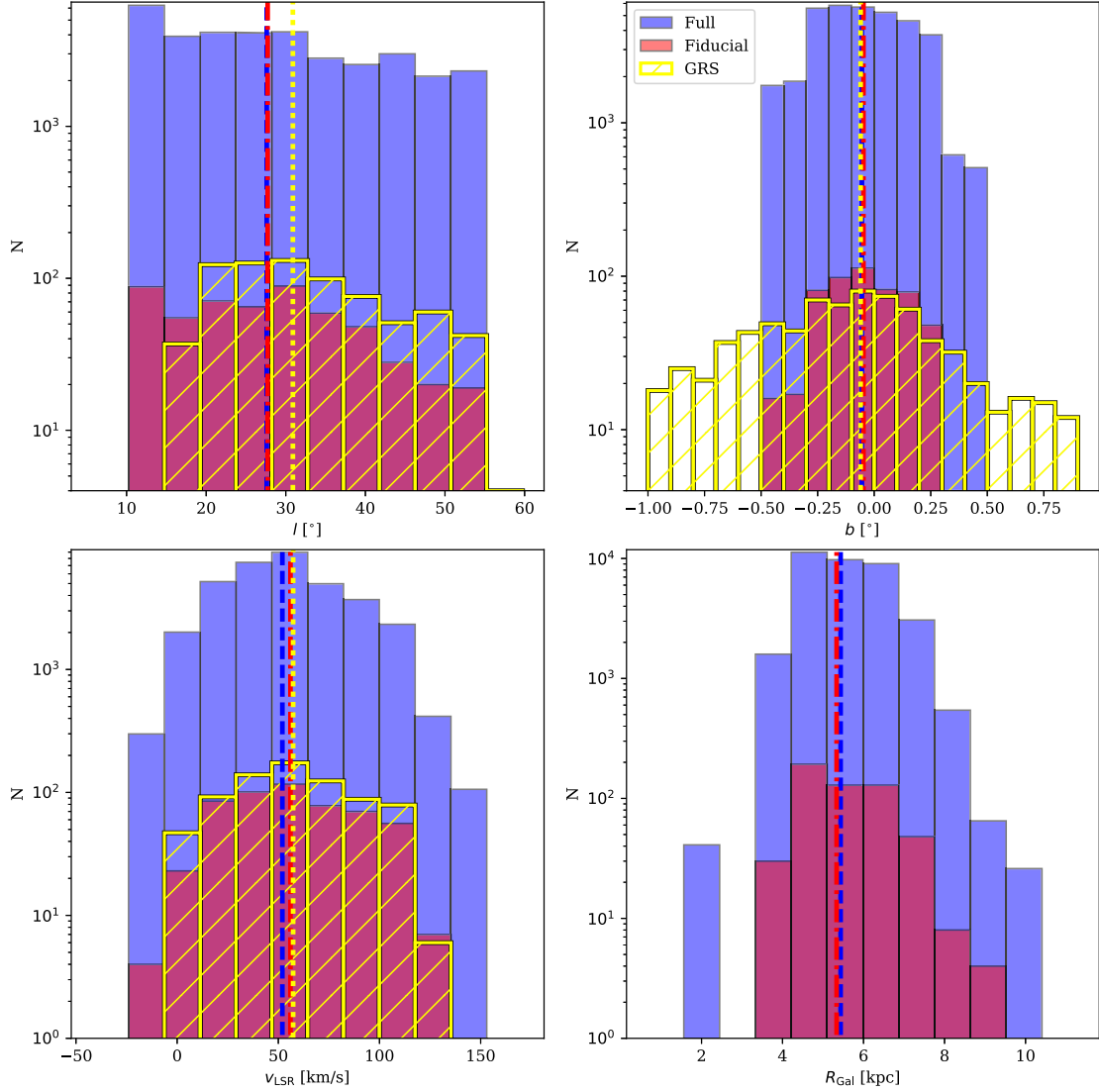


Figure 4. Distribution of the COHRS clouds with respect to Galactic longitude (l), Galactic latitude (b), systemic velocity (v_{LSR}), and Galactocentric distance (R_{Gal}) for the fiducial sample (red), the full sample (blue), and GRS catalogue from Roman-Duval et al. (2010) (yellow). The vertical red dash-dotted lines indicate the median of the fiducial sample, the blue dashed lines the median of the full sample, and yellow dotted lines the median of GRS property distributions. The COHRS fiducial catalogue is comparable in scope to the GRS catalogue and the differences between these distributions can be attributed to differences in the survey coverages.

The larger-than-expected linewidths will constitute one of the main biases of our study, and even with the deconvolution shown in Section 7.4 we are not able to compensate for this effect. This will potentially influence our measured virial masses and virial parameters. The average virial mass and virial parameters of the COHRS fiducial sample ($\sim 6.4 \times 10^5 M_{\odot}$ and $\alpha \sim 2$) are a factor 4 larger than the respective GRS median values ($\sim 1.4 \times 10^5 M_{\odot}$ and $\alpha \sim 0.5$).

The average masses derived from CO luminosity and surface densities are fairly consistent between the two catalogues: $\langle M_{\text{lum}} \rangle \sim 3.2 \times 10^5 M_{\odot}$ (COHRS fiducial) versus $\langle M_{\text{lum}} \rangle \sim 2.5 \times 10^5 M_{\odot}$ (GRS). Nevertheless, the fiducial sub-sample contains clouds more massive than GRS. Similarly, the average cloud mass surface densities between the GRS catalogue and the COHRS fiducial sample are approximately the same ($\Sigma_{\text{mol}} \sim 130 M_{\odot} \text{ pc}^{-2}$ and $\Sigma_{\text{mol}} \sim 140 M_{\odot} \text{ pc}^{-2}$, respectively), but we have clouds with larger surface densities in our sample.

8.2 Cumulative distributions of cloud masses and sizes

Fitting cloud mass and size distributions can provide basic information about the cloud population and the organization of the molecular ISM itself. In this section we use the cumulative distributions of cloud M_{lum} and R_{eff} that we model as a truncated power-law distribution (Williams & McKee 1997):

$$N(X > X') = N_0 \left[\left(\frac{X}{X_0} \right)^{\gamma+1} - 1 \right]. \quad (15)$$

Here, X_0 will correspond to M_0 for the mass distribution and to R_0 for the effective radius distribution, and represents the maximum value of the distribution, while N_0 is the number of clouds greater than $2^{1/(\gamma+1)} X_0$, i.e. where the distribution deviates from a power law. If $N_0 \gg 1$ there is strong evidence for a truncation in the power law which indicates that physical effects are at work to limit the maximum value of a given cloud property. The truncation in

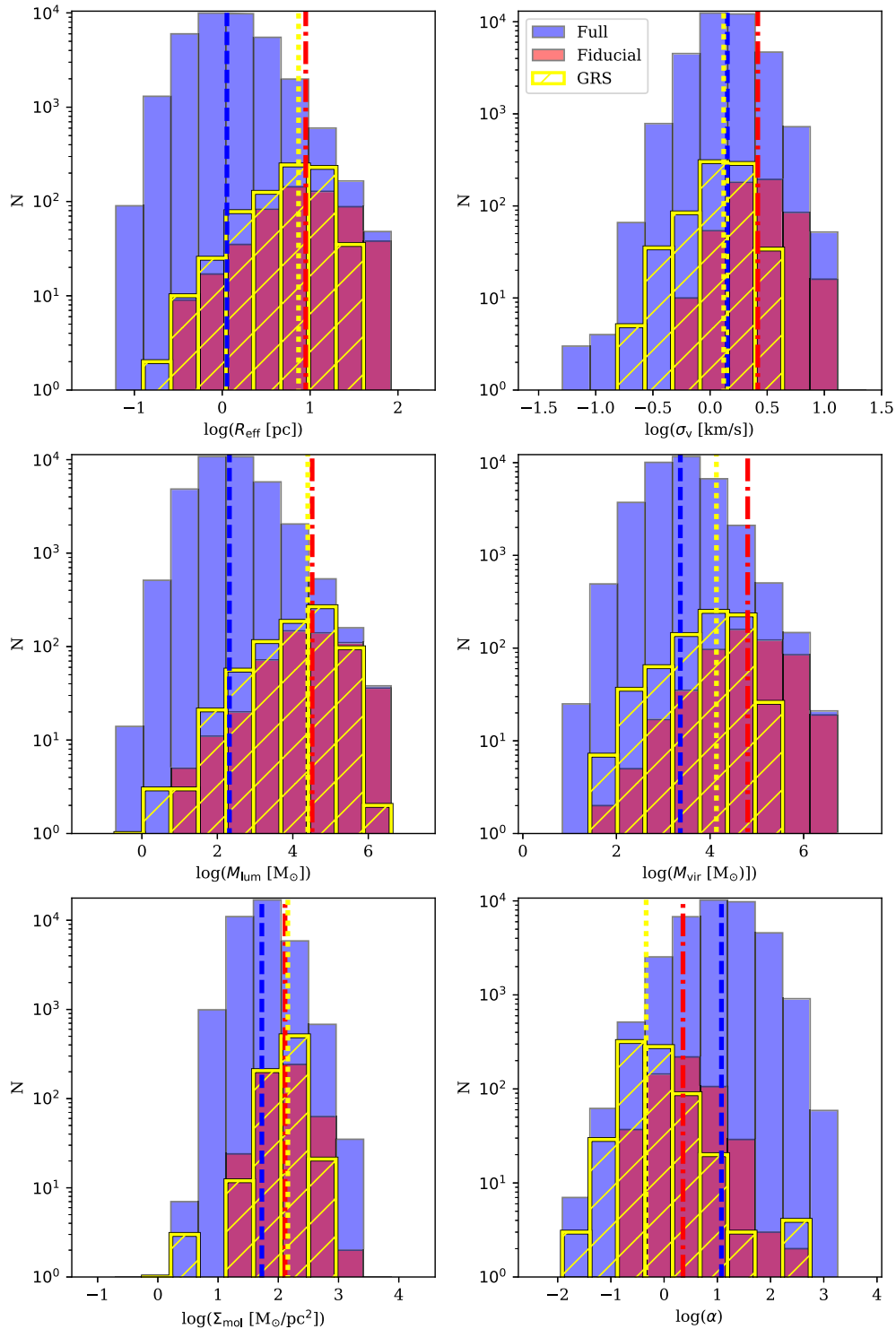


Figure 5. COHRS object ensemble properties (from left to right, top to bottom): effective radius (R_{eff}), velocity dispersion (σ_v), mass from CO luminosity (M_{lum}), virial mass (M_{vir}), molecular gas mass surface density (Σ_{mol}), virial parameter (α). Symbol conventions follow Fig. 4. The COHRS fiducial sample is comparable to the GRS catalogue but the coarse velocity resolution (1 km s^{-1}) of the COHRS data might strongly affect our results.

the cloud mass distribution has been observed in several occasions (Williams & McKee 1997; Rosolowsky 2005; Freeman et al. 2017; Jeffreson & Kruijssen 2018). The ‘cumulative’ form allows to fit the distributions considering the uncertainties on the cloud properties, and it is not influenced by the choice of the bin size that can bias binned distributions (Rosolowsky 2005).

Cumulative mass spectra are only well defined above the completeness limit of the survey. To estimate the mass completeness limit we consider the procedure illustrated in Heyer, Carpenter & Snell (2001, equations 2–4). For their outer Galaxy cloud catalogue

the authors suggest a minimum CO luminosity given by

$$L_{\text{CO}}^{\text{min}}(d) [\text{K km s}^{-1} \text{ pc}^2] = N_p N_c T_{\text{th}} \Delta v \Omega_{\text{mb}} d^2; \quad (16)$$

where N_p is the minimum number of pixel per object, N_c the minimum number of velocity channels, T_{th} the main-beam antenna temperature threshold, Δv the data set channel width, Ω_{mb} beam solid angle, and d the distance to the cloud. As explained by Heyer et al. (2001), the completeness limit is evaluated at 5σ confidence limit:

$$L_{\text{CO}}^c = L_{\text{CO}}^{\text{min}} + 5\sigma(L_{\text{CO}}); \quad (17)$$

where

$$\sigma_{L_{\text{CO}}} [\text{K km s}^{-1} \text{ pc}^2] = \sigma_{\text{RMS}} \sqrt{N_p N_c} \Delta v \Omega_{\text{mb}} d^2; \quad (18)$$

and σ_{RMS} is the median RMS noise value across the full survey.

In our case we assume that the minimal object contains $N_p = 18$ pixels (3 beams \times 6 pixels per beam), $N_c = 2$ channels (required by the dendrogram generation), minimum brightness $T_{\text{th}} = \text{SNR} \sigma_{\text{RMS}}$ (where we assume $\sigma_{\text{RMS}} = 1 \text{ K}$ as an conservative value across the COHRS fields, see Dempsey et al. 2013; and $\text{SNR} = 3$ as imposed by our masking method), channel width $\Delta v = 1 \text{ km s}^{-1}$ (COHRS data cube channel width), beam size of $\theta_{\text{FWHM}} = 17 \text{ arcsec}$ (COHRS data beam). In this way, we calculate a luminosity mass completeness of $\sim 2 \times 10^3 M_{\odot}$ by assuming our $\alpha_{\text{CO}} = 8.7 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$.

For the cloud size, considering the COHRS beam of $\theta_{\text{FWHM}} = 17 \text{ arcsec}$ and that a cloud must span at least three beams to be regarded as an independent structure in the dendrogram, we get an effective radius completeness of 3 pc.

These estimated completeness limits are conservative estimates for M_{lum} and R_{eff} at 15 kpc, the largest distance in our catalogue. SCIMES does not extract objects at a fixed T_{mb} threshold. Instead, the masking level is set by the local noise properties and the $\text{SNR} = 3$ threshold from the masking and dendrogram generation parameters. Moreover, we use extrapolation and deconvolution which renders the measurement of the radius distribution more complex. Thus, it is possible to find several objects at 15 kpc with masses and effective radii below $\sim 2 \times 10^3 M_{\odot}$ and 3 pc, respectively.

We fit equation (15) to our spectra above these completeness limits using orthogonal distance regression (ODR) as implemented in SCIPY,⁴ which takes into account the uncertainties on both dependent and independent quantities. Fig. 6 shows the result of this experiment for both mass (top) and size (bottom) distributions.

For the mass distribution of the full catalogue, we find a power-law slope of $\gamma = -1.76 \pm 0.01$, $N_0 = 20.2 \pm 0.1$, $M_0 = (3.14 \pm 0.03) \times 10^6 M_{\odot}$. The truncation indicates that clouds with mass above $M_0 \sim 3 \times 10^6 M_{\odot}$ are significantly absent in the region of the Milky Way surveyed by the COHRS. A truncation around $10^6 M_{\odot}$ has been observed for the first Galactic quadrant by other studies which use the cumulative representation of the mass spectrum (e.g. Rosolowsky 2005; Rice et al. 2016). Such a truncation mass is a critical feature for testing GMC evolution theories in the context of galaxy environment. Reina-Campos & Kruijssen (2017) developed a model for the maximum mass scale of GMCs in galaxies as a function of local environment, finding a near constant upper limit mass of $\sim 10^6 M_{\odot}$ over the Galactocentric radii of 4–8 kpc. The physical effects that govern this mass scale are gravitational collapse, on scales allowed by the Toomre stability criterion, and stellar feedback. Our mass distributions,

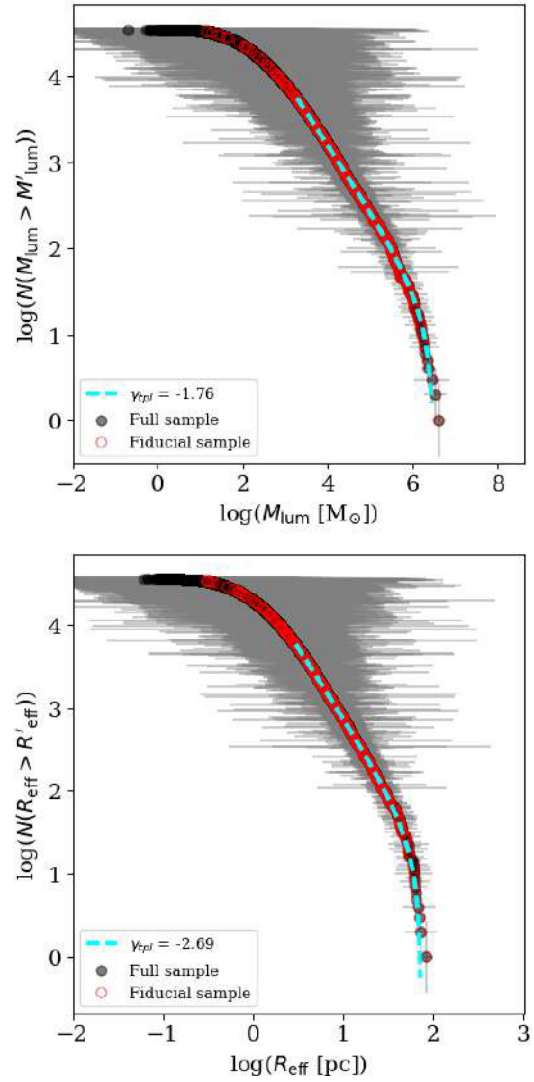


Figure 6. COHRS mass (top) and size (bottom) spectra from the M_{lum} and R_{eff} of all the clouds in the catalogue (black circle). The fiducial sample is indicated with red empty circles. Cyan dashed lines indicate the truncated power-law fit of the spectra above the estimated completeness limits. Grey lines indicate the error bars. For the y-axis the uncertainty is given by the counting error: \sqrt{N} .

when separated into bins of Galactocentric radius show a nearly constant truncation mass at all radii, consistent with those models.

A spectral index $\gamma \sim -1.76$ indicates that most of the molecular gas mass is contained in large objects. Our measurement is largely consistent to the $\gamma \sim -1.7$ observed for the inner Milky Way (e.g. Roman-Duval et al. 2010; Heyer & Dame 2015; Rice et al. 2016). Our slightly steeper index can arise from the significantly higher resolution of the survey compared to the data supporting previous catalogues. With coarse resolution, small clouds will be blended with large clouds, suppressing the number of small clouds recovered and increasing the mass of the larger clouds, thereby biasing the index to shallower values.

For the size distribution we get $\gamma = -2.70 \pm 0.01$, $N_0 = 26.4 \pm 0.2$, $R_0 = 73.6 \pm 0.2 \text{ pc}$. The latter two quantities indicate that the R_{eff} distribution shows a truncation as well: the size of the clouds also reaches a maximum value in the inner Galaxy. In contrast to the mass spectrum, our size distribution appears significantly

⁴<https://docs.scipy.org/doc/scipy/reference/odr.html>

shallower than the one observed in the same region through ^{13}CO observation ($\gamma = -3.9$; Roman-Duval et al. 2010). This difference can be attributed to the action of the SCIMES algorithm, which can recover objects significantly larger than the resolution element. The GRS catalogue is largely established by the size of the smoothing element, leading to a sharp cutoff in object sizes larger than 8 pc (i.e. the 6 arcmin smoothing kernel projected to 5 kpc). This is supported by Fig. 5, which shows that our COHRS catalogue recovers a tail of larger objects than those in the GRS catalogue.

9 COHRS CLOUDS IN THE CONTEXT OF THE MILKY WAY

In the previous section, we have looked at the global distribution of cloud properties using the COHRS data set. In this section, we shall investigate if these properties change as a function of Galactic environment.

9.1 Comparison between Galactic centre, inner Galaxy, and outer Galaxy clouds

In order to have a first glance at how the Galactic environment might be affecting cloud properties, we compare our clouds to those seen in other Galactic regions, where the environment is potentially different from the inner Galaxy. In particular we analyse our data with respect to the catalogues built for the outer Galaxy (Heyer et al. 2001) and the Galactic centre (Oka et al. 2001), which have been constructed starting from ^{12}CO observations⁵ that are similar in term of spatial and spectral resolution to COHRS (however bias effects introduced by the segmentation methods can apply, see Section 10.2).

In Fig. 7 we compare the clouds in the three surveys using two distance-independent properties: CO integrated intensity (I_{CO} , upper panel) and velocity dispersion (σ_v , lower panel). The cloud data are colour-encoded by their respective properties and plotted on the face-on view of the Milky Way. In terms of I_{CO} , the clouds in the three Galactic regions are starkly different: objects in the outer Galaxy reach maximum integrated intensities around $5 \times 10^3 \text{ K km s}^{-1}$, while clouds in the Galactic centre have $3 \times 10^3 \text{ K km s}^{-1} < I_{\text{CO}} < 6 \times 10^6 \text{ K km s}^{-1}$. COHRS objects show values of CO integrated intensity in between these two extremes. Similar conclusions can be drawn from the velocity dispersion comparison, even if the difference is less sharp. In Fig. 8 we show the data points colour-encoded by their mass from CO luminosity. In this case, for COHRS we plot only the clouds from the fiducial sample, for which we have good estimations of their distances. This experiment highlights the fact that the fiducial sample is mostly constituted by massive objects, and their masses appear to be similar to the clouds identified in the Galactic centre.

While drawing the clouds on a face-on view of the Milky Way is useful to visualize their locations across the Galactic disc, the large overlap between the data points does not allow to derive firm conclusions about the radial gradient of the cloud properties. Therefore, in Fig. 9 we display violin plots of CO integrated intensity, velocity dispersion, and luminosity mass within bins of 2 kpc Galactocentric radius for the cloud within the three surveys.

⁵Note, however, that Heyer et al. (2001) and Oka et al. (2001) uses $^{12}\text{CO}(1-0)$ which has different sensitivity with respect to $^{12}\text{CO}(3-2)$ ($\sim 1.4 \times 10^3 \text{ cm}^{-3}$ versus $\sim 7 \times 10^4 \text{ cm}^{-3}$), respectively. Therefore, $^{12}\text{CO}(1-0)$ would potentially trace more extended structures with respect to $^{12}\text{CO}(3-2)$, and give broader linewidths.

This representation confirms that, on average, clouds segmented out from COHRS data have properties intermediate between Galactic centre and outer Galaxy objects. Nevertheless, intrinsic biases of the different segmentation methods might play a role here (see Section 10.2). For instance, outer Galaxy cloud masses in the most external bin reach estimates closer to the ones of COHRS objects, but the monotonic increment of the average cloud mass in the outer Galaxy suggests that the extracted properties are scaling with the distance, as would be the case when the segmentation method extracts clouds around a specific angular scale, rather than to the actual physics involved in the different Galactic regions.

9.2 Galactocentric dependence of properties in the COHRS sample

Given that interpreting the environmental dependencies of cloud properties with the intercomparison of different surveys is not straightforward, we explore if we can detect any trends within the COHRS sample alone. To do so, we have divided the full sample into four bins according to the Galactocentric distance of the clouds, each containing an equal number of clouds, and estimated their individual cumulative distributions of cloud masses and radius (see Fig. 11, top row). All bins show mass spectrum slopes consistent with the full sample one ($\gamma \sim -1.75$). This indicates that most of the molecular gas is contained in massive clouds, for all radial bins. Nevertheless, the innermost annulus (where $1.6 \text{ kpc} < R_{\text{Gal}} < 4.8 \text{ kpc}$) exhibits a distribution shifted towards higher masses than the other bins. This annulus contains the most massive clouds of the sample, corresponding to rich reservoir of molecular gas in the Galactic ring. Furthermore, the dynamics likely favour the formation of large cloud complexes at the end of the Milky Way's stellar bar. The two outer annuli we consider ($5.5 \text{ kpc} < R_{\text{Gal}} < 6.3 \text{ kpc}$ and $6.3 \text{ kpc} < R_{\text{Gal}} < 10.4 \text{ kpc}$) show similar mass distributions. However, the region between $4.8 \text{ kpc} < R_{\text{Gal}} < 5.5 \text{ kpc}$ appears to contain less massive clouds compared to other Galactocentric annuli.

The radial variations in the size distributions constructed within the same annuli do not always reflect the radial changes in the mass spectra. The innermost ring and the two outermost annuli are almost indistinguishable. However, the spectrum of the ring between $4.8 \text{ kpc} < R_{\text{Gal}} < 5.5 \text{ kpc}$ appears bended towards lower effective radii, reflecting the behaviour of the corresponding mass distribution.

From this analysis, it appears that various environmental effects are at work in the surveyed Milky Way region that contribute to create mass and size distributions with different shapes. Zetterlund et al. (2018) find a significant steepening of the mass distribution of dense gas clumps in the range $4.6 \text{ kpc} < d_{\text{Gal}} < 6.3 \text{ kpc}$. We note that variations in the mass spectra remain when examined in a distance-limited sample, and are not likely to be attributable to distance scalings alone.

9.3 Spiral arm versus inter-arm clouds in the COHRS sample

From Figs 7 to 8 it appears that a significant number of clouds in the COHRS sample might be associated with the inter-arm regions of the Milky Way. It is interesting now to verify this rudimentary visual impression with a more rigorous test. For this experiment and to draw the arms in Figs 7 and 8 we use the spiral arm models defined by the lognormal spiral:

$$\log(R/R_{\text{ref}}) = -(\beta - \beta_{\text{ref}}) \tan(\psi), \quad (19)$$

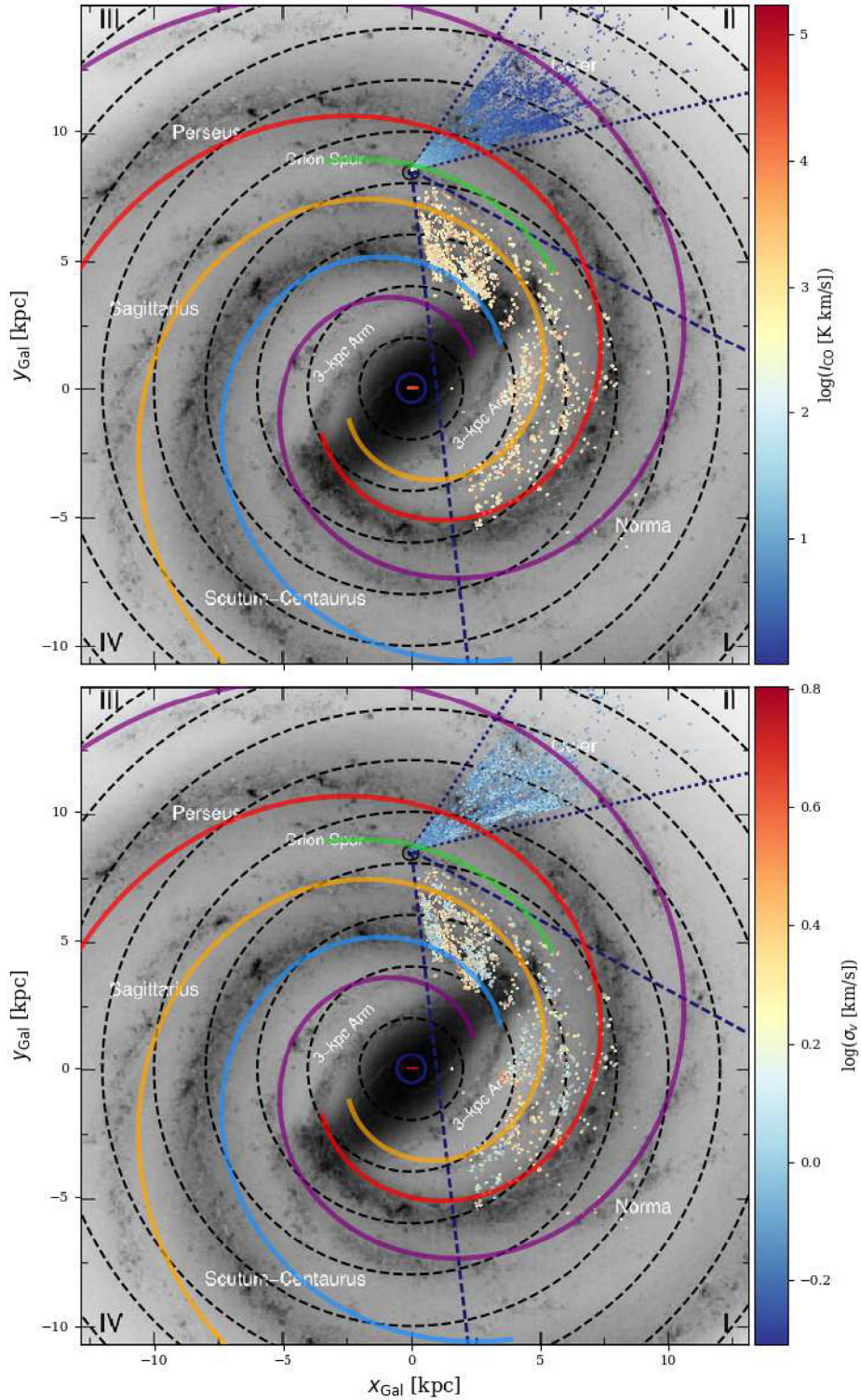


Figure 7. COHRS, outer Galaxy, and Galactic centre cloud positions with respect to the Milky Way artistic impression [courtesy of NASA/JPL-Caltech/R. Hurt (SSC/Caltech)]. Markers are colour-encoded with a given properties: CO integrated intensity (top) and velocity dispersion (bottom). Those properties are independent from distance. The dark blue circle, dashed, and dotted lines indicate, approximately, the Milky Way regions observed by Oka et al. (2001), Dempsey et al. (2013), and Heyer et al. (2001) surveys, respectively. Note that the data present the Oka et al. (2001) catalogue of the Galactic centre are assumed to all be at a common y_{Gal} position in the Galactic centre. Galactocentric dashed black circles are placed 2 kpc apart. The spiral arm positions are drawn from the results of (Vallée 2017) (see Section 9.3 for further details) Scutum (blue), Sagittarius (yellow), Perseus (red), Local (green), and Norma/outer (purple).

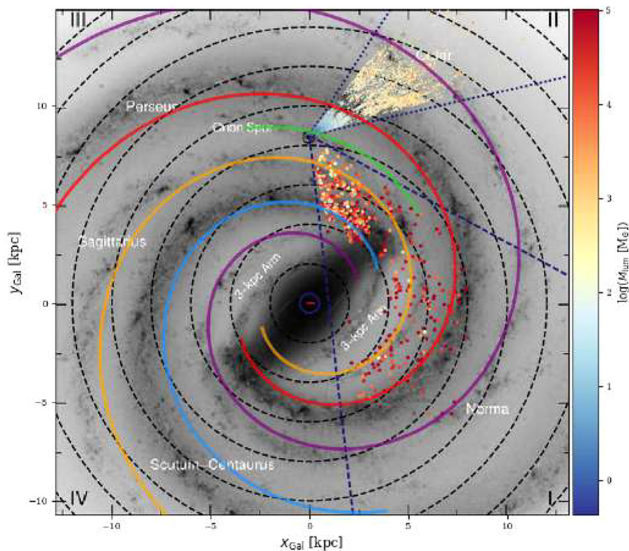


Figure 8. COHRS, outer Galaxy, and Galactic centre cloud positions with respect to the Milky Way artistic impression [courtesy of NASA/JPL-Caltech/R. Hurt (SSC/Caltech)]. Markers are colour-encoded with the cloud mass from CO luminosity. For the COHRS data only the fiducial sub-sample is indicated. Other symbol notations follow Fig. 7.

where the reference Galactocentric radius (R_{ref}), Galactocentric azimuth (β_{ref}), and pitch angle (ψ) are taken from the recent update of Vallée (2017). In the model of Vallée (2017), the pitch angle of each arm is kept constant ($\psi = 13^\circ$). We assume that the four main Milky Way arms originate from the tip of the long bar which have a semimajor axis of 5 kpc (Wegg, Gerhard & Portail 2015), therefore $R_{\text{ref}} = 5$ kpc for each arm. The values of β_{ref} are assumed to be 0° , 90° , 180° , 270° for Scutum, Sagittarius, Perseus, Norma arms, respectively. Given the average inclination of the bar with respect to the line that connects Sun and Galactic centre, $a = 30^\circ$ (Wegg et al. 2015), the starting Galactocentric azimuth of Scutum and Norma arms are fixed to $\beta = 60^\circ$, while for Sagittarius and Perseus arms $\beta = 240^\circ$. The parameters of the Local spurs have been precisely defined by the VLBI maser parallaxes measurements of Reid et al. (2014) (see their table 2): $R_{\text{ref}} = 8.4$ kpc, $\beta_{\text{ref}} = 8.9^\circ$, $\psi = 12.8^\circ$.

To quantify the number of clouds associated with a spiral arm we calculate the Euclidean distance between the cloud centroid position in Galactocentric coordinates (x_{Gal} and y_{Gal} in the catalogue) and the closest point of each spiral arm ridge line described by the model.

The result of the analysis is reported in Fig. 10. Most of the objects appear almost equally distributed between Sagittarius and Scutum arms, followed by Perseus, Local, and Norma arms. Similarly to the visual impression obtained by looking at Figs 7 and 8, a small fraction of the clouds (~ 35 per cent) are found within the spiral arms, if we consider an average arm width of 600 pc (Vallée 2017). By assuming a larger average width, 800 pc as calculated by the same author in an earlier paper (Vallée 2014), almost 50 per cent of the clouds appear to be enclosed within spiral arms. Clouds within the spiral arms encompass almost 50 per cent of the cloud flux in the catalogue (considering also the small objects excluded from most of the analysis in the paper). These fractions are very similar to the findings in nearby galaxies (for M51 only ~ 60 per cent of the flux comes from spiral arms; Colombo et al. 2014).

Using these sub-samples, we explored whether there are any noticeable differences between the cloud properties in either sub-sample. The cumulative distributions of the mass and radius are

shown in Fig. 11 (bottom row). From there, we can see that the distributions look overall very similar, independently of the arm/interarm assignment, particularly when looking at the slope of the distributions. A two-sided KS test on the full sample has low p -values, which would suggest that the two are not drawn from the same distribution. However, the KS test on the fiducial sample shows that the distributions are very similar. Interestingly, despite the lower numbers of clouds in the spiral arm regions, we find that both distributions reach similar maximum sizes and masses.

We note, however, that while it is interesting to see that we do not recover a strikingly different cumulative mass or size distribution in these two different environments (as opposed to what has been observed in, e.g. M51; Colombo et al. 2014), there are a couple of caveats to this study that we should bear in mind.

First, as noted in Section 8.1, optical depth effects could play a key role, particularly in our ability to trace the higher density regions within molecular clouds, which could potentially underestimate the masses of our clouds, especially for the most massive complexes where more of the mass is enclosed in high-density regions. This effect could potentially alter the shape of the tails of the distribution, which is where we would expect to see a different behaviour between the two distributions (e.g. Koda et al. 2009; Duarte-Cabral & Dobbs 2016).

Secondly, the specific assignment of each cloud into arms or interarms is highly uncertain, partly due to our limited knowledge of the position and extent of each arm in the Milky Way. Indeed, Reid et al. (2014) calculated a variable width for the spiral arms of the Milky Way: 170 pc for Scutum, 260 pc for Sagittarius, 380 pc for Perseus, 630 pc for Norma, and 330 pc for the Local arms (see their table 2). Using this width the number of clouds and the flux within the spiral arms appear strongly reduced: only 12 per cent of the clouds are found within the spiral arms, carrying a similar percentage of fluxes with respect to the total catalogued cloud fluxes. Nevertheless, the derivation of Reid et al. (2014) are valid for the second Galactic quadrant. Moreover, the authors noticed that the arm width tends to increase with Galactocentric radius. Therefore, the spiral arm widths might be different than the assumed ones in the region surveyed by COHRS. But even if the model derived by Vallée (2017) appears to be the more probable given the variety of tracers analysed and summarized by the author, our position within the Galactic plane makes difficult to define the spiral arm clouds with good precision. For instance, we have assumed that a cloud whose centroid falls within 300 pc from the assigned spiral arm is fully contained within it, but the clouds' extension and morphology is not taken into account. Some of the clouds might straddle the arm and inter-arm regions (as in the case of feather or spurs observed in nearby grand-designed spirals, e.g. Schinnerer et al. 2017), and this is not accounted for.

Lastly, the kinematic distances used to place the COHRS clouds in a face-on view of the Galaxy, are affected by large uncertainties, especially when considering clouds that are in or close to a spiral arm, where the velocities deviate from the circular motions assumed for the gas when using a model of the rotation curve (e.g. Ramón-Fox & Bonnell 2018; Duarte-Cabral et al. 2015). This effect can place clouds in an arm while they should be inter-arm clouds, and vice versa, making this exercise non-trivial.

10 SCALING RELATIONS BETWEEN CLOUD PROPERTIES

So far we have only analysed each cloud property on its own. However, the correlations between different properties have been

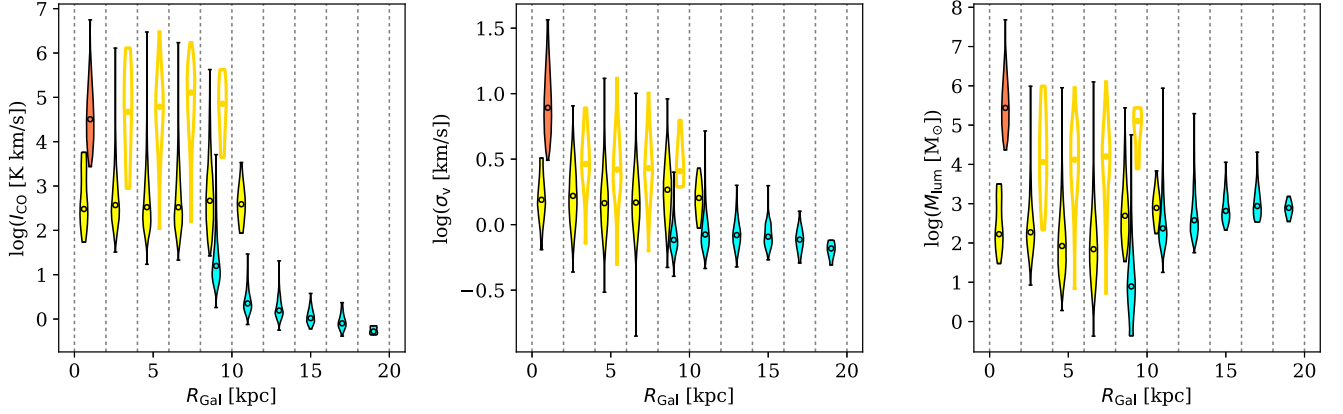


Figure 9. Violin plot representations of CO integrated intensity (left), velocity dispersion (middle), and luminosity mass (right) for the clouds in the outer Galaxy (cyan, Heyer et al. 2001 catalogue), Galactic Centre (red, Oka et al. 2001 catalogue), COHRS full catalogue (yellow full violins), and fiducial catalogue (yellow empty violins). Violin plots are histograms, where the width along the x -axis indicates the normalized fraction of data at the corresponding y -axis value. The violins extend from the minimum to the maximum of the distributions. Circles indicate the median of the property distributions within each Galactocentric radial bins of 2 kpc.

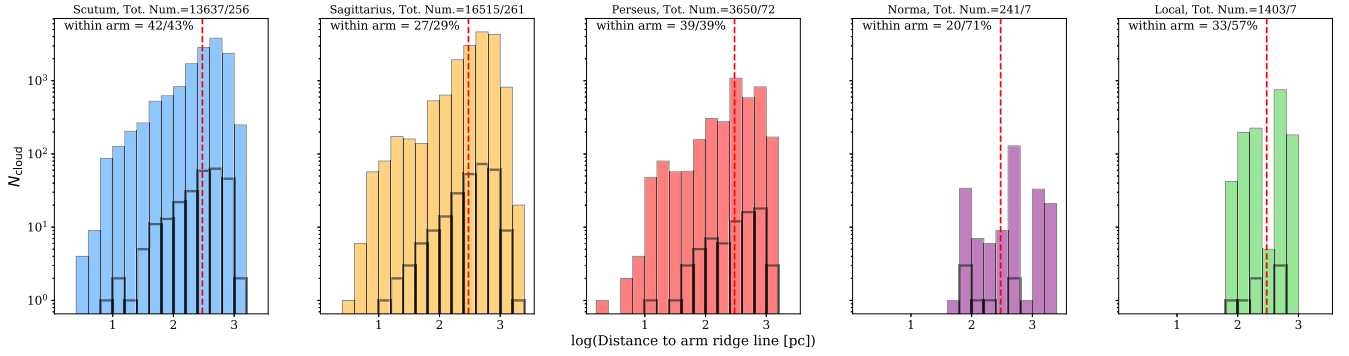


Figure 10. Histograms of the number of clouds from the ‘full’ sample associated with (from left to right) Scutum, Sagittarius, Perseus, Norma/outer, and Local arms described as lognormal spirals as drawn in Figs 7 and 8. Black transparent histograms highlight the clouds in the ‘fiducial’ sample. The total number of full and fiducial sample clouds associated with a specific arm is indicated in the title of each panel. The percentage of full and fiducial sample clouds located within the spiral arms is given in the upper left side of each panel. The vertical, red, dashed lines mark the average spiral arm half width (300 pc; Vallée 2017).

the primary channel for our understanding of the dynamical state and evolution of the molecular ISM. The seminal work in this area is from Larson (1981) who identified three fundamental scaling relations between the molecular cloud properties, often called ‘Larson’s laws’. Larson (1981) measured a correlation between velocity dispersion and size among molecular objects from sub-parsec to few hundreds parsecs size (the size–linewidth relation) as $\sigma_v = 1.10L^{0.38}$. The author interpreted this relation as the manifestation of Kolomogorov’s incompressible turbulence, which was proposed to be the main agent for creating molecular overdensities. The second relation measured by Larson (1981) was between velocity dispersion and mass ($\sigma_v = 0.42M^{0.20}$) which implies that the clouds are in approximate virial equilibrium. Consistent with this hypothesis, Larson (1981) demonstrated there was no discernible relationship between the virial parameter, α , and cloud size. An anticorrelation between cloud mean density and size ($n = 3400L^{-1.10}$) was consistent with clouds having a constant surface density. Solomon et al. (1987) then used a homogeneous cloud sample in the inner Galaxy to remeasure the Larson’s relations, and found a constant mass surface density $\Sigma_{\text{mol}} = 170 M_{\odot} \text{pc}^{-2}$.

However, these scalings are not independent (Heyer et al. 2009; Wong et al. 2011). The second Larson’s relation implies that the clouds are in virial equilibrium; in terms of the virial parameter, $\alpha \sim$

1, or $M_{\text{lum}} \sim 5\sigma_v^2 R_{\text{eff}}/G$. Together with the definition of cloud mass surface density, $\Sigma_{\text{mol}} = M_{\text{lum}}/(\pi R_{\text{eff}}^2)$, the cloud velocity dispersion can be expressed as:

$$\sigma_v = \sqrt{\frac{\pi}{5} G \Sigma_{\text{mol}} R_{\text{eff}}}. \quad (20)$$

For a $\Sigma_{\text{mol}} \sim 200 M_{\odot} \text{pc}^{-2}$ which is typical for the inner Galaxy (e.g. Heyer & Dame 2015), we find a formulation of the first Larson’s relation:

$$\sigma_v = 0.74 R_{\text{eff}}^{0.50}, \quad (21)$$

similar to the one observed by Solomon et al. (1987). In this aspect the scaling between size and linewidth of the cloud

$$\sigma_0 \equiv \sigma_v \left(\frac{R_{\text{eff}}}{1 \text{ pc}} \right)^{-1/2} \quad (22)$$

should be constant. Heyer et al. (2009) reanalysed Solomon et al. (1987) clouds using ^{13}CO data drawn from the GRS to independently derive the cloud mass through an LTE analysis and found a scaling between σ_0 and the molecular gas mass surface density of the clouds. This implies that the clouds in the first Galactic quadrant do not follow the original Larson relations: they cannot be defined

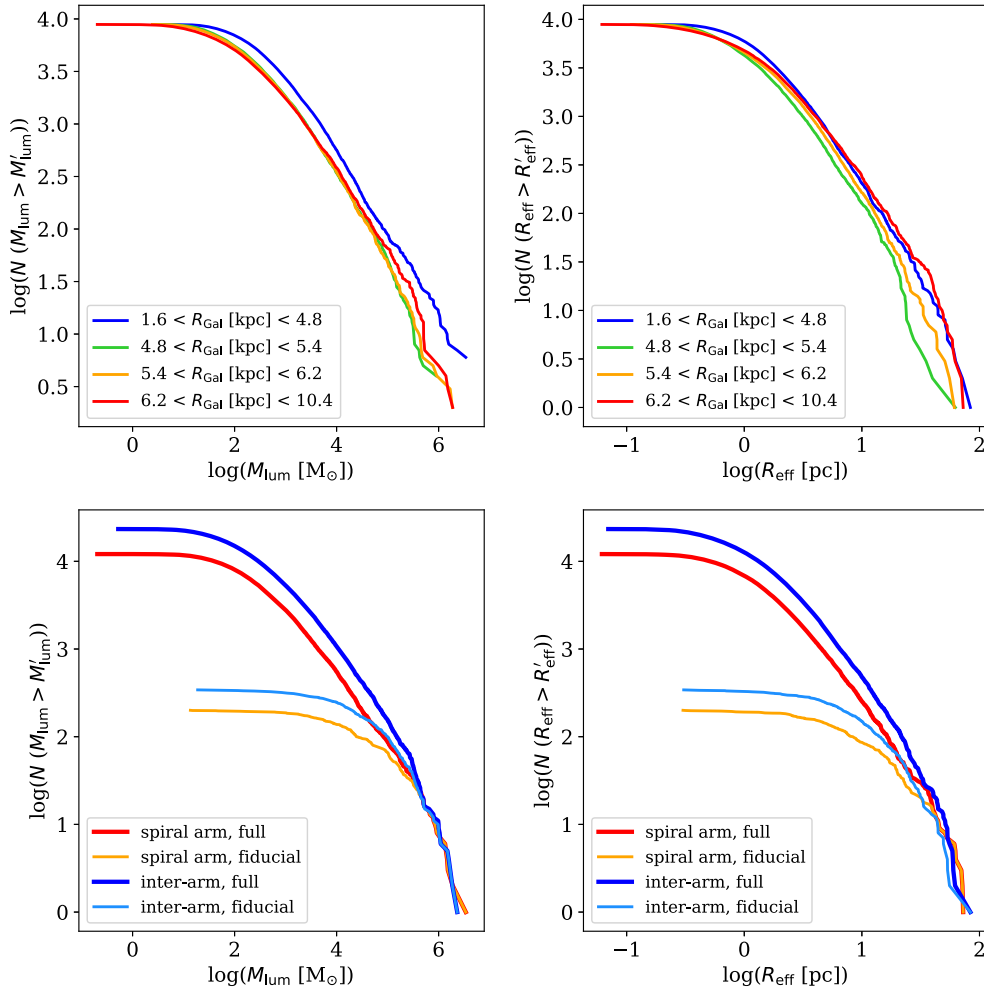


Figure 11. Cumulative mass (left) and size (right) distributions for the clouds in COHRS as function of the environment. *Top*: coloured spectra are calculated for the full sample within various Galactocentric radii that contain the same number of clouds. *Bottom*: distributions of cloud mass and size within the spiral arms and in the inter-arm regions for the full and fiducial samples separately.

by a single scaling between σ_v and R_{eff} , as they do not have constant Σ_{mol} , and they are not necessarily gravitationally bound.

In this Section we analyse the correlations between cloud integrated properties using the principal component analysis (PCA; Pearson 1901) technique applied to the bivariate relationships between cloud properties (see Appendix C for more details).

10.1 Scaling relations from the COHRS data set

We investigate the various scaling relations within the COHRS sample. Table 2 summarizes the results from PCA fitting of the scaling relations analysed here. In Fig. 12, the results of our PCA analysis are visualized through ellipses. We plot 1σ and 2σ confidence ellipses, which contain approximately 68 per cent and 95 per cent of the data points respectively, and we consider the full COHRS cloud sample and the fiducial sample separately. Dashed lines in Fig. 12 indicate loci where a given set of parameters are constant around a certain estimate, as would be expected from the Larson’s relations.

Fig. 12(a) shows the relationship between cloud size and linewidth, which are the only two independent integrated properties we can measure for the clouds. For the fiducial sample we obtain

a formulation of the size linewidth relation close to the original work of Larson (1981), but the two quantities are only moderately correlated. For the full sample the two quantities are very weakly correlated. As observed in Section 8.1, the coarse spectral resolution of our data set and the optically thick CO tracer used by COHRS may play a role in rendering the relationship between σ_v and R_{eff} shallower than predicted by equation (21). Nevertheless, the relationship between σ_v and R_{eff} shows a large amount of scatter from both full and fiducial catalogues, which suggests that the clouds in our catalogue cannot be described as simple virialized objects with a single value of molecular gas mass surface density. A similar conclusion can be drawn by the second Larson’s law relation which connects σ_v and cloud mass from CO luminosity.

The scatter in the first Larson’s relation, as well as the large velocity dispersions measured for the clouds, could be driven by surface density variations (Heyer et al. 2009). In that case, σ_v and $\Sigma_{\text{mol}}R_{\text{eff}}$ should show a larger degree of correlation than other relations that involve the velocity dispersion (e.g. Miville-Deschênes et al. 2017). For our catalogue, however, this is not the case (see Fig. 12c), but we also measure a much shallower correlation than the one imposed by self-gravitation, as shown by the lines of constant α in Fig. 12(c). The clouds in the fiducial

Table 2. Summary of the Larson’s laws fitting performed using PCA. r indicates the Pearson’s correlation coefficient value.

Relation	β	Fiducial γ	Scatter (dex)	r	β	Full γ	Scatter (dex)	r
σ_v [km s ⁻¹] = $\beta(R_{\text{eff}} [\text{pc}])^\gamma$	1.79 ± 1.05	0.24 ± 0.02	0.41	0.40	1.91 ± 1.00	0.11 ± 0.01	0.41	0.16
σ_v [km s ⁻¹] = $\beta(M_{\text{lum}} [M_\odot])^\gamma$	1.05 ± 1.10	0.10 ± 0.01	0.41	0.42	1.45 ± 1.01	0.05 ± 0.01	0.40	0.21
σ_v [km s ⁻¹] = $\beta(\Sigma_{\text{mol}} R_{\text{eff}} [M_\odot \text{ pc}^{-1}])^\gamma$	0.80 ± 1.10	0.19 ± 0.01	0.41	0.40	1.24 ± 1.01	0.11 ± 0.01	0.40	0.23
$\alpha = \beta(M_{\text{lum}} [M_\odot])^\gamma$	237.86 ± 1.28	-0.43 ± 0.02	0.82	-0.62	512.86 ± 1.02	-0.58 ± 0.01	0.75	-0.69
$L_{\text{CO}} [\text{K km s}^{-1} \text{ pc}^2] = \beta(R_{\text{eff}} [\text{pc}])^\gamma$	32.47 ± 2.30	2.17 ± 0.38	0.28	0.95	19.05 ± 1.02	2.26 ± 0.06	0.29	0.92
$M_{\text{vir}} [M_\odot] = \beta(L_{\text{CO}} [\text{K km s}^{-1} \text{ pc}^2])^\gamma$	188.19 ± 2.12	0.75 ± 0.09	0.73	0.82	544.99 ± 1.06	0.65 ± 0.02	0.73	0.73

catalogue are globally closer to the virial equilibrium $\alpha \sim 1-2$, but the bulk of objects in the full sample seem to have $\alpha \gg 1$ (as observed in Section 7). The clouds in our fiducial sample are generally the most massive ones in the catalogue, so it is not surprising that those are closer to virialization, given the covariance between α and M_{lum} (Fig. 12d).

Plotting the cloud luminosities from CO against their effective radii (in Fig. 12e) provides an additional diagnostic for the surface density of the clouds. Both measurements from fiducial and full samples appear tightly clustered along the $\Sigma_{\text{mol}} = 100 M_\odot \text{ pc}^2$, however the average for the fiducial catalogue clouds is slightly above this line, while the average from the full catalogue is slightly below. The two quantities are covariant and result strongly correlated.

The virial parameter and cloud mass surface density that somehow shape the appearance of the Larson’s relations depend upon M_{mol} which in turn depends on the global value of the CO-to-H₂ conversion factor, α_{CO} . In Section 5 we estimated an $\alpha_{12\text{CO}(3-2)} = 8.7 M_\odot (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ by correcting the standard Galactic $\alpha_{12\text{CO}(1-0)}$ by the $^{12}\text{CO}(3-2)/^{12}\text{CO}(1-0)$ ratio, R_{31} . R_{31} looks highly variable across our survey area and the $\alpha_{12\text{CO}(3-2)}$ value we calculated is an approximation spanning both bright and faint emission. The $M_{\text{vir}}-L_{\text{CO}}$ diagram can be a useful diagnostic to test both cloud dynamical state and variations in $\alpha_{12\text{CO}(3-2)}$. The correlation between these two quantities appears significant considering that both fiducial and full sample register quite high values of Pearson’s correlation coefficients ($r \sim 0.8$). If the clouds are virialized, the relationship between M_{vir} and L_{CO} appears to be clustered across the $\alpha_{12\text{CO}(3-2)} = 10 M_\odot (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ constant line, which is close the value of $\alpha_{12\text{CO}(3-2)}$ we estimated independently. This set of observations argues that the clouds being near virial equilibrium is consistent with the values of $\alpha_{12\text{CO}(3-2)}$ that we argued for previously. This relationship is clearest for the massive clouds in the fiducial sample. Taking the full sample as virialized would imply the global average $\alpha_{12\text{CO}(3-2)}$ is much closer to $100 M_\odot (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ and the slope of the correlation is shallower than linear ($\gamma \sim 0.65$). Given the small scale of these objects, we are inclined to believe these small objects are unbound molecular clouds rather than our $\alpha_{12\text{CO}(3-2)}$ being grossly inappropriate (Heyer, Carpenter & Snell 2001).

The scaling relations we observe for COHRS clouds all have a significant amount of scatter (~ 0.5 dex) and are shallower than typically measured in the inner Galaxy despite our significantly larger sample sizes. This is particularly true if we consider the full catalogue instead of the fiducial sub-sample. Part of the reason might be attributed to the coarse spectral resolution of our data set and the optically thick tracer we use, but we can also interpret this as the fact that we are genuinely sensitive to a cloud population that show large linewidths (with respect to their sizes) and that are truly in a gravitationally unbound state.

10.2 COHRS cloud scaling relations in the context of other surveys

In this section we have gathered data from a variety of CO surveys in order to compare the scaling relations from the clouds in the COHRS sample to those derived from other Milky Way regions and nearby galaxies, as listed in Table 3.⁶ All of these studies define the identified objects as ‘molecular clouds’ (or depending on their sizes and masses, to ‘clumps’ or ‘Giant Molecular Clouds’) and they all use CO transition or isotopologues for the calculation of cloud properties. The data set parameters from where those clouds have been extracted are illustrated in Table 3. Average properties of the cloud ensembles are collected in Table 4. Fig. 13 illustrates the differences between cloud samples drawn from various galaxies in scaling relationship diagrams. As per Section 10, we use PCA to describe the relationship between the cloud properties in the different surveys and to illustrate where these objects are found in the various parameter spaces.

The left-hand panels of Fig. 13 show the ‘size–linewidth’ relation expressed through the effective radius and velocity dispersion of the clouds, for the compilation of Galactic studies on the top, and extragalactic studies on the bottom.

In Galactic studies (Fig. 13 top left), we can see that clouds identified in the Galactic Centre (Oka et al. 2001) and the outer Galaxy (Heyer et al. 2001) show the most extreme variations when compared to the bulk of the MW (as already suggested from Section 9.1). Our fiducial sample overlaps quite well with the ellipses defined by the clouds from the GRS by Roman-Duval et al. (2010) (in green) and those from Heyer et al. (2009) (in blue), all of which have been identified in a similar region of the Galaxy. Nevertheless, both our clouds and those of Heyer et al. (2009) show slightly larger velocity dispersions than the GRS for a same effective radii. This behaviour can potentially be attributed to a difference in tracer, given that both our sample and that of Heyer et al. (2009) clouds have been segmented from ^{12}CO emission, whilst Roman-Duval et al. (2010) objects are drawn from ^{13}CO data. Indeed, ^{12}CO can suffer from high-optical depth effects which can artificially broaden the line emission, and can also suffer from severe blending due to their higher abundance and ability to trace lower density regimes (Hughes et al. 2013) – effects that are reduced for the less abundant ^{13}CO . Another cause of the mismatch could be attributed to the lower spectral resolution of both our data set and that of Heyer et al. (2009), compared to the GRS data, unabling us from detecting lower velocity dispersion objects. Finally, the segmentation method could also play a role, but the effects of the different decomposition methods are more difficult to quantify. Nevertheless, we note that

⁶Some of these data were obtained via the Cloud Archive for Metadata, Library and Online Toolkit, CAMELOT (Ginsburg et al. 2016, <http://came-lot-project.herokuapp.com/>)

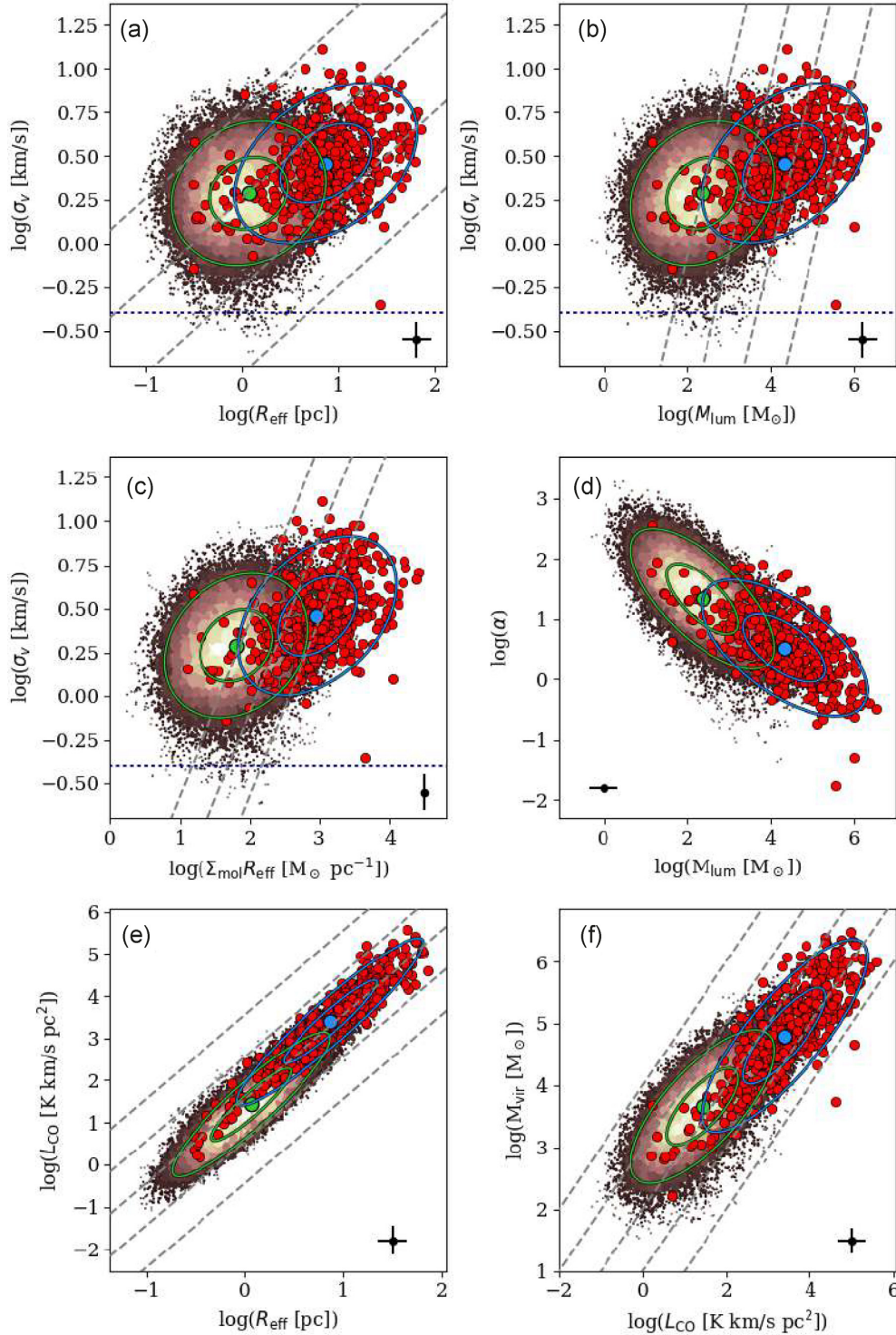


Figure 12. COHRS cloud scaling relations. The scaling relations are shown as bi-dimensional histograms, where the colours get brighter proportional to the number of data. The fiducial data are highlighted with red symbols. Green and cyan ellipses represent the PCA of the full and fiducial samples, respectively. Green and cyan points within the inner ellipses show the mean of the properties for full and fiducial sample, respectively. The ellipses contain ~ 68 per cent and ~ 95 per cent of the data. The horizontal, dotted line marks the spectral resolution of the COHRS data. The straight, dashed, grey lines indicate (from top to bottom, left to right): (a) virialized clouds ($\alpha = 1$) with $\Sigma_{\text{mol}} = 10^4, 10^3, 10^2, 10 M_{\odot} \text{ pc}^{-2}$; (b) virialized clouds ($\alpha = 1$) with $R_{\text{eff}} = 0.1, 1, 10, 100 \text{ pc}$ for a constant $\alpha_{\text{CO}} = 8.7 M_{\odot} / (\text{K km s}^{-1} \text{ pc}^2)$; (c) $\alpha = 10, 3, 1$; (e) $\Sigma_{\text{mol}} = 10^4, 10^3, 10^2, 10, 1 M_{\odot} \text{ pc}^{-2}$; (f) $\alpha_{\text{CO}} = 10^4, 10^3, 10^2, 10, 1 M_{\odot} / (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ assuming virialized clouds. In the bottom of each panel the typical error bars are given.

Table 3. Summary of the observation from which the clouds in Section 10.2 comparison have been identified: (1) Galactic region or nearby galaxy related to the observation; (2) CO isotopologue and transition used for the observation; (3) spatial resolution of the observation in arcsecond, degrees, or pc; (4) spectral resolution of the observation in km s^{-1} ; (5) reference where the data sets have been originally presented. Notes: Heyer et al. 2009 use a combination of $^{12}\text{CO}(1-0)$ data from the University of Massachusetts-Stony Brook Galactic Plane Survey (Sanders et al. 1986) and the $^{13}\text{CO}(1-0)$ data from the Boston University-FCRAO Galactic Ring Survey (Jackson et al. 2006) to redefine the clouds properties previously identified by Solomon et al. 1987. Data from nearby dwarfs (Bolatto et al. 2008) and nearby spirals (Donovan-Meyer et al. 2013) are obtained from a variety of data sets (see the cited works for further references and details); the values reported in the table are averages from the parameters of the different data sets involved in the paper.

Region	Tracer	θ_{FWHM}	Δv	Reference
(1)	(2)	(3)	(4)	(5)
		Milky Way	km s^{-1}	
MW – 1st quadrant	$^{12}\text{CO}(3-2)$	16.6 arcsec	1	Dempsey et al. (2013)
MW centre	$^{12}\text{CO}(1-0)$	34 arcsec	0.65	Oka et al. (1998)
Outer MW	$^{12}\text{CO}(1-0)$	50 arcsec	0.98	Heyer et al. (1998)
MW – 1st quadrant	$^{13}\text{CO}(1-0)$	45 arcsec	1	Sanders et al. (1986)
MW – 1st quadrant	$^{12}\text{CO}/^{13}\text{CO}(1-0)$	46 arcsec	0.212	Jackson et al. (2006)
Whole MW	$^{12}\text{CO}(1-0)$	$0.125^\circ/0.25^\circ$	0.65/1.3	Dame et al. (2001)
		Dwarfs		
Nearby dwarfs	$^{12}\text{CO}(1-0)/(2-1)$	37 pc	1.8	Bolatto et al. (2008)
LMC	$^{12}\text{CO}(1-0)$	8 pc	0.53	Hughes et al. (2010)
NGC300	$^{12}\text{CO}(2-1)$	37 pc	1.056	Faesi et al. (2016)
NGC6822	$^{12}\text{CO}(2-1)$	2 pc	5	Schruba et al. (2017)
		Spirals		
M64	$^{13}\text{CO}(1-0)$	75 pc	4.25	Rosolowsky & Blitz (2005)
M33	$^{12}\text{CO}(1-0)$	48 pc	2.6	Gratier et al. (2010)
Nearby spirals	$^{12}\text{CO}(1-0)$	<78 pc	5.8	Donovan-Meyer et al. (2013)
M51	$^{12}\text{CO}(1-0)$	40 pc	5	Colombo et al. (2014)
M100	$^{12}\text{CO}(1-0)$	216 pc	5	Pan et al. (2017)
NGC1068	$^{13}\text{CO}(1-0)$	98 pc	1.5	Tosaki et al. (2017)
NGC253	$^{12}\text{CO}(1-0)$	37 pc	5	Leroy et al. (2015)
M83	$^{12}\text{CO}(1-0)$	22.8 pc	2.57	Freeman et al. (2017)
		Peculiar		
Antennae	$^{12}\text{CO}(2-1)$	160 pc	4.9	Wei et al. (2012)
		Lenticular		
NGC4526	$^{12}\text{CO}(2-1)$	18 pc	10	Utomo et al. (2015)

the Heyer et al. (2009) and Roman-Duval et al. (2010) catalogues are constructed by watershed segmentation methods which tend to decompose objects with sizes closer to the resolution element (Hughes et al. 2013, see also Table 4). We, instead, consider the full dynamical range of the survey. Therefore it is not surprising that our sampling in effective radius is larger, especially for the full catalogue. We can also see that the clouds identified in the Dame survey data by Rice et al. (2016) and Miville-Deschênes et al. (2017), which span the entire Galactic plane, sample larger effective radii than the other samples, but this may be a consequence of the coarser resolution of the Dame survey. Interestingly, the catalogue construction techniques developed by the authors of those two works have been designed to overcome the data limitations,⁷ but they produce quite dissimilar results.

Looking at the extragalactic works (Fig. 13 bottom left panel), clouds from the different galaxy groups occupy different parts of the size–linewidth plane. Objects from our fiducial sample span

sizes that range from the smaller clouds in dwarf galaxies to average clouds in spirals. Velocity dispersions between those sub-samples are also comparable, however clouds in our fiducial and full catalogues show larger σ_v values with respect to similar sized objects in dwarf galaxies. As for the Milky Way catalogues, this can be simply due to differences in spatial and/or velocity resolution (Hughes et al. 2013).

The slope of the relationship between σ_v and R_{eff} calculated through PCA of our fiducial sample, the GRS catalogue and full Milky Way catalogue from Miville-Deschênes et al. (2017) are quite comparable (~ 0.3 – 0.4), while the slope for the outer Galaxy clouds is shallower (~ 0.05), and the Galactic centre clouds and the full Galaxy cloud catalogue of Rice et al. (2016) are steeper (~ 0.7). The slope of the relationships from the various galaxy groups, deducible from the PCA ellipse orientation, are also quite different between each other: we obtain ~ 0.65 for ‘Spirals’, ~ 0.8 for ‘Peculiar’, and ~ 0.1 for ‘Dwarfs’ and ‘Lenticular’ types. Nevertheless, these orientations could also be set by the resolution biases as suggested by Hughes et al. (2013). Note, also, that PCA results can be different from the fit between σ_v and R_{eff} performed on the various papers where the data are taken from, due to different fitting techniques, sub-sample considered, etc. However, a

⁷The two catalogues have been built following two different philosophies: Rice et al. (2016) catalogue collects only Milky Way GMCs, while Miville-Deschênes et al. (2017) work aims to assemble all of the flux into discrete structures of any mass, see Table 4.

Table 4. Summary of the cloud properties used for the comparison in Section 10.2: (1) Galactic region or nearby galaxy related to the observation; (2) number of objects identified; (3) effective radius; (4) velocity dispersion; (5) molecular gas surface density; (6) scaling parameter; (7) method used to identify clouds: GAUSSCLUMPS (Stutzki & Guesten 1990), CLUMPFIND (Williams et al. 1994), CPROPS (Rosolowsky & Leroy 2006), dendrogram (Rosolowsky et al. 2008), SCIMES (Colombo et al. 2015), CPROPSOO (Leroy et al. 2015); (8) catalogue reference. All properties are presented as median, 25th and 75th percentiles of their distributions.

Region	N. objects	R_{eff} pc	σ_v km s $^{-1}$	Σ_{mol} $M_{\odot} \text{ pc}^{-2}$	σ_0^2 km $^2 \text{ s}^{-2}$	Method	Reference
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Milky Way							
MW – 1st quadrant	542	8.93 $^{19.19}_{4.51}$	2.95 $^{4.39}_{2.17}$	129.27 $^{216.48}_{79.86}$	1.06 $^{2.36}_{0.51}$	SCIMES	This work – fiducial sample
MW – 1st quadrant	35446	1.13 $^{2.16}_{0.60}$	1.94 $^{2.66}_{1.44}$	53.30 $^{92.15}_{32.10}$	3.23 $^{7.33}_{1.41}$	SCIMES	This work – full sample
MW centre	165	9.10 $^{14.60}_{5.70}$	7.80 $^{12.70}_{5.40}$	1171.95 $^{1922.26}_{690.73}$	8.12 $^{17.21}_{3.38}$	Sigma clipping	Oka et al. (2001)
Outer MW	10156	2.69 $^{7.39}_{1.04}$	0.79 $^{0.94}_{0.68}$	1.14 $^{1.45}_{0.94}$	0.23 $^{0.60}_{0.09}$	Sigma clipping	Heyer et al. (2001)
MW – 1st quadrant	158	3.20 $^{5.47}_{1.63}$	2.00 $^{2.60}_{1.60}$	189.54 $^{372.64}_{109.29}$	1.34 $^{2.48}_{0.73}$	Sigma clipping	Heyer et al. (2009)
MW – 1st quadrant	749	7.40 $^{11.90}_{3.40}$	1.32 $^{1.70}_{0.98}$	144.00 $^{184.60}_{103.60}$	0.23 $^{0.44}_{0.14}$	CLUMPFIND	Roman-Duval et al. (2010)
Whole MW	8107	25.07 $^{44.70}_{12.81}$	3.63 $^{5.50}_{2.25}$	16.50 $^{42.57}_{6.62}$	0.57 $^{1.19}_{0.26}$	Own	Miville-Deschenes et al. (2017)
Whole MW	1038	27.07 $^{41.48}_{17.70}$	2.08 $^{2.96}_{1.47}$	21.22 $^{50.87}_{10.99}$	0.17 $^{0.29}_{0.09}$	Dendrogram	Rice et al. (2016)
Dwarfs							
Nearby dwarfs	110	27.54 $^{41.69}_{18.73}$	3.39 $^{4.65}_{2.63}$	74.64 $^{132.69}_{37.40}$	0.40 $^{0.69}_{0.23}$	CPROPS	Bolatto et al. (2008)
LMC	524	12.05 $^{15.80}_{8.57}$	1.40 $^{1.79}_{1.07}$	60.78 $^{96.03}_{38.19}$	0.17 $^{0.28}_{0.10}$	CPROPS	Wong et al. (2011)
NGC300	45	4.15 $^{4.58}_{3.69}$	1.83 $^{2.38}_{1.49}$	686.59 $^{1303.61}_{352.29}$	0.82 $^{1.27}_{0.61}$	CPROPS	Faesi et al. (2016)
NGC6822	111	2.60 $^{3.30}_{2.05}$	1.15 $^{1.41}_{0.80}$	7.47 $^{11.93}_{3.80}$	0.45 $^{0.73}_{0.27}$	CPROPSOO	Schruba et al. (2017)
Spirals							
M64	25	86.00 $^{115.00}_{62.00}$	9.79 $^{14.89}_{7.66}$	57.14 $^{69.89}_{43.20}$	1.14 $^{2.22}_{0.73}$	CLUMPFIND	Rosolowsky & Blitz (2005)
M33	308	37.00 $^{56.00}_{24.75}$	8.85 $^{11.00}_{7.20}$	31.66 $^{64.39}_{16.23}$	2.10 $^{3.57}_{1.29}$	CPROPS	Gratier et al. (2012)
Nearby spirals	96	87.90 $^{108.75}_{61.33}$	4.77 $^{7.84}_{3.79}$	188.05 $^{272.61}_{138.50}$	0.32 $^{0.54}_{0.19}$	CLUMPFIND	Donovan-Meyer et al. (2013)
M51	1506	35.00 $^{55.00}_{32.00}$	5.90 $^{8.00}_{4.20}$	139.38 $^{258.87}_{82.87}$	0.80 $^{1.58}_{0.43}$	CPROPS	Colombo et al. (2014)
NGC253	10	30.50 $^{40.75}_{24.25}$	9.36 $^{10.53}_{7.55}$	4760.04 $^{9018.37}_{2218.98}$	2.61 $^{3.00}_{2.34}$	CPROPSOO	Leroy et al. (2015)
M100	165	301.70 $^{403.90}_{229.40}$	7.30 $^{9.90}_{5.20}$	32.29 $^{52.75}_{17.53}$	0.17 $^{0.36}_{0.09}$	CPROPS	Pan et al. (2017)
NGC1068	187	100.00 $^{133.00}_{74.50}$	4.13 $^{5.33}_{3.08}$	26.12 $^{59.07}_{15.21}$	0.18 $^{0.26}_{0.11}$	CLUMPFIND	Tosaki et al. (2017)
M83	873	46.10 $^{56.89}_{36.68}$	3.24 $^{4.61}_{2.36}$	79.53 $^{146.65}_{54.30}$	0.22 $^{0.45}_{0.12}$	CPROPSOO	Freeman et al. (2017)
Peculiar							
Antennae	67	168.00 $^{233.50}_{118.00}$	11.90 $^{16.25}_{9.30}$	224.78 $^{387.46}_{112.18}$	0.88 $^{1.31}_{0.59}$	CLUMPFIND	Wei et al. (2012)
Lenticular							
NGC4526	103	16.06 $^{22.53}_{11.86}$	8.20 $^{9.14}_{7.24}$	1237.36 $^{1740.62}_{907.78}$	3.95 $^{6.15}_{2.58}$	CPROPS	Utomo et al. (2015)

uniform analysis of the data, drawn without homogenization from different surveys, shows significant variation between the different catalogues. This highlights the fact that these scaling relations are rather sensitive to the segmentation method used as well as the specifications of each data set, and should therefore be used with caution.

So far we have considered clouds as self-contained objects, however, the Galactic environment could potentially shape the properties of the clouds in different ways as we noticed in Section 9. This is particularly evident in nearby spirals (Colombo et al. 2014; Leroy et al. 2017; Sun et al. 2018). Comparing the Milky Way catalogues on the $\sigma_v^2/R_{\text{eff}} - \Sigma_{\text{mol}}$ plane provides a different interpretation of the cloud dynamical state and their interaction with the environment. Clouds bound simply by their self-gravity would tightly cluster across the solid black line in Fig. 13 (upper and lower right panels). However, clouds seem almost always shifted toward higher values of $\sigma_v^2/R_{\text{eff}}$ with respect to the line imposed by virial equilibrium. Dobbs, Burkert & Pringle (2011) interpret this fact as evidence that clouds are simply gravitationally unbound entities (even though they could still be confined by ram/thermal pressure, as later suggested by e.g. Duarte-Cabral & Dobbs 2017).

Indeed, Field et al. (2011, and references therein) consider the case of clouds being confined by an external pressure (P_e) and suggest an alternative version of equation 20 that takes into account this parameter:

$$\sigma_v = \sqrt{\frac{1}{3} \left(\pi \Gamma G \Sigma_{\text{mol}} + \frac{4P_e}{\Sigma_{\text{mol}}} \right) R_{\text{eff}}}, \quad (23)$$

where $\Gamma = 0.6$ for clouds with constant density. Constant pressure loci are indicated as dashed lines in Fig. 13 (upper and lower right panels). Looking at the upper right panel of Fig. 13 it appears that proceeding from the outer Galaxy to the Galactic centre passing from the inner Galaxy, clouds could be bound if subjected to a crescent ambient pressure. The lower right panel of Fig. 13 shows that most of the clouds in the various groups appear bound by self-gravity; only objects in the ‘Dwarfs’ group clearly require ambient pressure confinement to explain their linewidths if they are bound. Clouds have also similar values of molecular gas mass surface densities, on average quite close to $\Sigma_{\text{mol}} \sim 200 M_{\odot} \text{ pc}^{-2}$, except for the ‘Lenticular’ type which has cloud surface densities an order of magnitude larger than the other groups. Again, drawing firm conclusions about the intrinsic physics

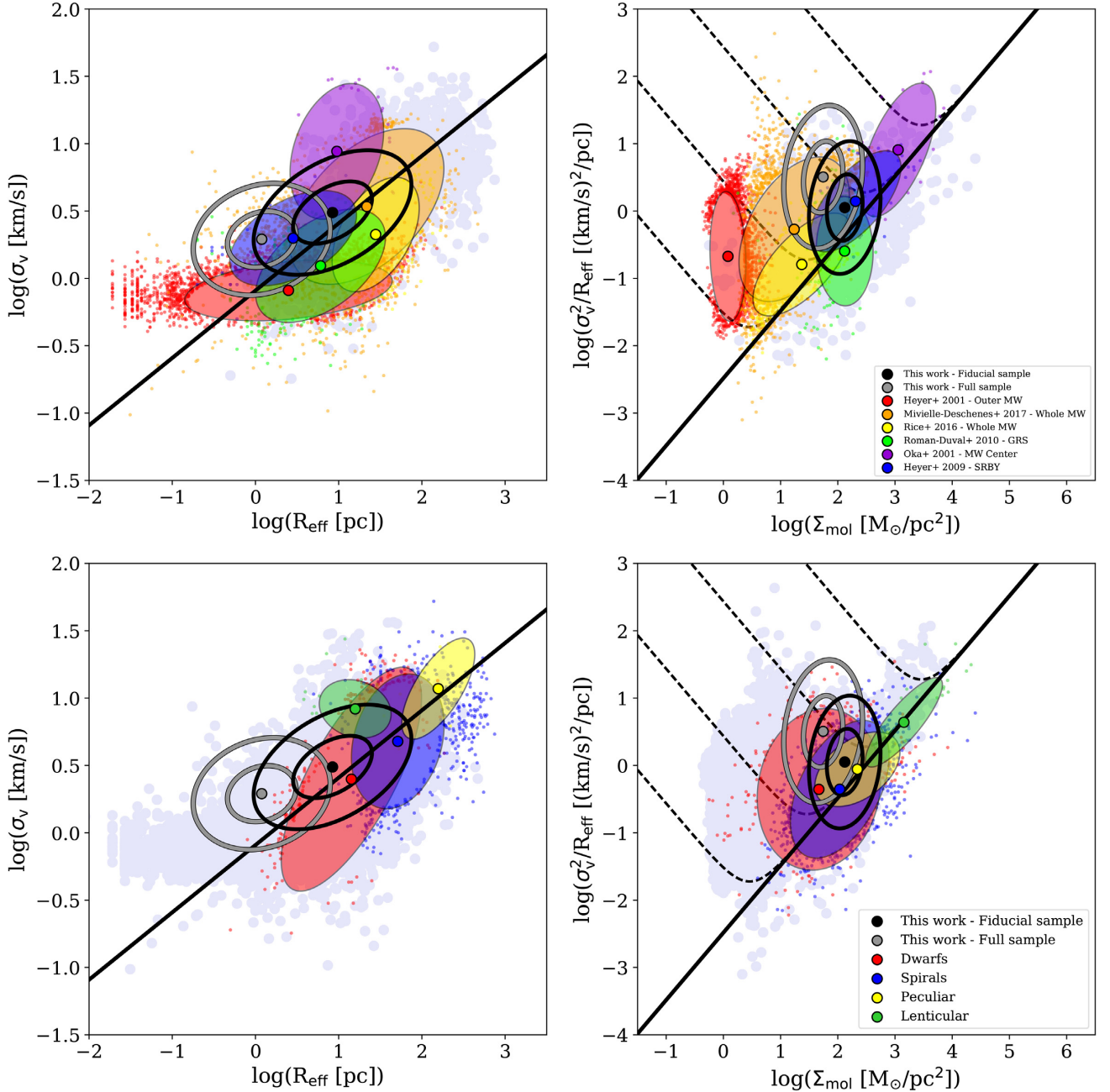


Figure 13. Comparison of scaling relationships between cloud properties from different catalogues. The coloured ellipses show the density distribution of the data approximated with PCA and contain ~ 95 per cent of the data. For the COHRS catalogue, the inner ellipse contains ~ 68 per cent of the data. The central dots show the mean of the cloud property distributions. Small coloured markers illustrate data outside the 95 per cent confidence ellipses, except for the COHRS clouds. The upper panels include Milky Way cloud catalogues as PCA ellipses, while grey markers indicate the cloud properties from the collection of nearby galaxy catalogues. The opposite applies for the lower panels. Galaxy groups and references are summarized in Tables 3 and 4. In the right-hand panels, the black solid line shows equation 20 for $\Sigma_{\text{mol}} = 200 \text{ M}_{\odot} \text{ pc}^{-2}$. In the left-hand panels, the black solid line represents equation (23) with $P_e = 0$. Dashed black lines show equation (23) where the ambient medium pressure is included, from bottom to top: $P_e/k_B = 10^2, 10^4, 10^6$, and 10^8 K cm^{-3} .

described by these diagrams is hampered by the various catalogue creation techniques and survey designs which need to be taken with caution.

To summarize, it is challenging to derive firm conclusions from cloud catalogues built from heterogeneous data sets. Cloud definition and their derived properties can be strongly influenced by the combination of survey designs and cloud identification methods. Nevertheless across the scaling relation diagrams of

Milky Way and nearby galaxies a few common features can be discerned. Most of the cloud samples show a sub-linear scaling between their velocity dispersions and effective radii which tend to virialization prescription. Small objects (of typically a pc in size), however, do not follow this description and can be considered simply gravitationally unbound or bound by ambient pressure. For the Milky Way, those clouds appear to have a little contribution to

the total molecular gas budget, since mass spectra from COHRS clouds and from other surveys (Heyer et al. 2001; Roman-Duval et al. 2010; Heyer & Dame 2015; Rice et al. 2016) all indicate that most of the molecular material is contained into few large GMC-like entities. The difficulties in the interpretation of the results of this analysis, in term of actual difference in cloud properties, underlines also the needs of having, at least for the Milky Way, a single cloud catalogue constructed with a consistent extraction method and from surveys with similar designs.

11 SUMMARY AND OUTLOOK

We presented an analysis of the molecular gas properties imaged by the COHRS through the cloud identification. We applied the SCIMES algorithm to obtain a catalogue of integrated properties from more than 85 000 clouds in the first Galactic quadrant. We corrected cloud properties for instrumental biases by applying the techniques described by Rosolowsky & Leroy (2006). Our main results can be summarized as follows.

(i) By comparing the University of Massachusetts Stony Brook Survey $^{12}\text{CO}(1-0)$ data with the smoothed and reprojected COHRS data we obtained an average $^{12}\text{CO}(3-2)$ over $^{12}\text{CO}(1-0)$ flux ratio, $R_{31} = 0.5$. This translates into a $^{12}\text{CO}(3-2)$ -to- H_2 conversion factor: $X_{^{12}\text{CO}(3-2)} = 4 \times 10^{20} (\text{cm}^2 \text{K km s}^{-1})^{-1}$ or $\alpha_{^{12}\text{CO}(3-2)} = 8.7 M_{\odot} (\text{K km s}^{-1} \text{pc}^2)^{-1}$.

(ii) We calculated the distance to the clouds using the distance measurements from Zetterlund et al. (2018) which applied the methods of Ellsworth-Bowers et al. (2013) starting from the Bolocam Galactic Plane Survey. This establishes a grid of 2202 pixels in the position–position–velocity coordinate system of COHRS data set. In this way, we obtained 406 clouds with a distance pixel within them, 41 896 clouds where a parental structure of the dendrogram have a distance point within it, and 42 718 isolated clouds for which we choose the distance pixel closest to the outer boarder of the object.

(iii) We separated the analysis between two sub-samples: $\sim 35\,500$ well-resolved objects, defined ‘full’ sample, and a ‘fiducial’ sample constituted by 542 clouds with well-defined distances. The latter shows properties quite similar to GRS catalogue of Roman-Duval et al. (2010). Nevertheless, velocity dispersion and the properties that depend on it (as virial mass and virial parameter) are shifted towards higher values with respect to GRS. We attributed this to: (1) the coarse velocity resolution of the COHRS data relative to the structures that COHRS can resolve spatially and (2) the different tracers used to image the molecular medium by COHRS and GRS.

(iv) We analysed the cumulative distributions of mass derived from CO luminosity and effective radius in detail. We found a $\gamma \sim -1.75$ and $\gamma \sim -2.80$ for the spectral index of mass and size spectra, respectively. Both distributions show high-end truncations at $M_{\text{lum}} \sim 3 \times 10^6 M_{\odot}$ and $R_{\text{eff}} \sim 70 \text{ pc}$, respectively. Mass and size spectra show subtle differences when calculated within different Galactocentric annuli.

(v) COHRS clouds show CO integrated intensity and velocity dispersion values intermediate between the Galactic centre and the outer Galaxy. In particular we observed that clouds in the fiducial sample have masses similar to Galactic centre clouds.

(vi) Approximately 35 per cent of the clouds and the catalogued flux look embedded into spiral arms, by considering a model that foresees 4 main spiral arms and the Local spurs, where arms have a fixed width of 600 pc.

(vii) We used principal component analysis to study the scaling relations from ‘full’ and ‘fiducial’ samples separately. We found mainly moderate correlations (Pearson $r \sim 0.5$) between most of the properties only when the fiducial sample clouds are considered. Only the $L_{\text{CO}}-R_{\text{eff}}$ and $M_{\text{vir}}-L_{\text{CO}}$ show high levels of correlation ($r > 0.8$ from both subsamples). However, these quantities are expected to be intrinsically correlated given the observables that contribute to them. For the fiducial sample, we observed a size–linewidth relation shallower (slope ~ 0.3) and weaker (Pearson $r \sim 0.4$) when compared to previous studies in the same Galactic region. We attribute this disparity to the different tracers (which have different optical depths), and the different velocity resolutions of the surveys.

(viii) By collecting catalogues from various, heterogeneous Galactic and nearby galaxy CO surveys we found that, on average, the scaling between cloud effective radii and velocity dispersions largely scatter around a power law of slope 0.5. Most of the cloud linewidths would be consistent with a state of virialization if the clouds are bound by an overpressurized environment. We warn that physical interpretations can be strongly shaped by methodological choices in term of survey designs, cloud identification methods, and the CO isotopologues observed.

In this work we have studied the properties of the clouds identified in the COHRS data through the approach of using integrated cloud properties. These properties reduce all the available information from the mapping of the clouds down to single numbers. These measurements can also be achieved when the objects are point sources, i.e. a few beams apart, as in the case of nearby galaxy observations. Even with the advent of ALMA the smallest scale resolvable outside the Local Group remains the size of a GMC ($\sim 20 \text{ pc}$). However, in recent years, the Galactic plane has been explored by a wealth of high-resolution spectroscopic large scale, blind surveys able to image down to the size of clumps (e.g. COHRS, CHIMPS, SEDIGISM, FUGIN; and the ongoing FQS, Benedettini et al. 2017, and OGHReS, PIC. König). The application of SCIMES on these surveys will allow us to obtain thousands of clouds with a well-resolved internal structure and defined outer edges. This will also give the opportunity to study the ‘resolved’ properties of the clouds as probability distribution functions (PDFs), turbulence, morphology (elongation versus roundness), kinematics (through moment maps), clump formation efficiency (CFE, Eden et al. 2013), and SF efficiencies (in combination with sub-mm surveys such as ATLASGAL: Schuller et al. 2009; Csengeri et al. 2014; Urquhart et al. 2014; Hi-GAL: Molinari et al. 2016; Elia et al. 2017; BGPS: Aguirre et al. 2011; Ginsburg et al. 2013) in a truly statistical fashion.

ACKNOWLEDGEMENTS

We thank the anonymous referee for carefully reading our manuscript. DC acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) through project number SFB956C. DC thanks Axel Weiß, Karl Menten, Friedrich Wyrowski, and Carsten König for the useful discussions. ER acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-03987. ADC acknowledges the support of the UK Science and Technology Facilities Council consolidated grants ST/N000706/1. This research made use of ASTROPY, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013).

REFERENCES

- Aguirre J. E. et al., 2011, *ApJS*, 192, 4
- Astropy Collaboration, 2013, *A&A*, 558, A33
- Barnes P. J., Muller E., Indermuehle B., O'Dougherty S. N., Lowe V., Cunningham M., Hernandez A. K., Fuller G. A., 2015, *ApJ*, 812, 6
- Barnes P. J., Hernandez A. K., Muller E., Pitts R. L., 2018, *ApJ*, 866, 19
- Benedettini M. et al., 2017, *Mem. Soc. Astron. Ital.*, 88, 722
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Bertoldi F., McKee C. F., 1992, *ApJ*, 395, 140
- Blitz L., Thaddeus P., 1980, *ApJ*, 241, 676
- Bolatto A. D., Leroy A. K., Rosolowsky E., Walter F., Blitz L., 2008a, *ApJ*, 686, 948
- Bolatto A. D., Leroy A. K., Rosolowsky E., Walter F., Blitz L., 2008b, *ApJ*, 686, 948
- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, 51, 207
- Brand J., Blitz L., 1993, *A&A*, 275, 67
- Colombo D. et al., 2014, *ApJ*, 784, 3
- Colombo D., Rosolowsky E., Ginsburg A., Duarte-Cabral A., Hughes A., 2015, *MNRAS*, 454, 2067
- Csengeri T. et al., 2014, *A&A*, 565, A75
- Dame T. M., Hartmann D., Thaddeus P., 2001, *ApJ*, 547, 792
- Dempsey J. T., Thomas H. S., Currie M. J., 2013, *ApJS*, 209, 8
- di Francesco J., Evans N. J., II, Caselli P., Myers P. C., Shirley Y., Aikawa Y., Tafalla M., 2007, in Reipurth B., Jewitt D., Keil K., eds, *Protostars and Planets V*, 17. University of Arizona Press, Tucson, AZ, USA, p. 17
- Dobbs C. L., Burkert A., Pringle J. E., 2011, *MNRAS*, 413, 2935
- Duarte-Cabral A., Dobbs C. L., 2016, *MNRAS*, 458, 3667
- Duarte-Cabral A., Dobbs C. L., 2017, *MNRAS*, 470, 4261
- Duarte-Cabral A., Acreman D. M., Dobbs C. L., Mottram J. C., Gibson S. J., Brunt C. M., Douglas K. A., 2015, *MNRAS*, 447, 2144
- Eden D. J., Moore T. J. T., Morgan L. K., Thompson M. A., Urquhart J. S., 2013, *MNRAS*, 431, 1587
- Elia D. et al., 2017, *MNRAS*, 471, 100
- Ellsworth-Bowers T. P. et al., 2013, *ApJ*, 770, 39
- Ellsworth-Bowers T. P., Rosolowsky E., Glenn J., Ginsburg A., Evans N. J., II, Battersby C., Shirley Y. L., Svoboda B., 2015, *ApJ*, 799, 29
- Ester M., Kriegel H.-P., Sander J., Xu X., 1996, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. KDD'96*, AAAI Press, p. 226
- Field G. B., Blackman E. G., Keto E. R., 2011, *MNRAS*, 416, 710
- Freeman P., Rosolowsky E., Kruijssen J. M. D., Bastian N., Adamo A., 2017, *MNRAS*, 468, 1769
- Fukui Y., Kawamura A., 2010, *ARA&A*, 48, 547
- Ginsburg A. et al., 2013, *ApJS*, 208, 14
- Ginsburg A. et al., 2016, *Astrophysics Source Code Library*, record ascl:1605.006
- Goodman A. A. et al., 2014, *ApJ*, 797, 53
- Heyer M., Dame T. M., 2015, *ARA&A*, 53, 583
- Heyer M., Krawczyk C., Duval J., Jackson J. M., 2009, *ApJ*, 699, 1092
- Heyer M. H., Carpenter J. M., Snell R. L., 2001, *ApJ*, 551, 852
- Houllahan P., Scalo J., 1992, *ApJ*, 393, 172
- Hughes A. et al., 2013, *ApJ*, 779, 46
- Jackson J. M. et al., 2006, *ApJS*, 163, 145
- Jeffreson S. M. R., Kruijssen J. M. D., 2018, *MNRAS*, 476, 3688
- Kauffmann J., Pillai T., Goldsmith P. F., 2013, *ApJ*, 779, 185
- Koda J. et al., 2009, *ApJ*, 700, L132
- Larson R. B., 1981, *MNRAS*, 194, 809
- Leroy A. K. et al., 2016, *ApJ*, 831, 16
- Leroy A. K. et al., 2017, *ApJ*, 846, 71
- Miville-Deschênes M.-A., Murray N., Lee E. J., 2017, *ApJ*, 834, 57
- Molinari S. et al., 2016, *A&A*, 591, A149
- Oka T., Hasegawa T., Sato F., Tsuboi M., Miyazaki A., Sugimoto M., 2001, *ApJ*, 562, 348
- Pearson K., 1901, *Phil. Mag.*, 2, 559
- Peñaloza C. H., Clark P. C., Glover S. C. O., Klessen R. S., 2018, *MNRAS*, 475, 1508
- Pineda J. E., Rosolowsky E. W., Goodman A. A., 2009, *ApJ*, 699, L134
- Ramón-Fox F. G., Bonnell I. A., 2018, *MNRAS*, 474, 2028
- Rathborne J. M., Johnson A. M., Jackson J. M., Shah R. Y., Simon R., 2009, *ApJS*, 182, 131
- Reid M. J. et al., 2014, *ApJ*, 783, 130
- Reid M. J., Dame T. M., Menten K. M., Brunthaler A., 2016, *ApJ*, 823, 77
- Reina-Campos M., Kruijssen J. M. D., 2017, *MNRAS*, 469, 1282
- Rice T. S., Goodman A. A., Bergin E. A., Beaumont C., Dame T. M., 2016, *ApJ*, 822, 52
- Rigby A. J. et al., 2016, *MNRAS*, 456, 2885
- Roman-Duval J., Jackson J. M., Heyer M., Johnson A., Rathborne J., Shah R., Simon R., 2009, *ApJ*, 699, 1153
- Roman-Duval J., Jackson J. M., Heyer M., Rathborne J., Simon R., 2010, *ApJ*, 723, 492
- Rosolowsky E., 2005, *PASP*, 117, 1403
- Rosolowsky E., Blitz L., 2005, *ApJ*, 623, 826
- Rosolowsky E., Leroy A., 2006, *PASP*, 118, 590
- Rosolowsky E. W., Pineda J. E., Kauffmann J., Goodman A. A., 2008, *ApJ*, 679, 1338
- Sanders D. B., Clemens D. P., Scoville N. Z., Solomon P. M., 1986, *ApJS*, 60, 1
- Schinnerer E. et al., 2017, *ApJ*, 836, 62
- Schuller F. et al., 2009, *A&A*, 504, 415
- Schuller F. et al., 2017, *A&A*, 601, A124
- Scoville N. Z., Solomon P. M., 1975, *ApJ*, 199, L105
- Scoville N. Z., Yun M. S., Sanders D. B., Clemens D. P., Waller W. H., 1987, *ApJS*, 63, 821
- Shirley Y. L. et al., 2013, *ApJS*, 209, 2
- Solomon P. M., Sanders D. B., Scoville N. Z., 1979, *ApJ*, 232, L89
- Solomon P. M., Rivolo A. R., Barrett J., Yahil A., 1987, *ApJ*, 319, 730
- Stutzki J., Giesten R., 1990, *ApJ*, 356, 513
- Sun J. et al., 2018, *ApJ*, 860, 172
- Umamoto T. et al., 2017, *PASJ*, 69, 78
- Urquhart J. S. et al., 2014, *MNRAS*, 443, 1555
- Vallée J. P., 2014, *AJ*, 148, 5
- Vallée J. P., 2017, *Astron. Rev.*, 13, 113
- Warren B. E. et al., 2010, *ApJ*, 714, 571
- Wegg C., Gerhard O., Portail M., 2015, *MNRAS*, 450, 4050
- Williams J. P., McKee C. F., 1997, *ApJ*, 476, 166
- Williams J. P., de Geus E. J., Blitz L., 1994, *ApJ*, 428, 693
- Wilson C. D., Scoville N., Madden S. C., Charmandaris V., 2003, *ApJ*, 599, 1049
- Wong T. et al., 2011, *ApJS*, 197, 16
- Zetterlund E., Glenn J., Rosolowsky E., 2018, *MNRAS*, 480, 893

APPENDIX A: NEW SCIMES VERSION

Together with this paper we release a new version of SCIMES (v.0.3.2) on the github page of the project,⁸ which includes several major upgrades. The new version is approximately 30 times faster than the version previously released online. A few new features have been added. In the final catalogue, the code can now retain dendrogram branches and leaves that cannot be uniquely associated with the identified clusters. The affinity matrix scaling parameters (see Colombo et al. 2015 for further details) can be restricted to only searching above a specified signal-to-noise ratio (default: SNR > 3). The ‘SNR affinity matrix’ is built by inputting the brightness temperature peak value of each structure considered in the matrix construction divided by a fixed (user-defined) noise estimation. The temperature peak value is collected in the attribute height of the `astrodendro.structure.Structure` class. The scaling parameter is now set by the largest gap within the affinity matrices above the given SNR. This increases the code stability and avoids considering large and incoherent objects. The algorithm updates

⁸<https://github.com/Astroua/SCIMES>

information about the cluster type within the dendrogram: ‘T’ for trunks (i.e. dendrogram branches without parent); ‘B’ for branches (structures with parent and children); ‘L’ for leaves (structures without children). The code now also reports the number of leaves and unique identifiers for the ID of the structure parent and ancestor. Finally, the SCIMES main class also provides mask cubes for the identified clusters, as well as for all leaves and trunks in the dendrogram.

APPENDIX B: COMPARISON BETWEEN ASSUMED AND KINEMATIC DISTANCES TO THE CLOUDS

The kinematic distance method consists of the derivation of Galactocentric radius and distance to a clump of gas through the knowledge of its spectroscopic radial velocity and the assumption of a given rotation curve (e.g. Roman-Duval et al. 2009). The Galactocentric radius can be uniquely determined through the following equation for a given longitude (l) and radial velocity (v_r):

$$r = R_{\odot} \sin(l) \frac{v(r)}{v_r + v_{\odot} \sin(l)}, \quad (\text{B1})$$

where R_{\odot} and v_{\odot} are the galactocentric radius and orbital velocity of the Sun, respectively; $v(r)$ indicates the rotation curve, and v_r is the spectroscopical radial velocity of the cloud.

In the inner Galaxy ($r < R_{\odot}$) the distance calculated through the kinematic method is not unique. Each galactocentric radius corresponds to two distances along the line of sight, called *near* and the *far* kinematic distances, which are located on either side of the *tangent point*. The tangent point is the region where the velocity vector of the cloud is aligned with the line of sight. The near and far distances can be calculated via

$$d = R_{\odot} \cos(l) \pm \sqrt{r^2 - R_{\odot}^2 \sin^2(l)}. \quad (\text{B2})$$

The duality of the distance calculated within the inner Galaxy is referred to as the *kinematic distance ambiguity* which is challenging to break without additional information.

Fig. B1 shows the comparison between the distance we assumed for the clouds via association with BGPS sources (as described in Section 6) and the kinematic distance calculated to the clouds through equations (B1) and (B2). The kinematic distance attributed to each cloud is given by the longitude and radial velocity of its centroid position (`glon_deg` and `vlsr_kms` parameters in the catalogue, respectively). As Galactic rotation curve we use the model defined by Brand & Blitz (1993) with the parameter calculated by Reid et al. (2014):

$$v(r) [\text{km s}^{-1}] = (240 \pm 9)(r/R_{\odot})^{0.00 \pm 0.02}. \quad (\text{B3})$$

For a given cloud the near or the far distance is plotted considering the one which is closer in value to the assumed one. Around 60 percent of the sample analysed in the paper can be attributed to the far distance, while the remaining 40 percent to near distances. Nevertheless, only 30 percent of the sources in the BGPS catalogue are at far distances. Most of the COHRS clouds at the far distances have ‘broadcasted’ or ‘closest’ distance attribution. Indeed, the exact values of the kinematic distances can be quite dissimilar to the assumed ones for the full sample. Nevertheless, the distances assumed for the clouds in the ‘fiducial’ sample are almost indistinguishable to the ones derived using the kinematic method, with only few exceptions. Moreover, 60 percent of ‘fiducial’ clouds

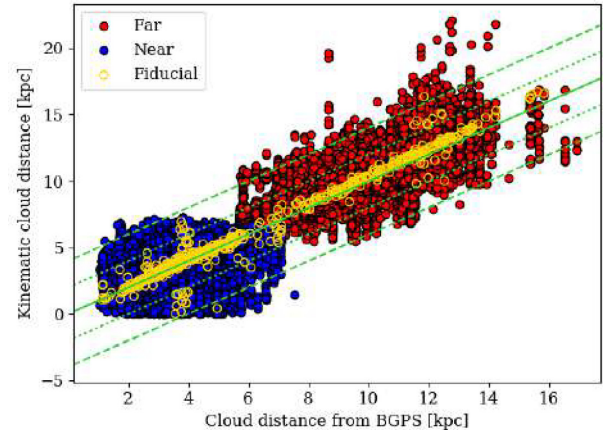


Figure B1. Comparison between cloud assumed distance from BGPS sources and calculated kinematic distance. Blue points indicate near cloud distance, while red points the far cloud distances. The fiducial sample is highlighted with yellow circles. The green full line represents unity, while the green dotted and dashed line a factor 2 and 4 of scatter, respectively.

is attributed to the near distances, the remaining 40 percent to the far distances.

APPENDIX C: PRINCIPAL COMPONENT ANALYSIS

In Section 10, we analysed the correlations between cloud integrated properties using the PCA technique (Pearson 1901) applied to the bivariate relationships between cloud properties. While PCA is a dimensionality reduction technique, we used it in the place of regression to identify the directions of maximal variance of the data and assumed that this direction defines the scaling relationship. The perpendicular directions define the scatter within the data. This approach is particularly useful for our analysis since the underlying correlations have large intrinsic scatter.

The PCA algorithm constructs the covariance matrix between quantities under analysis, calculating the largest and smallest eigenvectors of this matrix. The orientation of the largest eigenvector with respect to the x -axis defines the slope of the relation (i.e., the tangent of the angle between the largest eigenvector and the positive x -axis directions). By construction, the smaller eigenvector is perpendicular to the larger eigenvector. The scatter of the relation is provided by PCA as $\sigma_{\min} = 2\sqrt{\lambda_{\min}}$, where λ_{\min} is the smaller eigenvalue. Generally, two quantities can be considered as correlated if $\lambda_{\text{maj}} \gg \lambda_{\min}$, where λ_{maj} is the largest eigenvalue.

To obtain the uncertainty on the relationship slopes we generate 1000 random, synthetic data sets by redistributing the data within the x - and y -axis error bars. PCA is applied at each iteration, the slope uncertainty is calculated as standard deviation of the resulting slope distributions.

APPENDIX D: EXTRAPOLATION AND DECONVOLUTION EFFECTS ON CLOUD PROPERTIES

In this paper, we have applied extrapolation and deconvolution to correct the measured cloud properties. Nevertheless, this paradigm has been mostly used for nearby galaxy studies and only rarely in the Milky Way (e.g., Blitz & Thaddeus 1980). Here we explore the

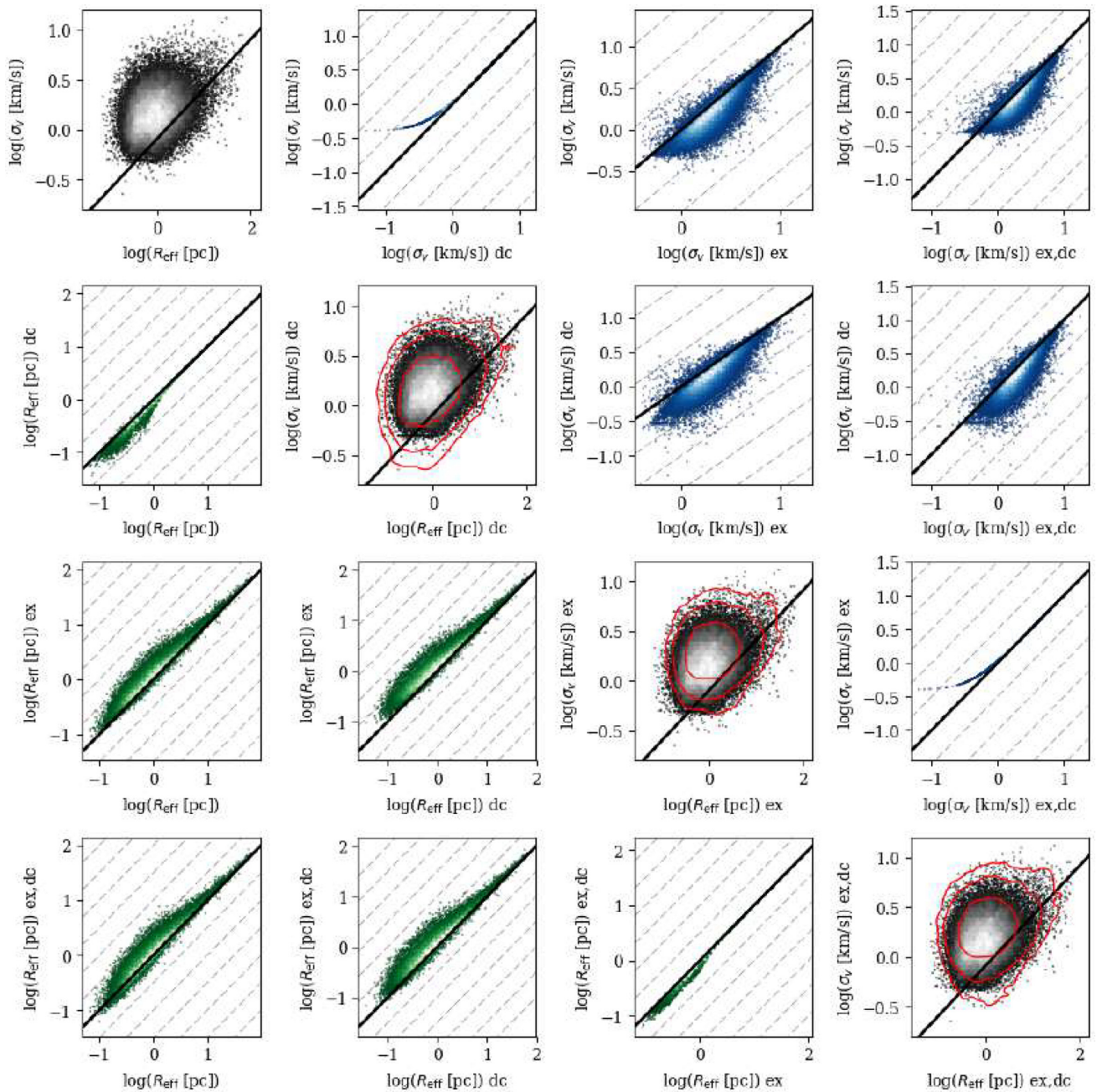


Figure D1. Comparison between extrapolated ('ex'), deconvolved ('dc'), extrapolated and deconvolved ('ex.dc'), and uncorrected R_{eff} and σ_v (no suffix), and their effects on the size–linewidth relation (along the figure main diagonal). Across the main diagonal the black full line indicates equation (21), while in the panels off diagonal the 1:1 relation. Across the main diagonal uncorrected properties are always indicated with grey markers, while the red markers indicate the corrected properties. Dashed grey lines are spaced of 0.5 dex.

differences between properties directly calculated by ASTRODENDRO and those corrected for instrumental biases.

Fig. D1 summarizes the differences between the various approaches for calculating the cloud properties in term of R_{eff} and σ_v . Generally, the deconvolution effectively reduces the cloud velocity dispersion starting from both the uncorrected and extrapolated σ_v . Large values of σ_v are not influenced by this correction. In contrast, the extrapolation almost always shifts the velocity dispersion toward values ~ 0.5 dex higher with respect to the uncorrected velocity dispersion. The effect is similar for the cloud size. Nevertheless, extrapolated and deconvolved R_{eff} appears more equally distributed

across the 1:1 line with respect to uncorrected and deconvolved only effective radii. The deconvolution alone does not significantly change the form of the size–linewidth relation we measured in the paper, but adds several points at low σ_v . The same is true for the extrapolation and deconvolution combined.

The extrapolation produces a more significant boost of the properties (up to 1 dex) when it is applied to the CO flux. However, this boost closes the gap between the emission in the catalogued structure and the total flux obtained by averaging over the entire data cube. Fig. D2 (left-hand panel) shows that extrapolated cloud

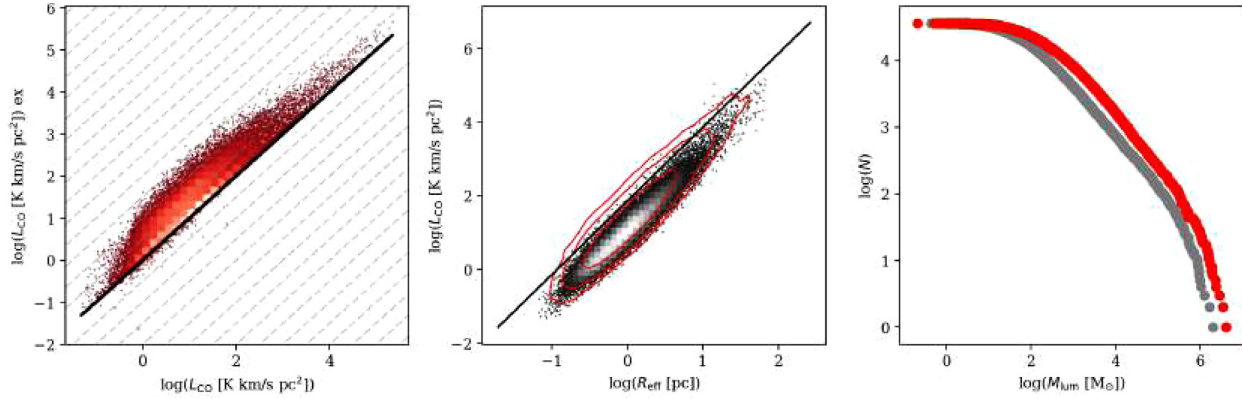


Figure D2. Comparison between extrapolated ('ex') and uncorrected (no suffix) CO luminosity (left). Effect of luminosity extrapolation on the $L_{\text{CO}}-R_{\text{eff}}$ relation (middle) and the cumulative mass spectrum (right). In the left-hand panel the black line represents the 1:1 relation while in the middle one, the locus defined by a constant $\Sigma_{\text{mol}} = 200 M_{\odot} \text{pc}^{-2}$ and $\alpha_{\text{CO}} = 8.7 M_{\odot} (\text{K km s}^{-1} \text{pc}^2)^{-1}$. In the last two panels, corrected properties (extrapolated for L_{CO} , extrapolated and deconvolved for R_{eff}) are indicated with red contours and markers, while uncorrected ones with grey markers. Dashed grey lines in the first panel are spaced of 0.5 dex.

L_{CO} can be even 1 dex higher than uncorrected CO luminosity. This effect is small for clouds with large L_{CO} values in the catalogue.

The instrumental correction on both L_{CO} and R_{eff} provides a relationship between the two quantities that is closer to the theoretical model described by constants $\Sigma_{\text{mol}} = 200 M_{\odot} \text{pc}^{-2}$ and $\alpha_{\text{CO}} = 8.7 M_{\odot} (\text{K km s}^{-1} \text{pc}^2)^{-1}$ (Fig. D2, middle). At the same time, the shape of the mass spectrum (Fig. D2, right) from uncorrected and extrapolated values of M_{lum} is similar. The fit of the mass spectrum from the uncorrected M_{lum} has a spectral index $\gamma = -1.7$, consistent with the one measured from the extrapolated mass reported in the Section 8.2. Nevertheless, the truncation mass given by the uncorrected property spectra is a factor 2 lower $M_0 \sim 1.7 \times 10^6 M_{\odot}$.

In conclusion, we regard these corrections for emission at low significance and for the effects of instrumental resolution as significant. While the models used for correction cannot completely overcome the limitations of the observations, the combination of corrections adopted and the uncertainties established by the bootstrapping better describe the molecular gas population compared to just using the emission in the catalogued objects alone.

APPENDIX E: COHRS IN MOLECULAR CLOUDS

Here we collect the maps of the sub-cubes from the full COHRS data overlaid with the identified cloud mask. Galactic longitude - latitude maps are shown in Fig. E1-E15, while Galactic longitude - velocity maps are reported in Fig. E16-E30.

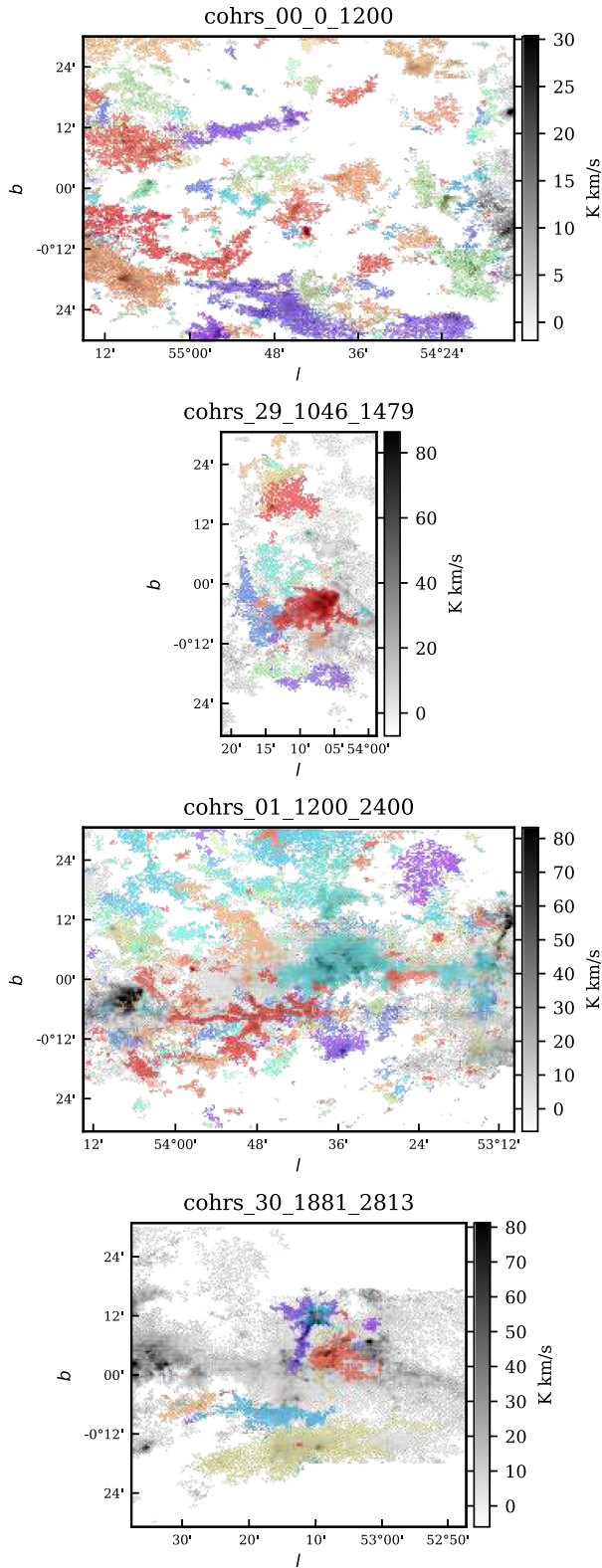


Figure E1. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

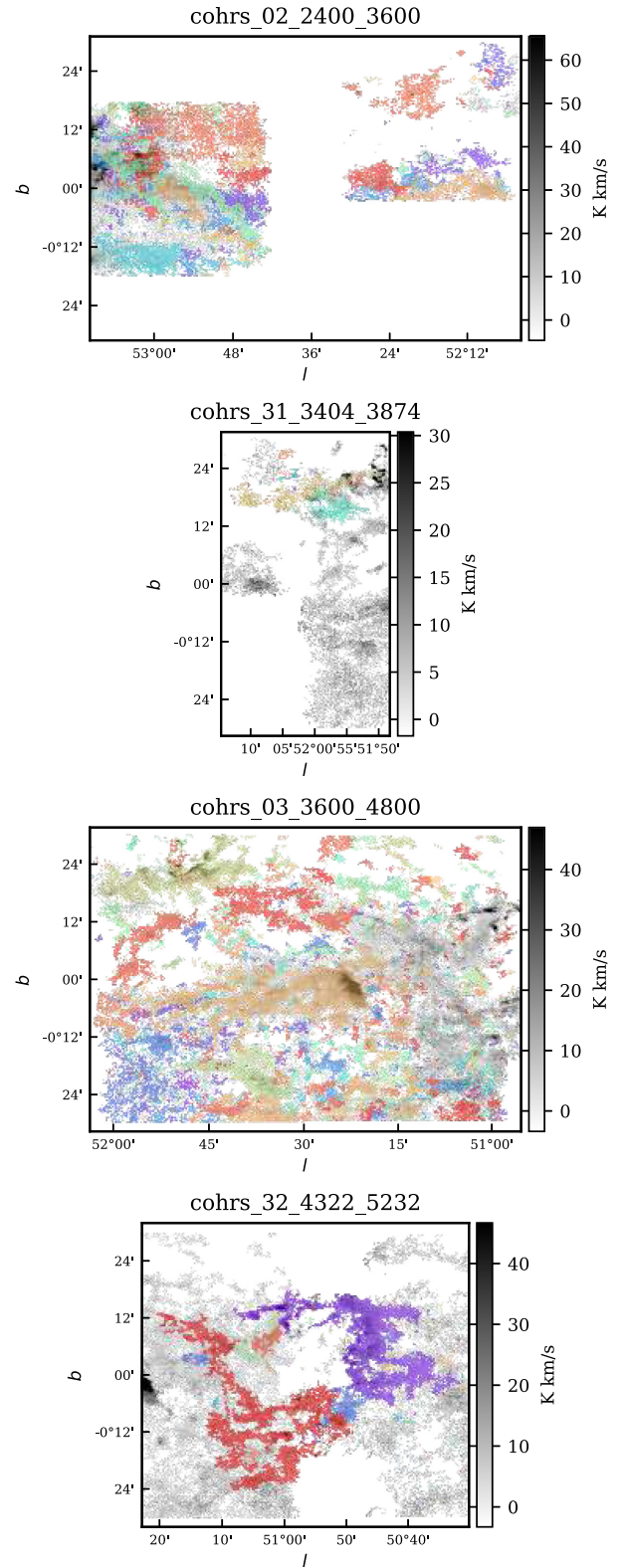


Figure E2. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

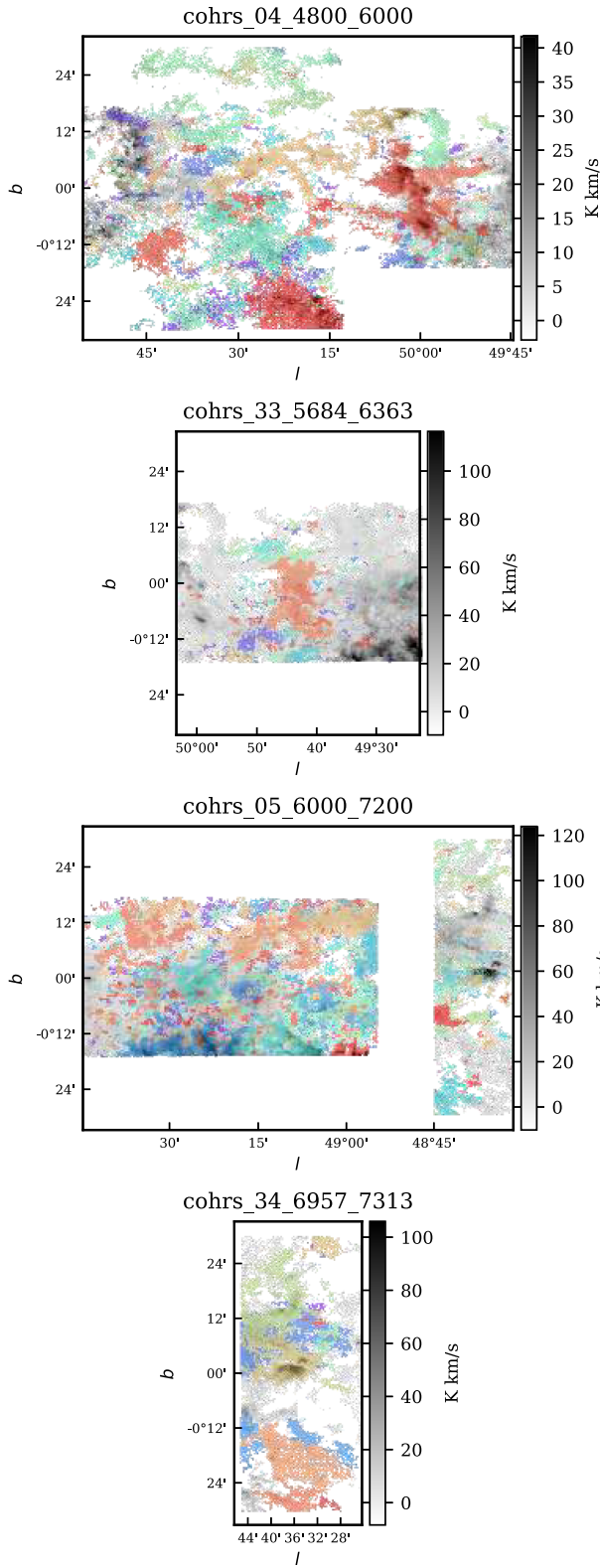


Figure E3. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

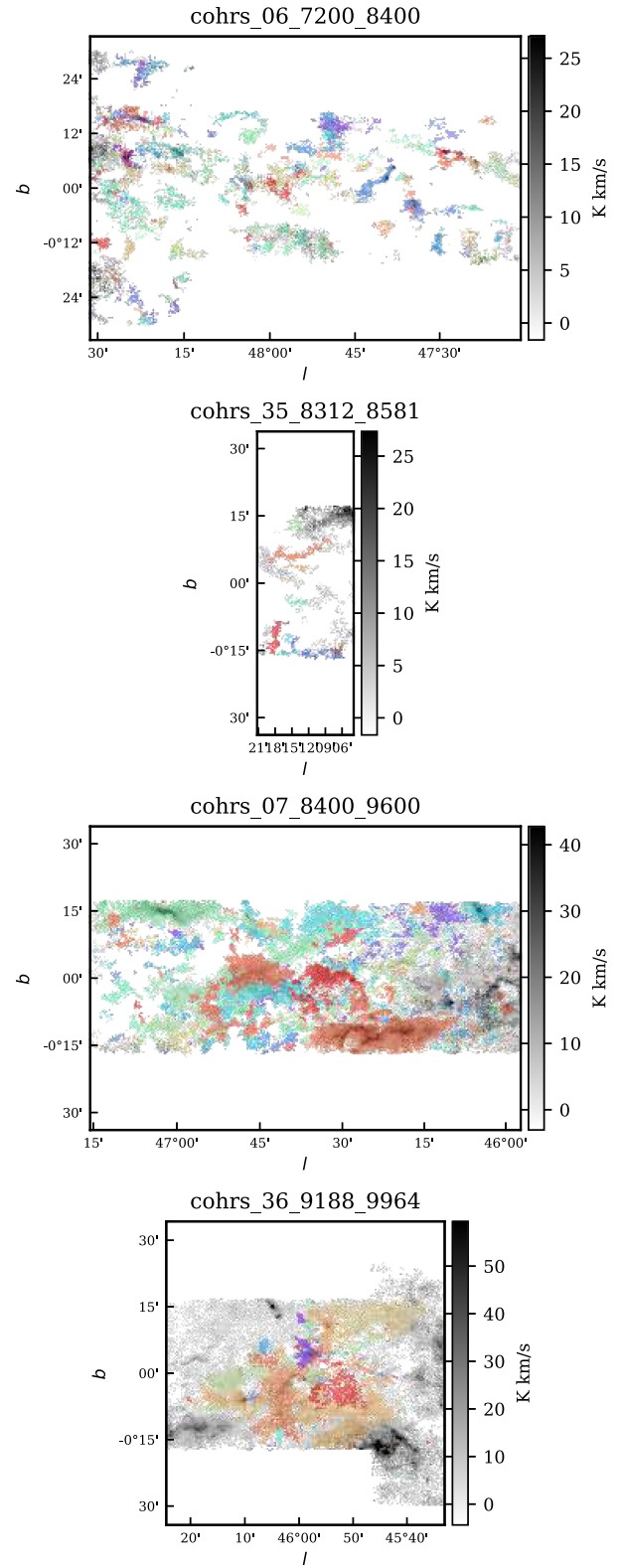


Figure E4. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

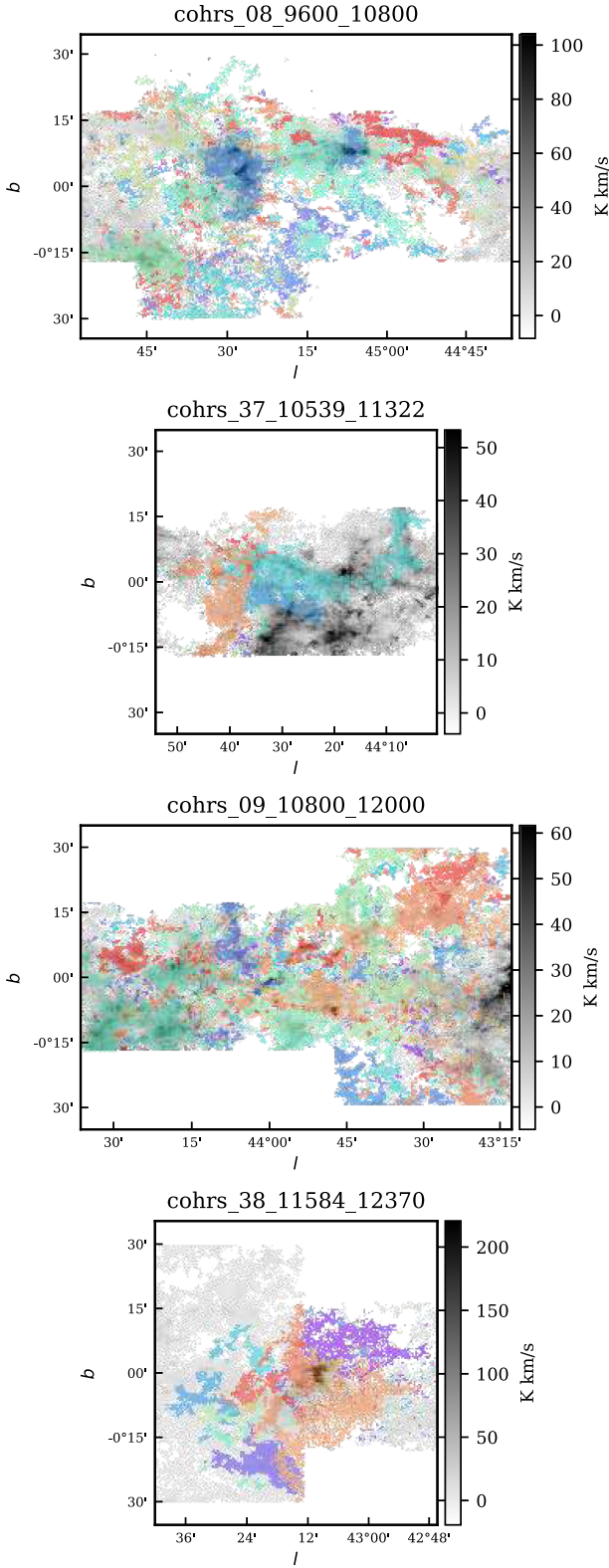


Figure E5. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

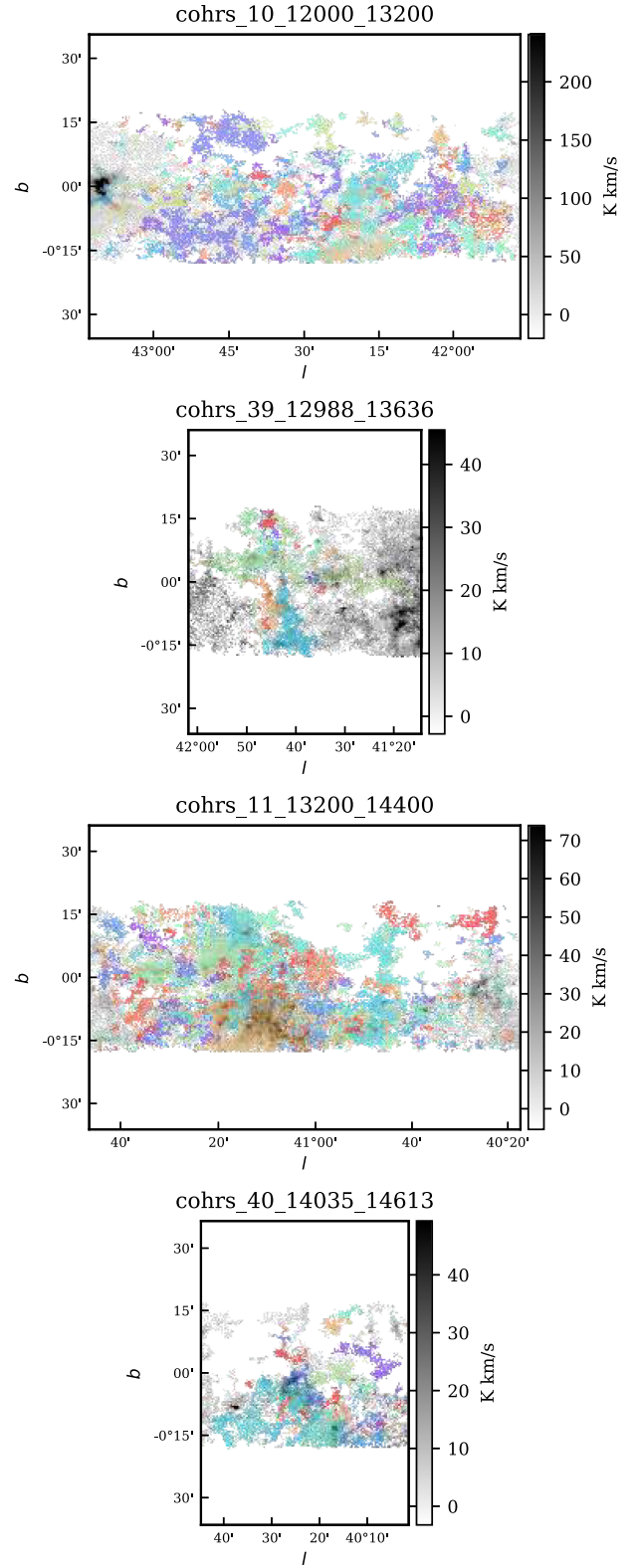


Figure E6. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

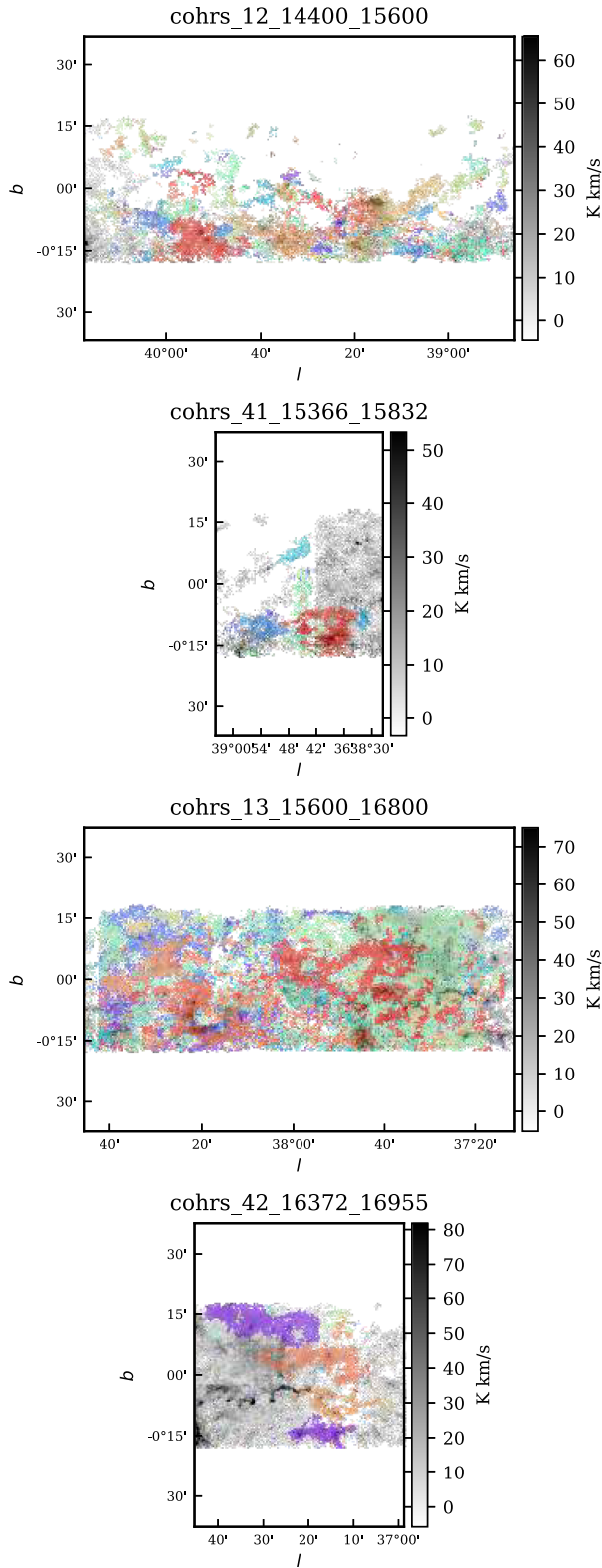


Figure E7. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

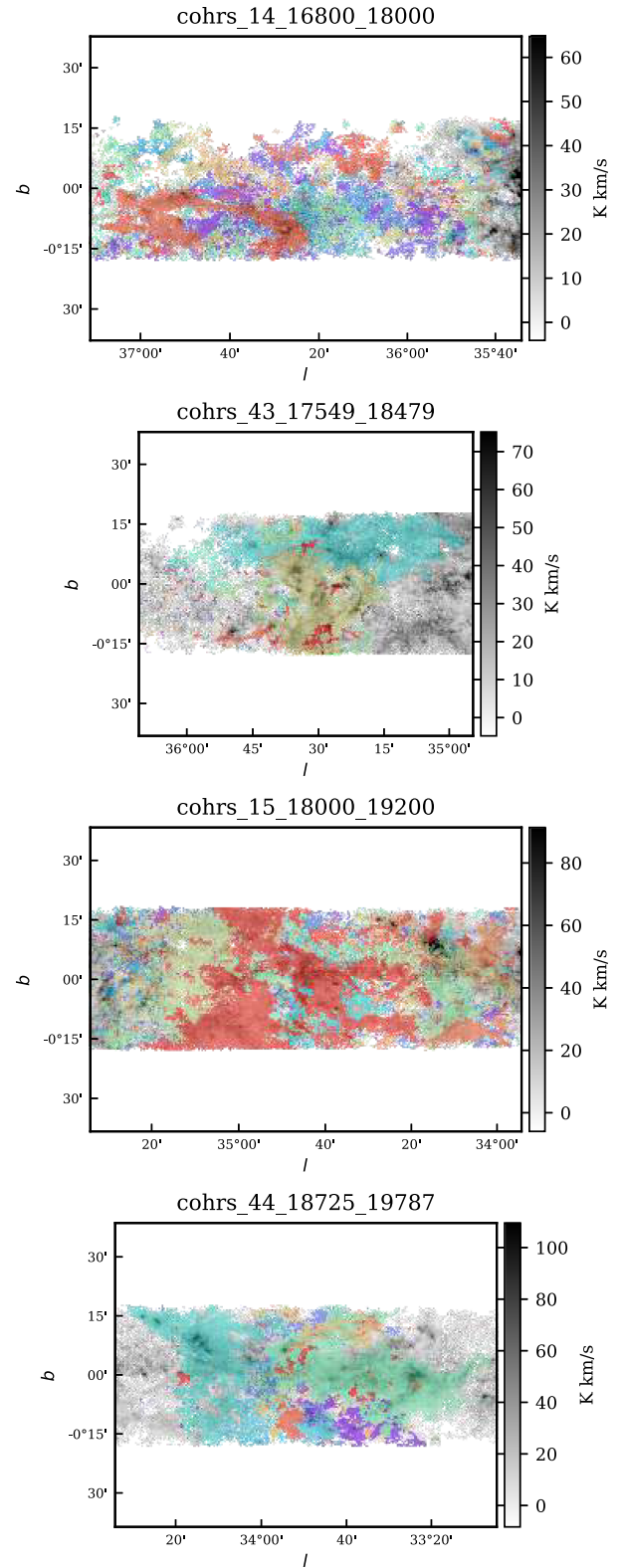


Figure E8. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

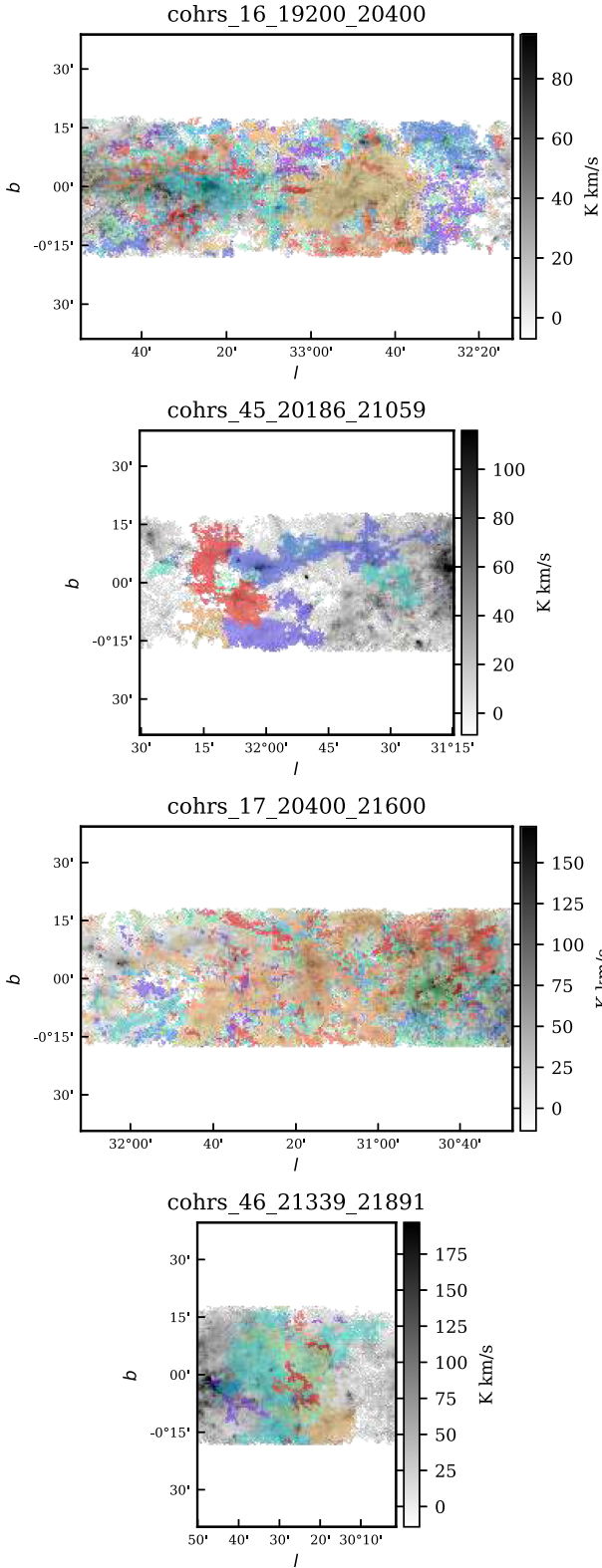


Figure E9. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

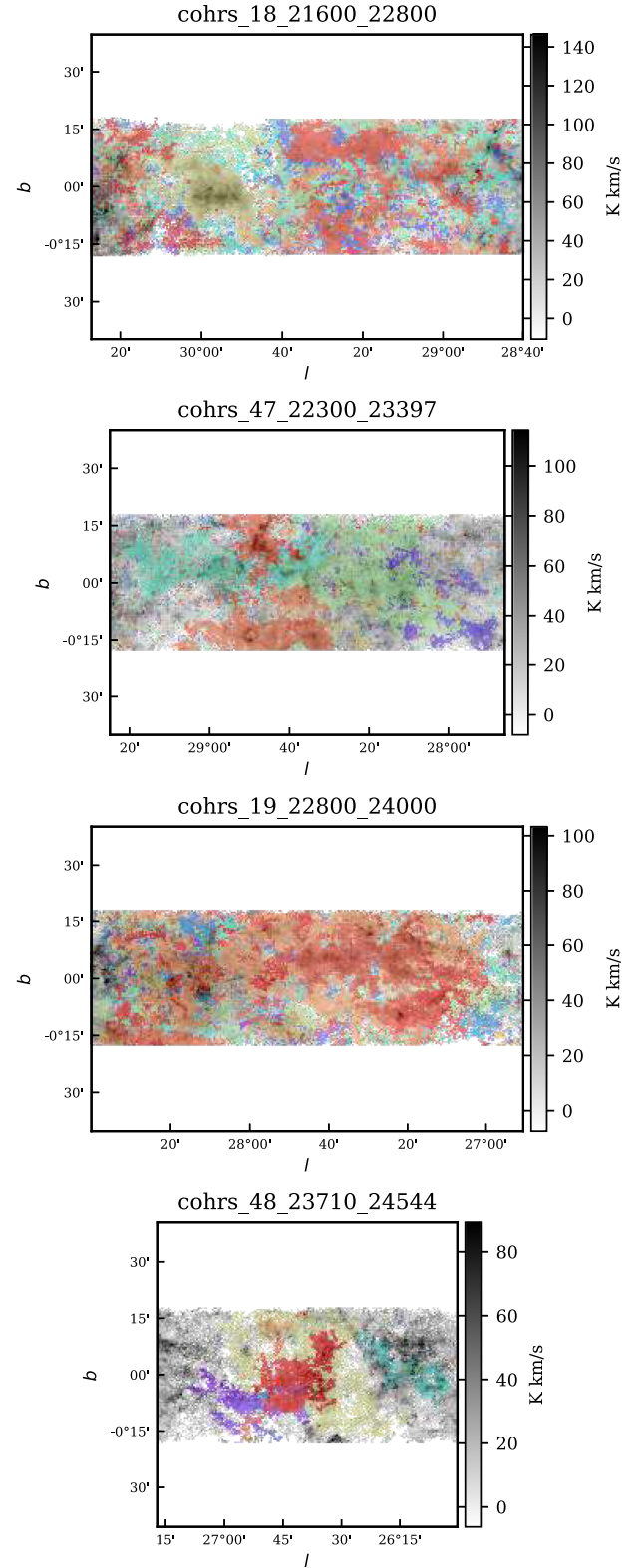


Figure E10. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

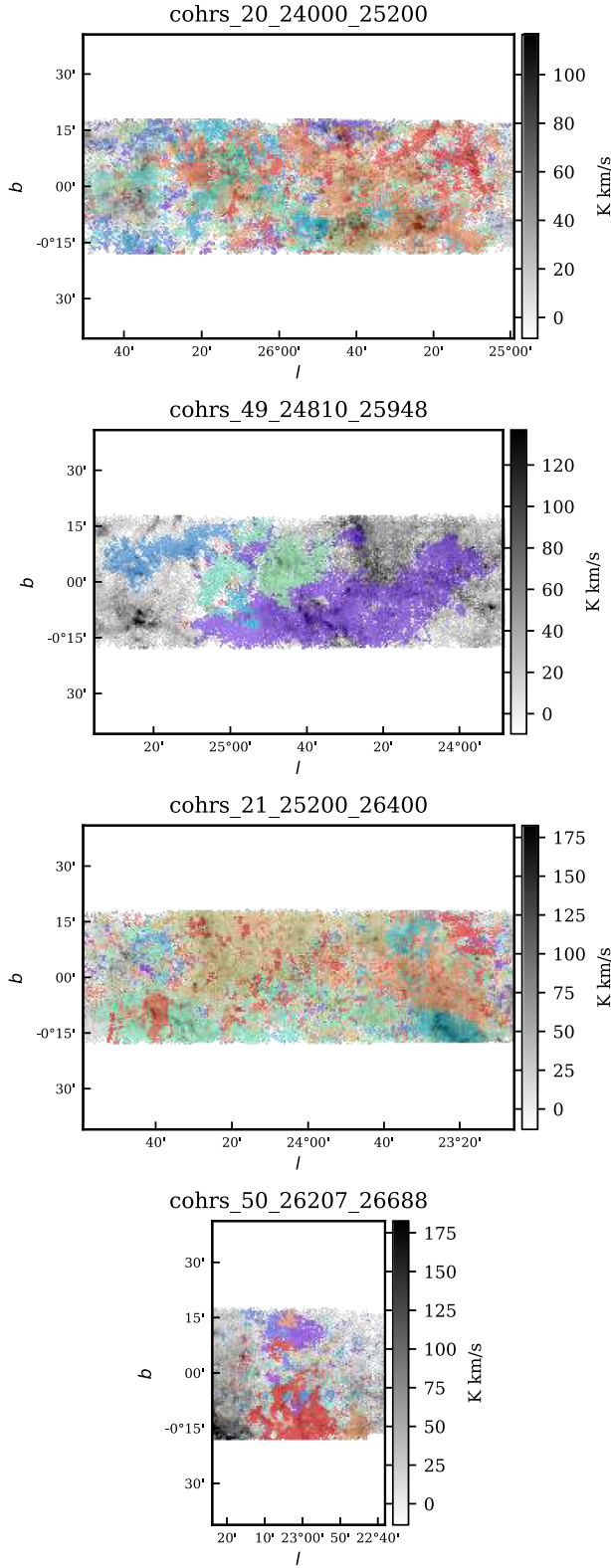


Figure E11. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

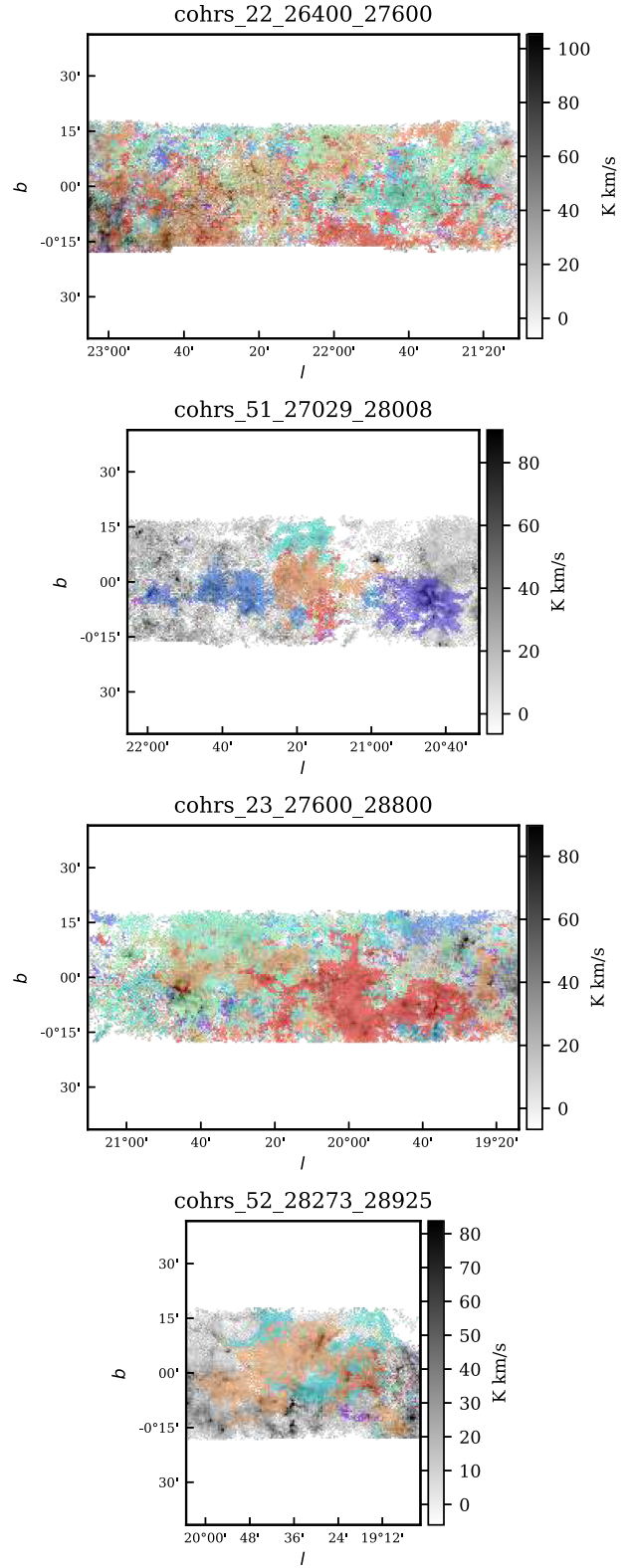


Figure E12. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

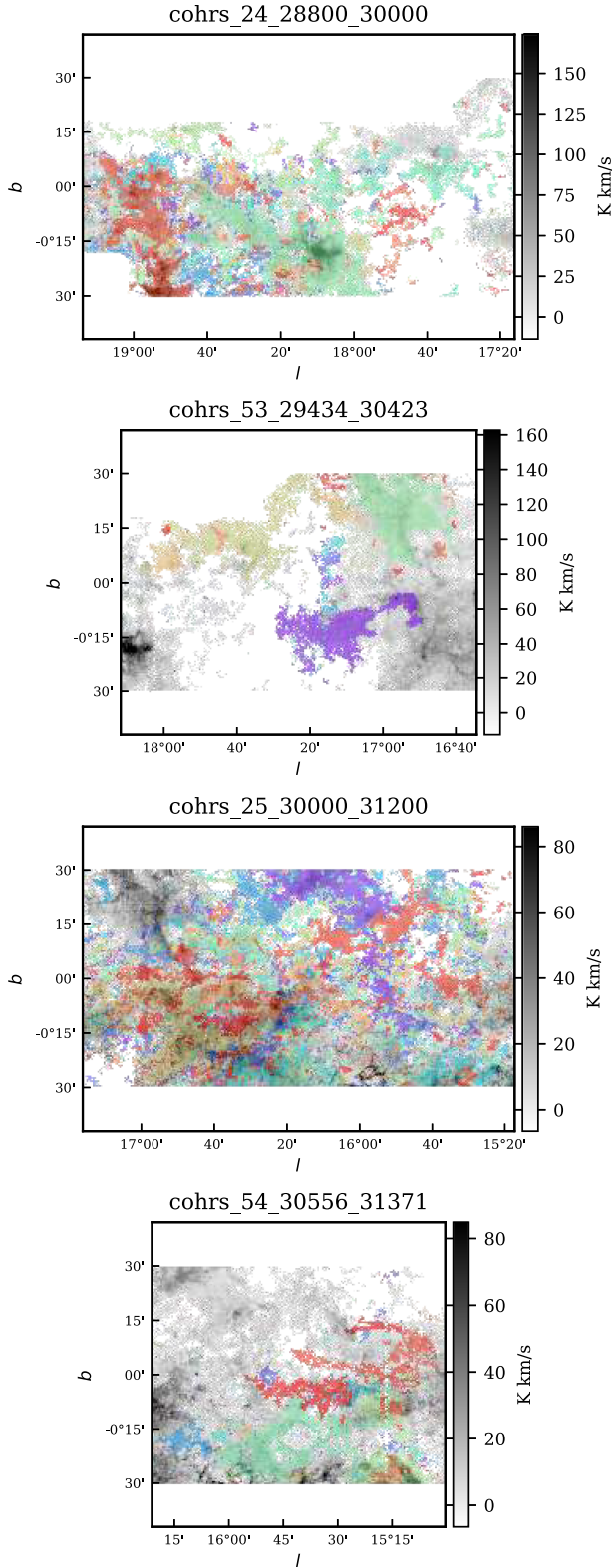


Figure E13. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

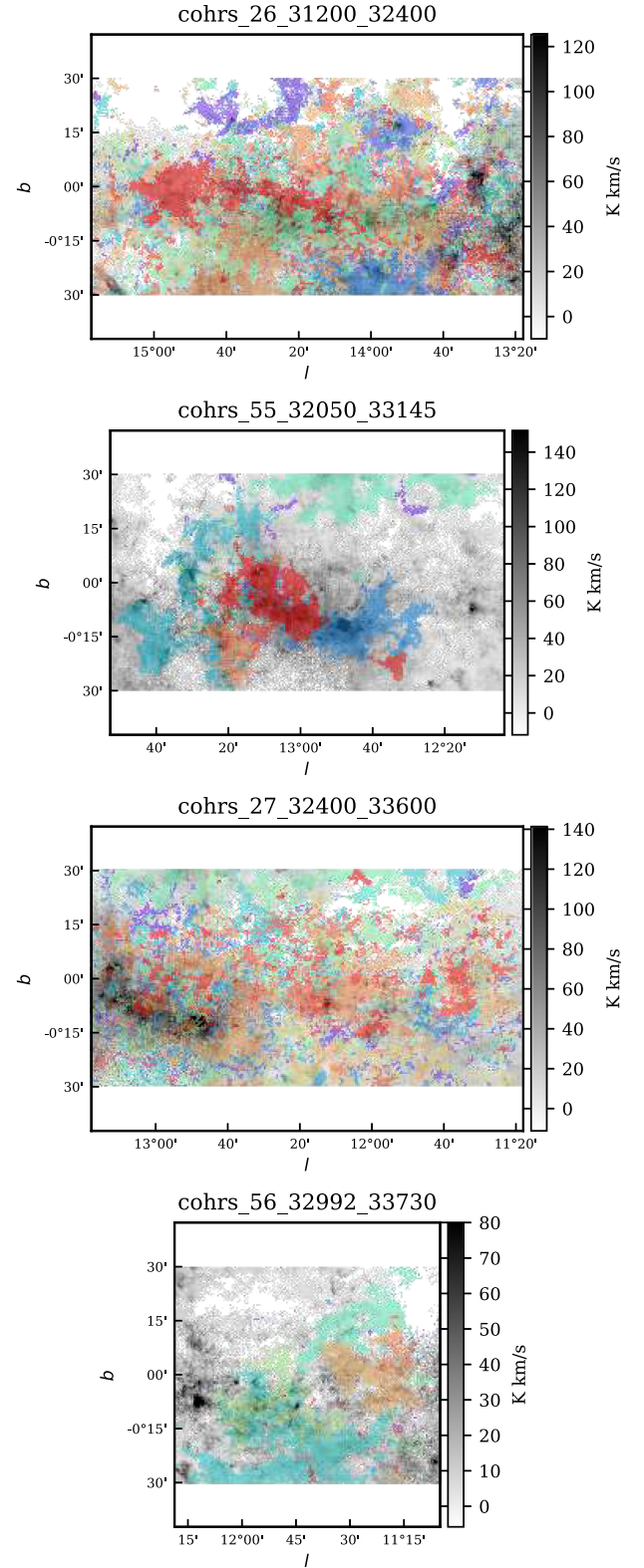


Figure E14. Longitude–latitude integrated intensity maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

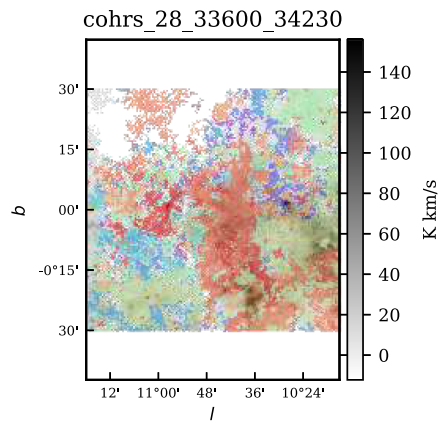


Figure E15. Longitude–latitude integrated intensity map of COHRS given sub-cube masked as explained in Section 4. In colour the identified clouds are indicated.

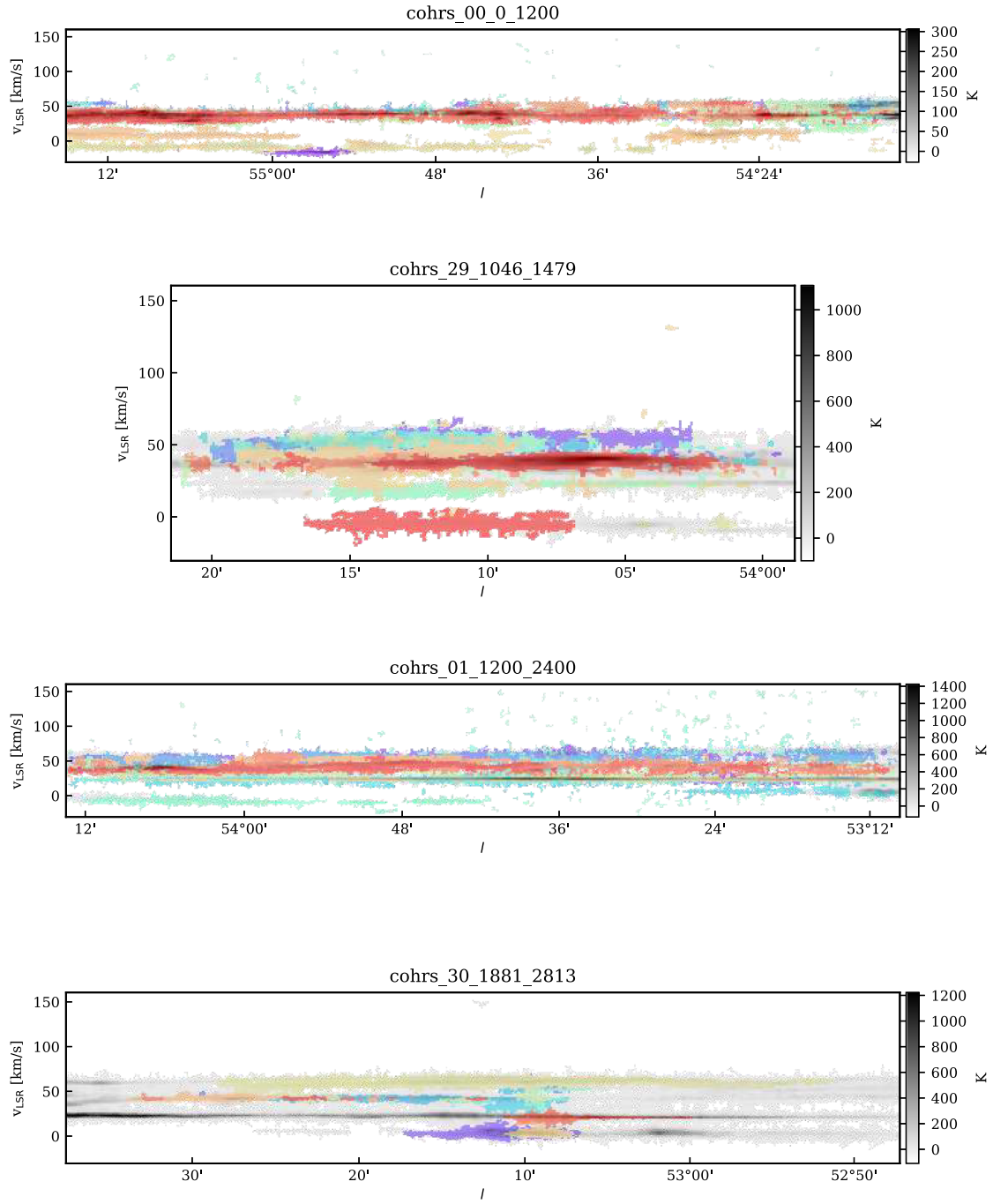


Figure E16. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

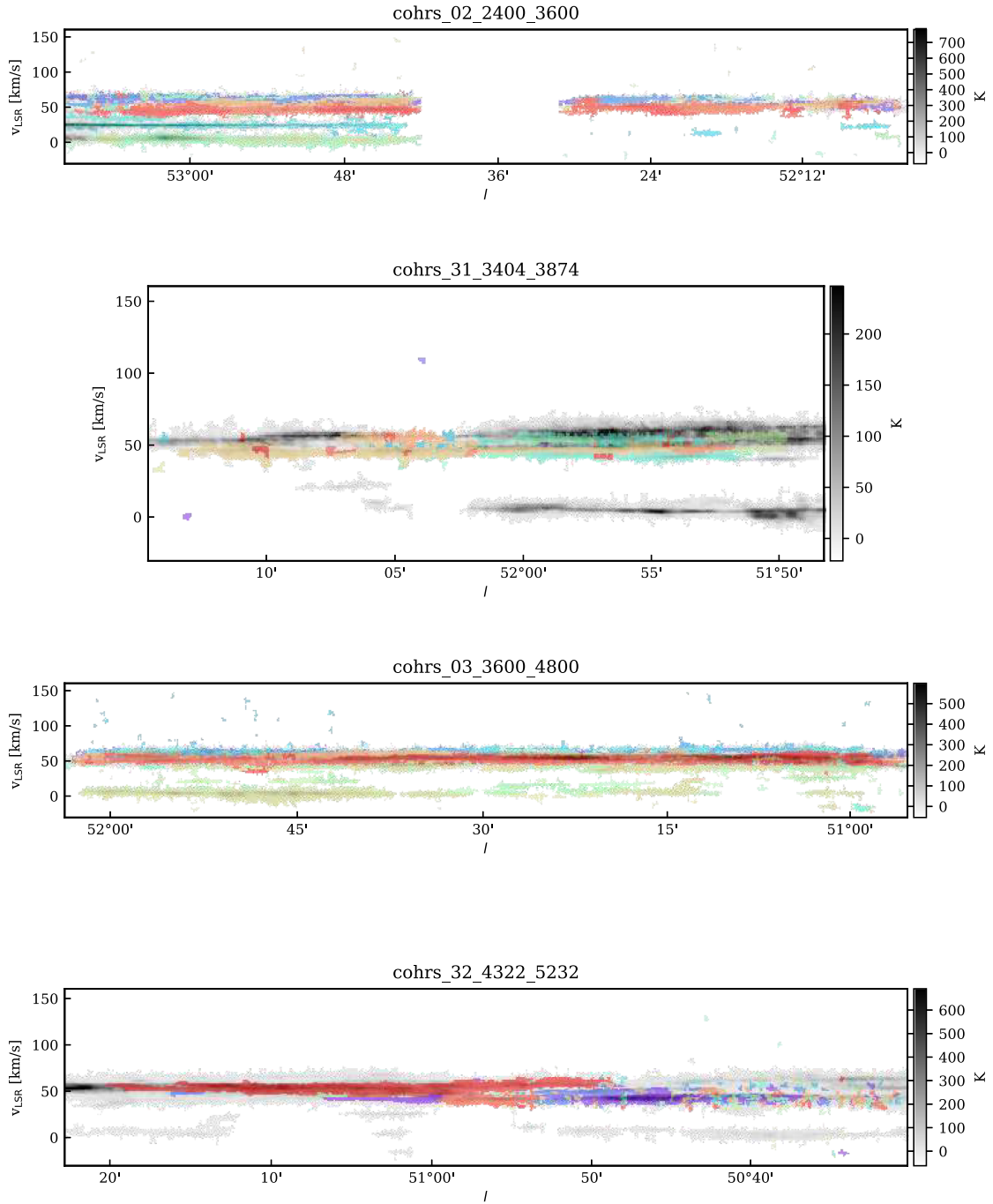


Figure E17. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

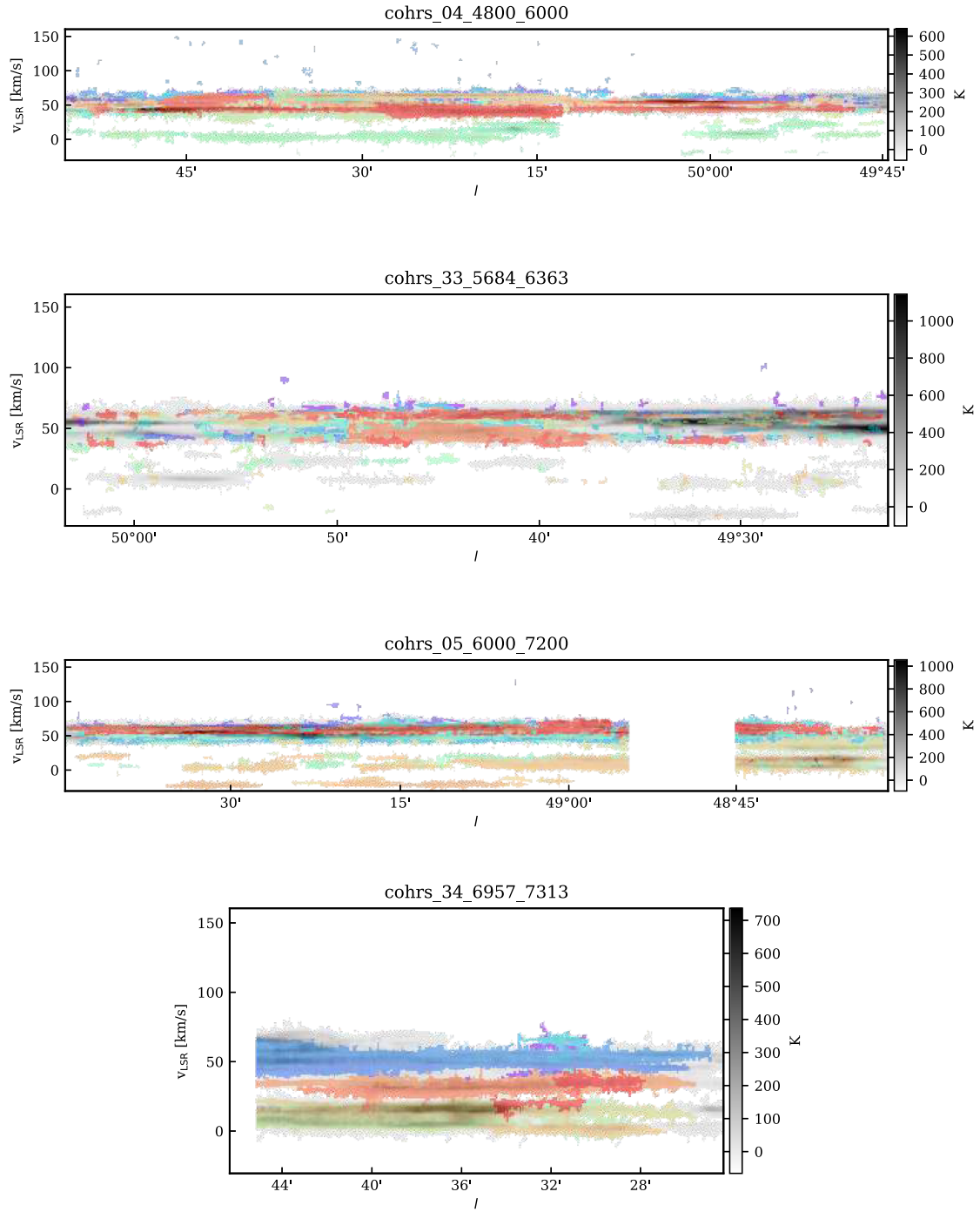


Figure E18. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

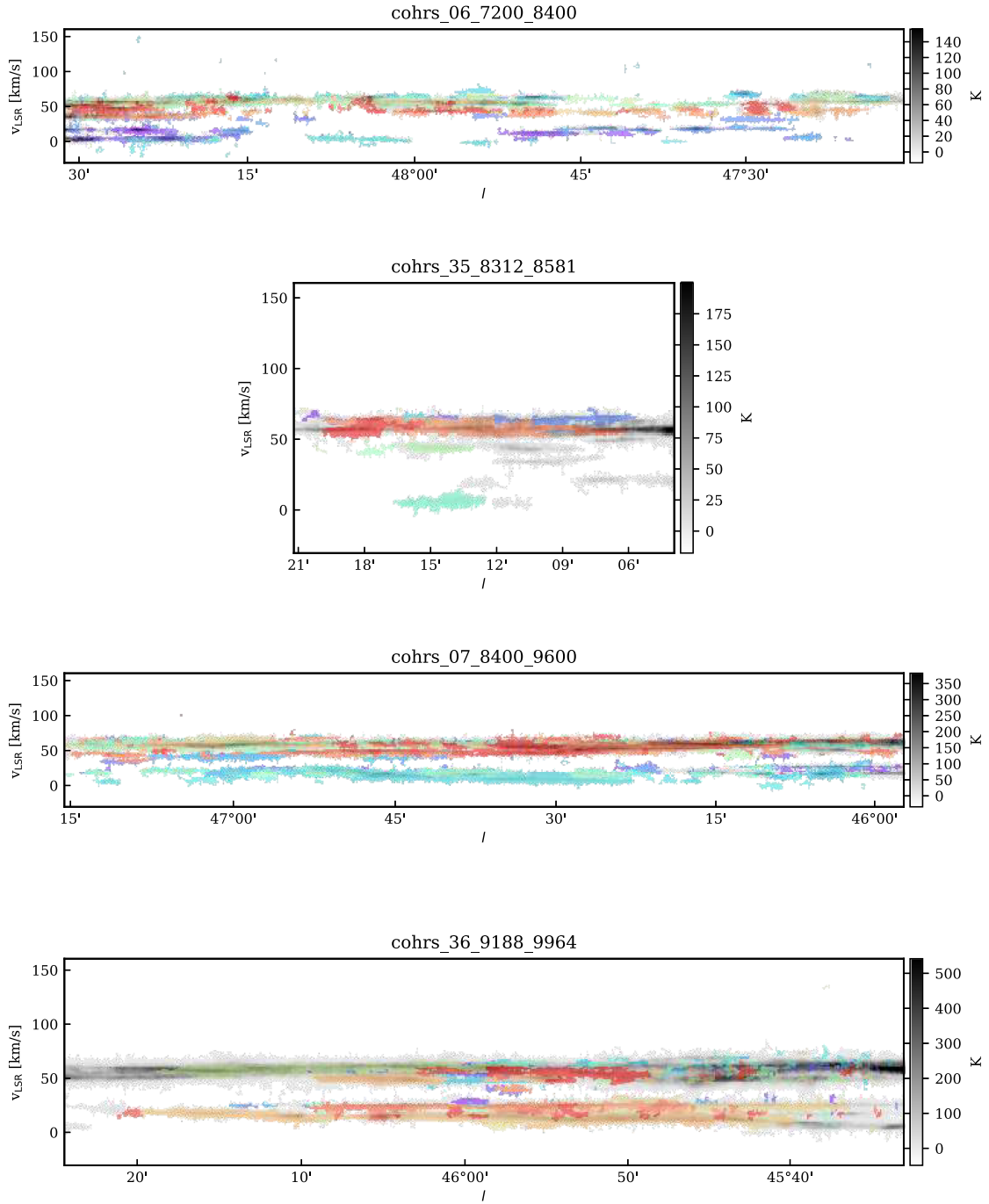


Figure E19. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

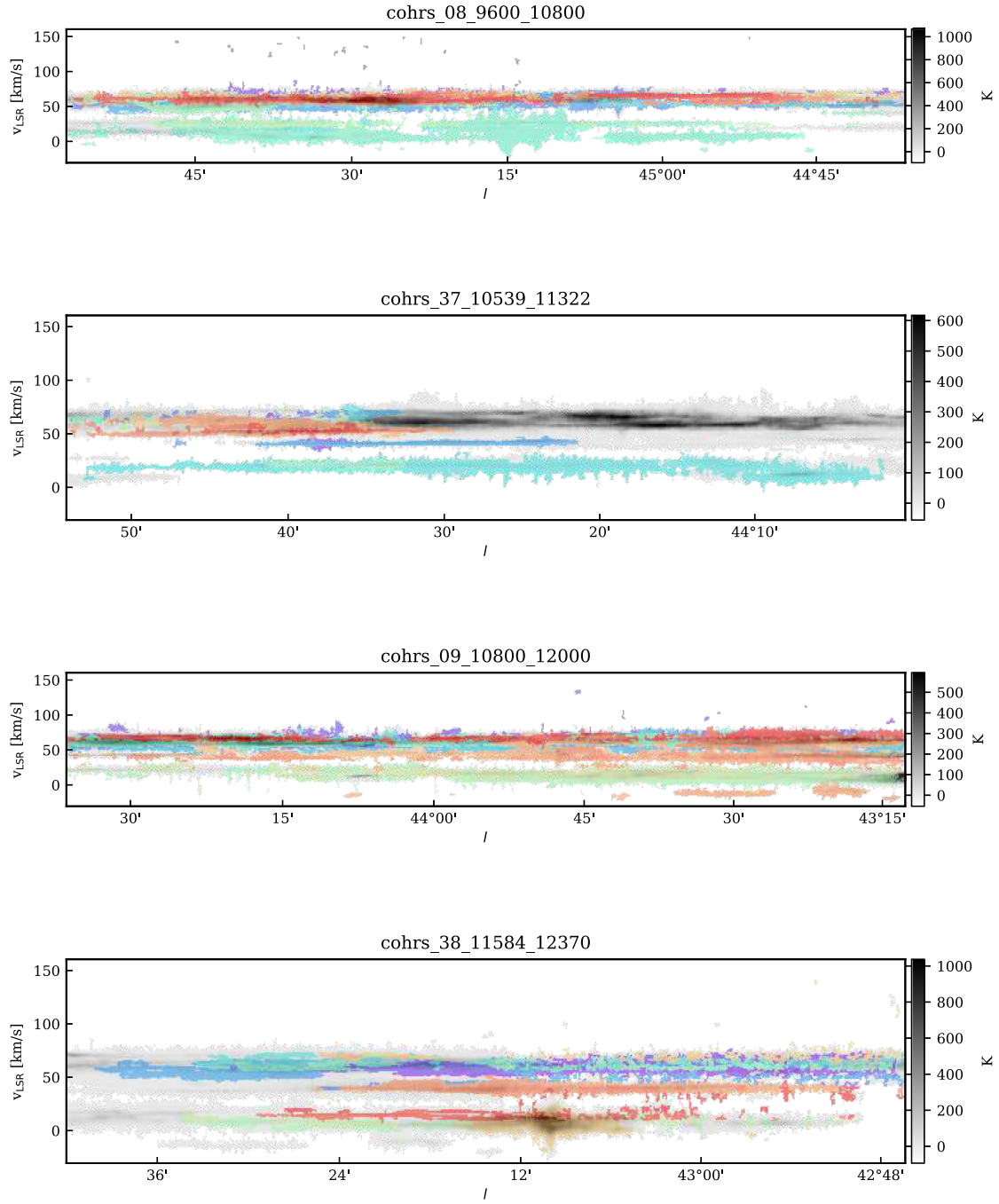


Figure E20. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

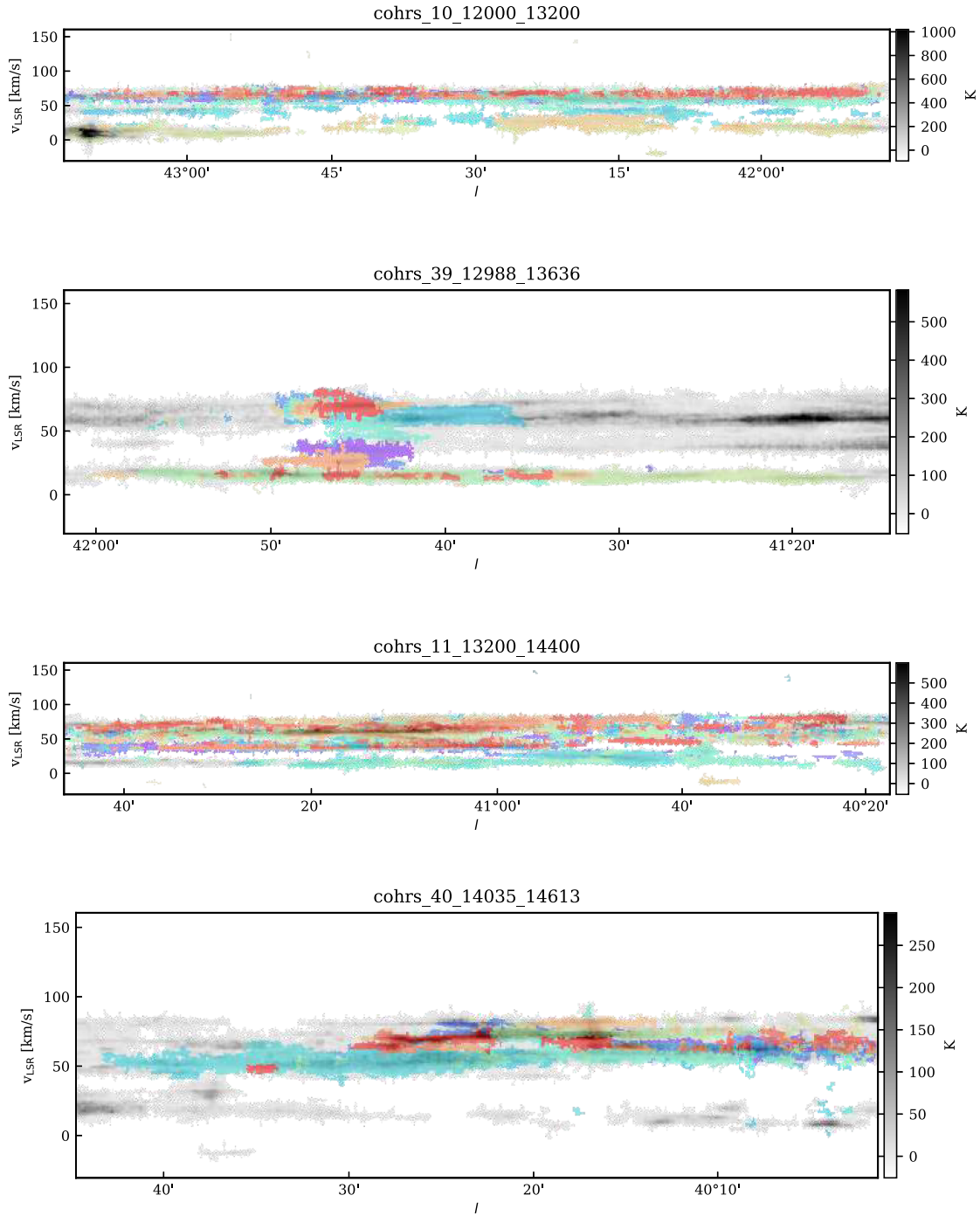


Figure E21. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

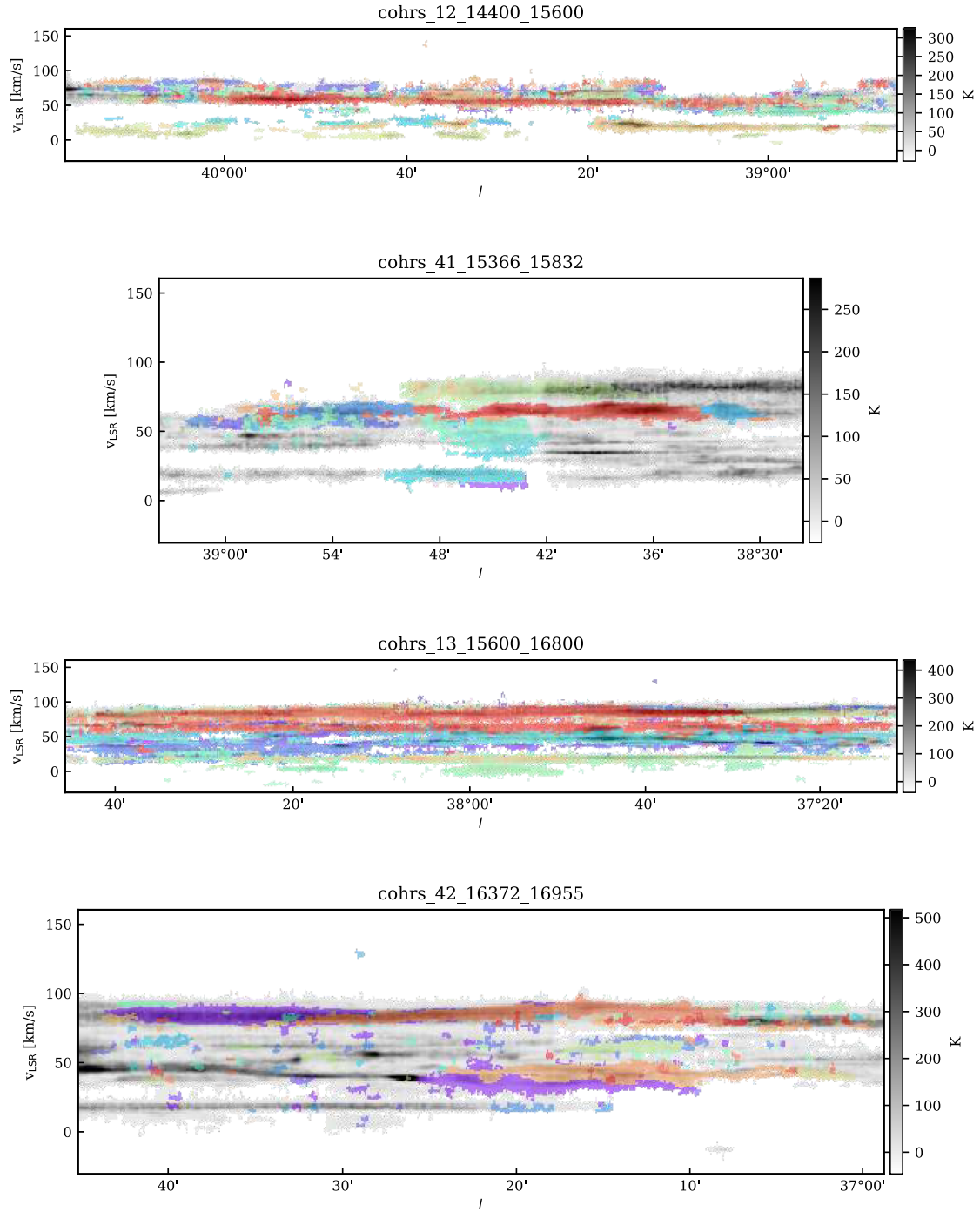


Figure E22. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

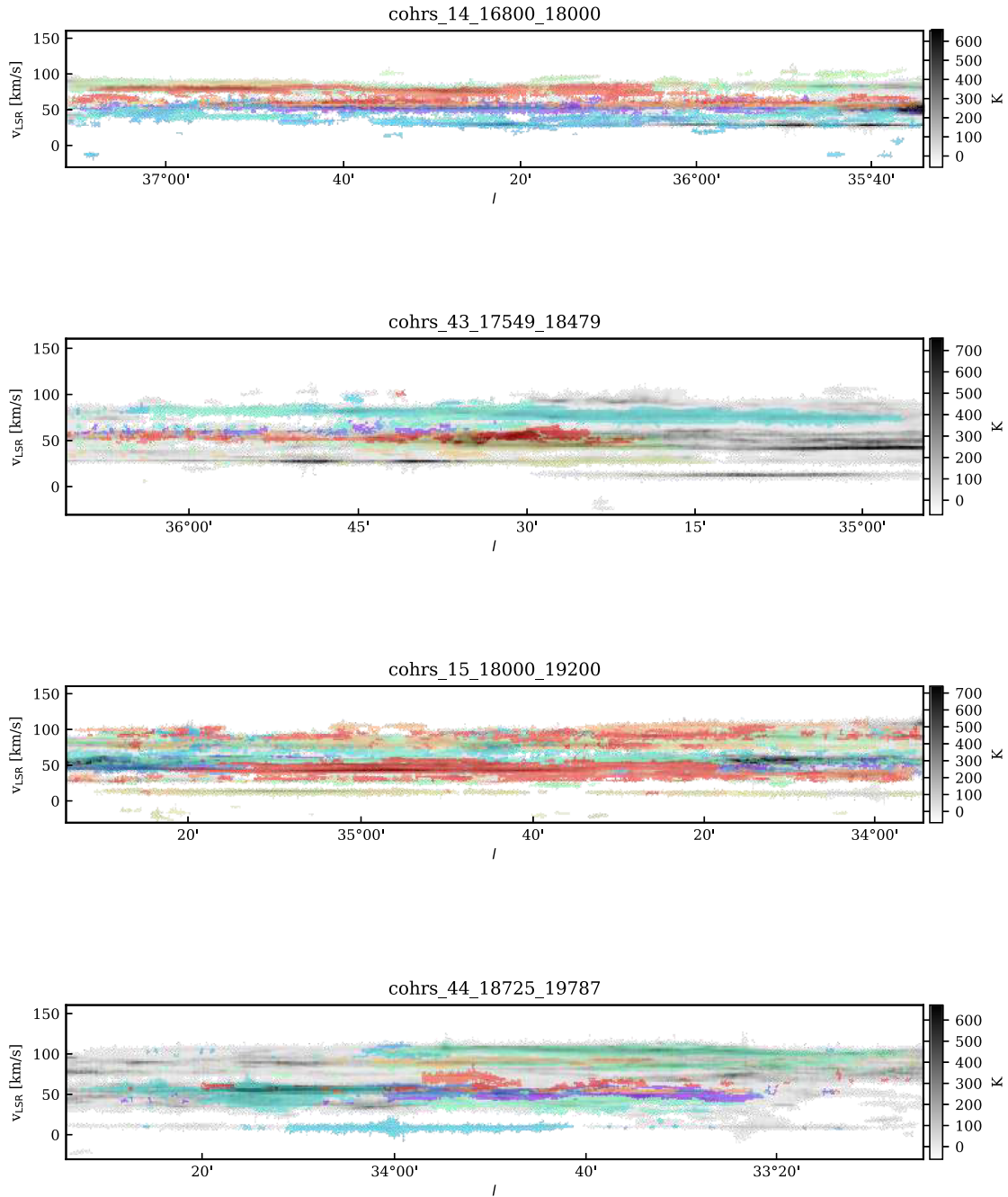


Figure E23. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

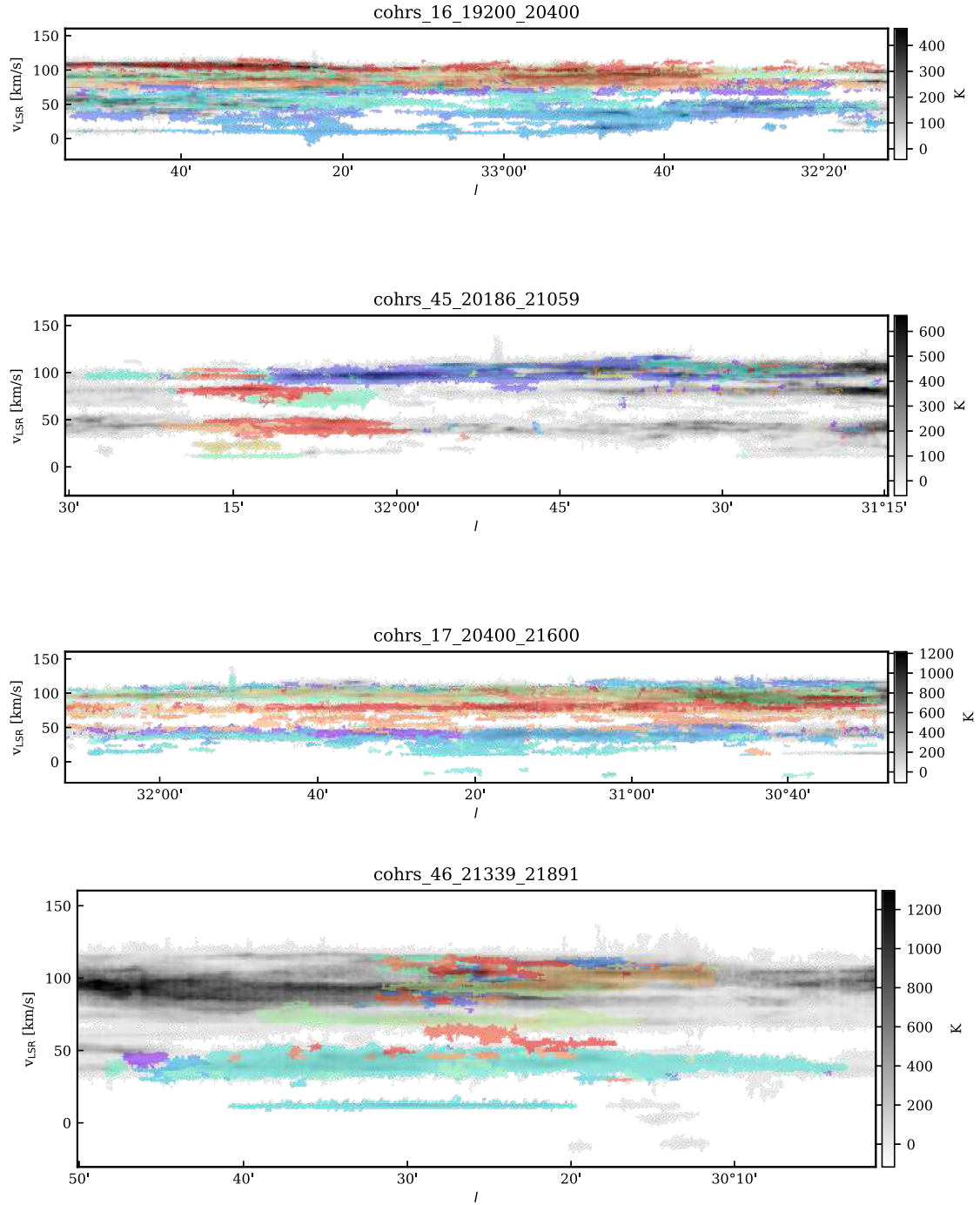


Figure E24. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

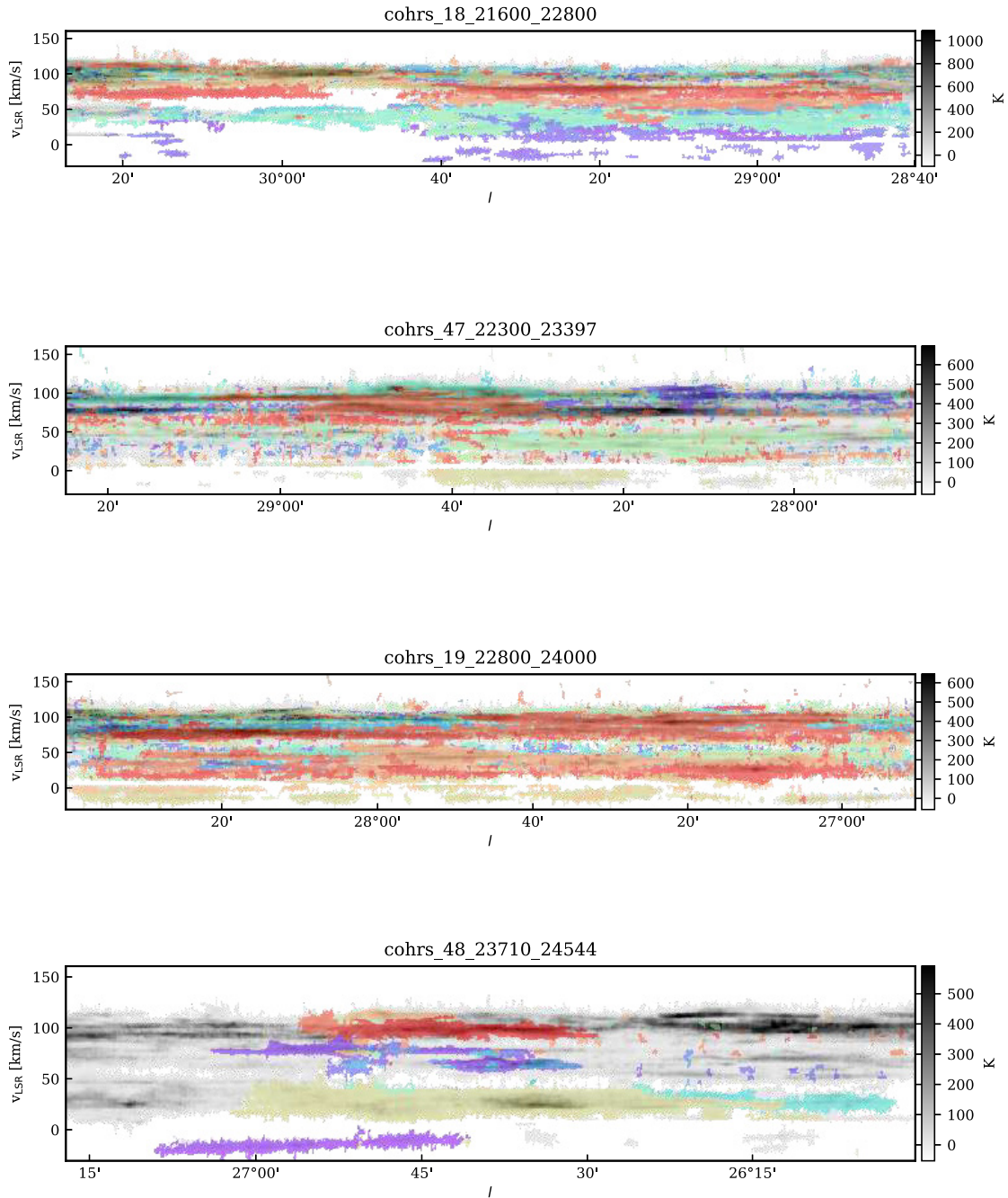


Figure E25. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

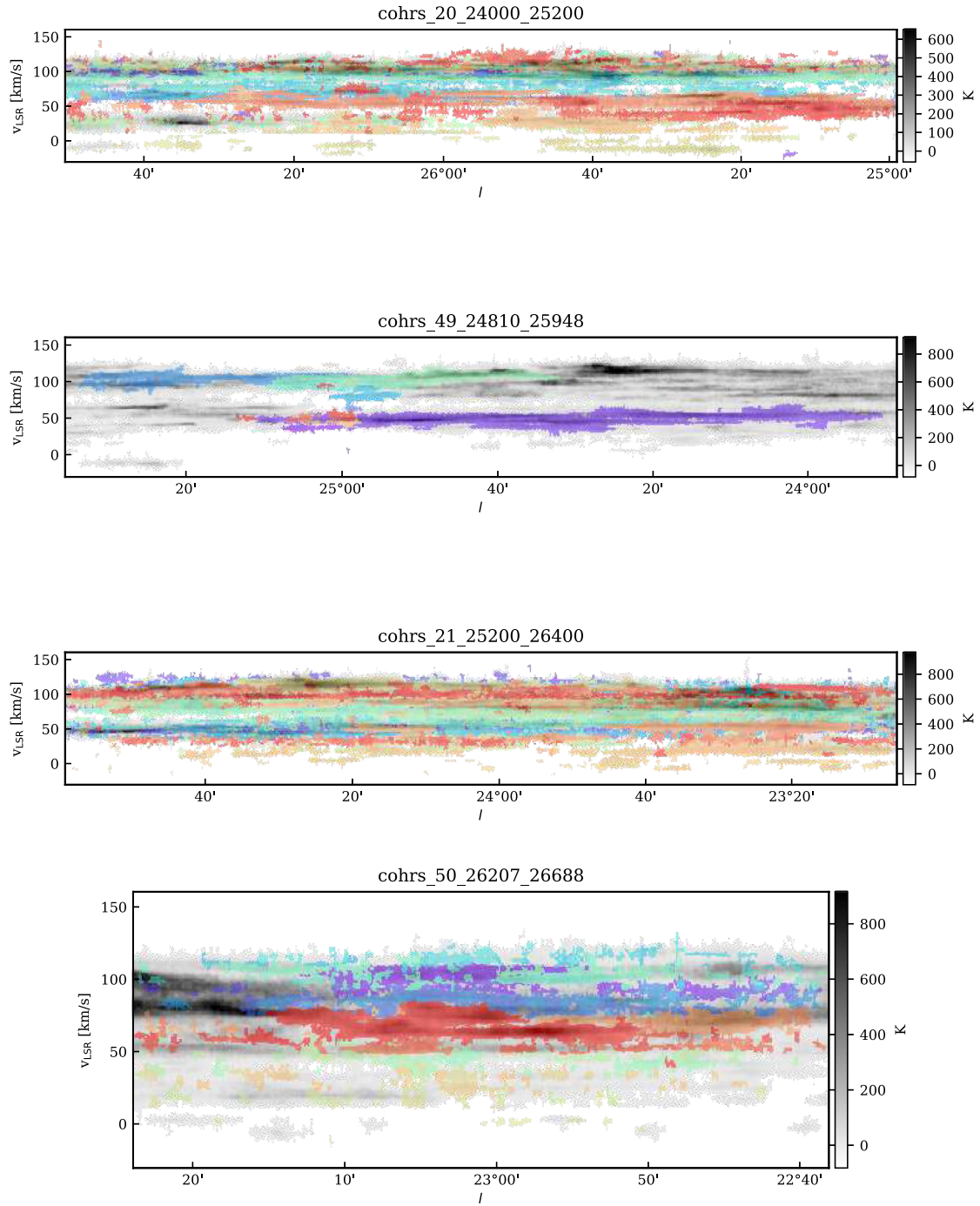


Figure E26. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

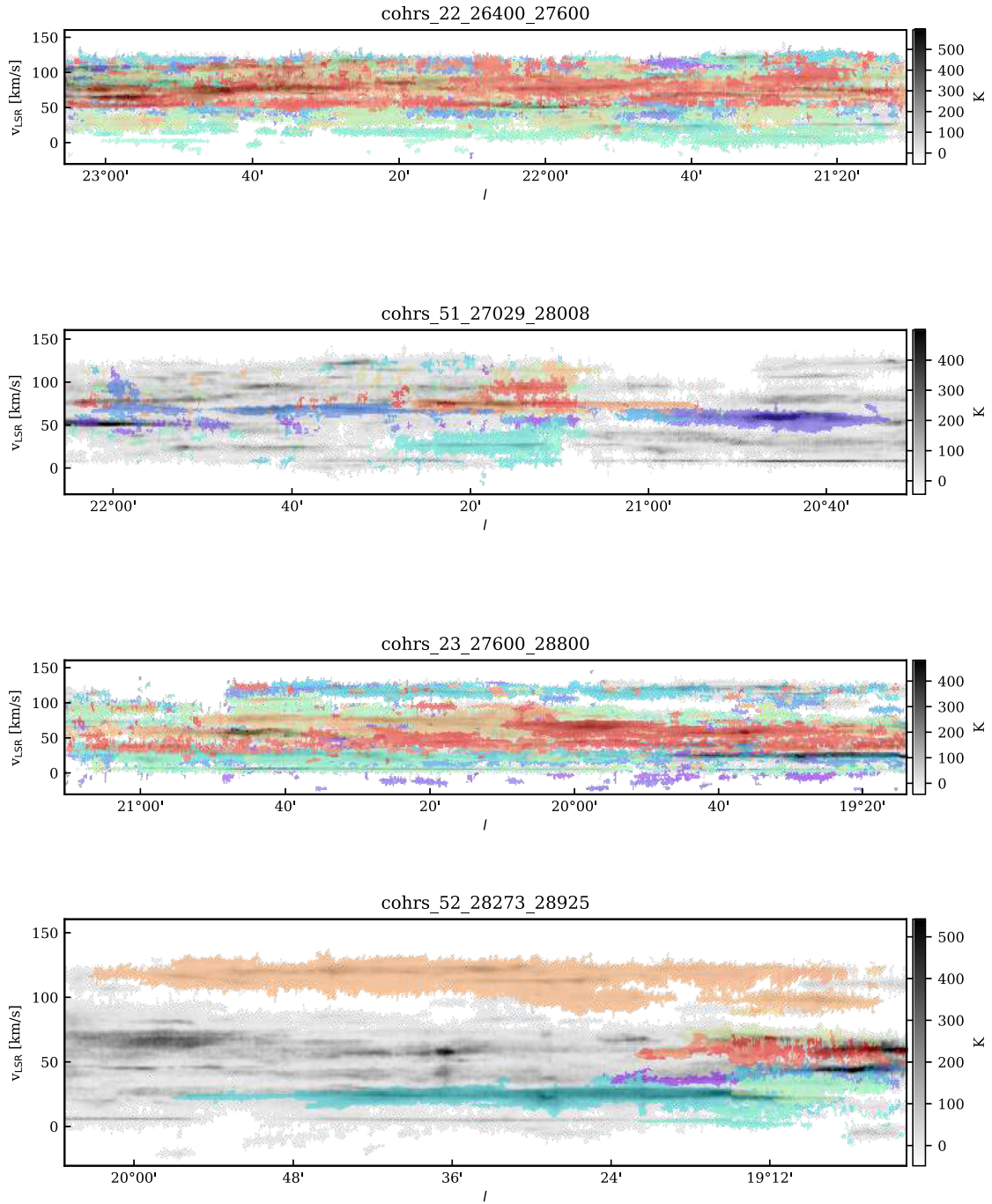


Figure E27. Longitude–velocity integrated maps of COHERS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

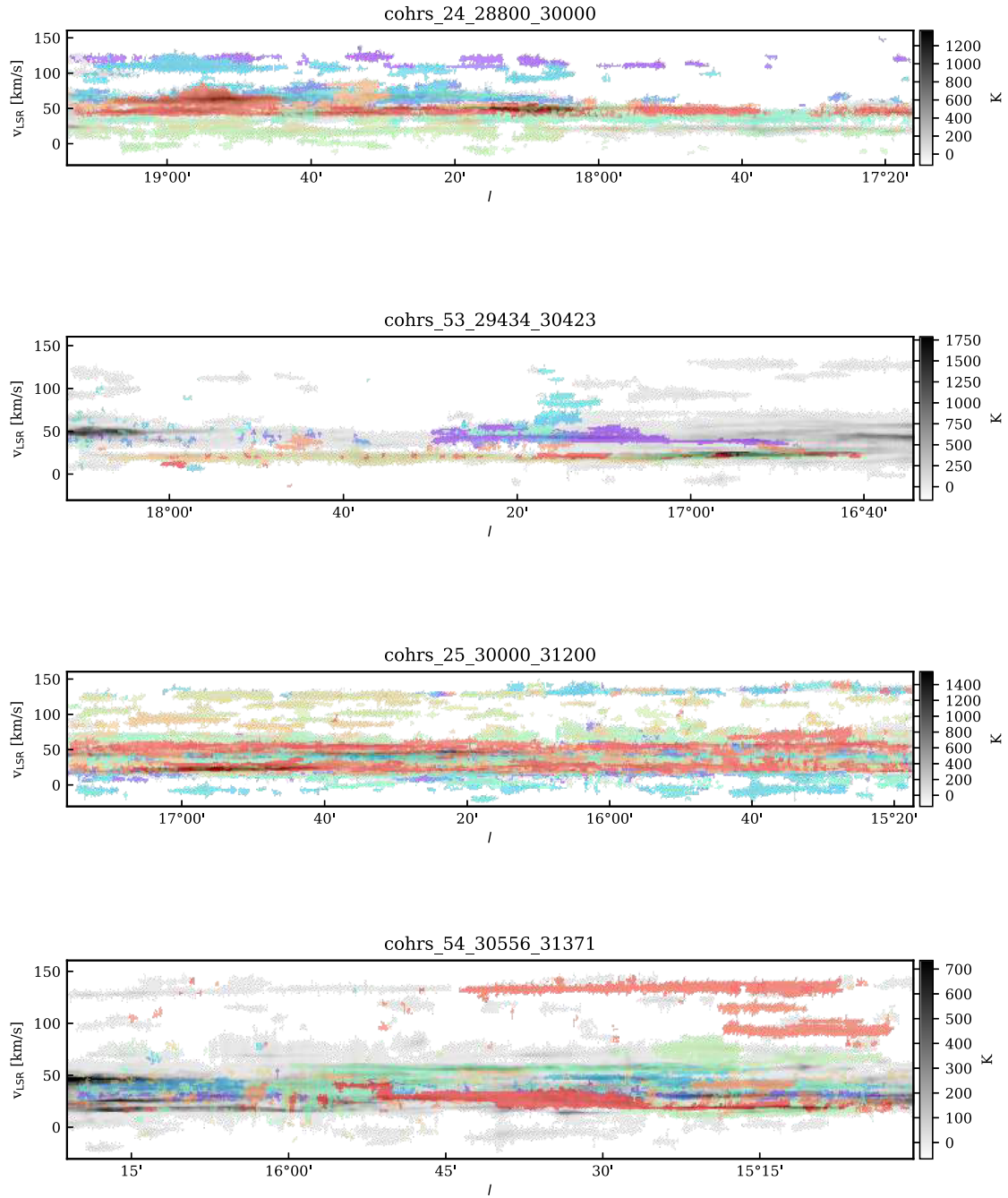


Figure E28. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

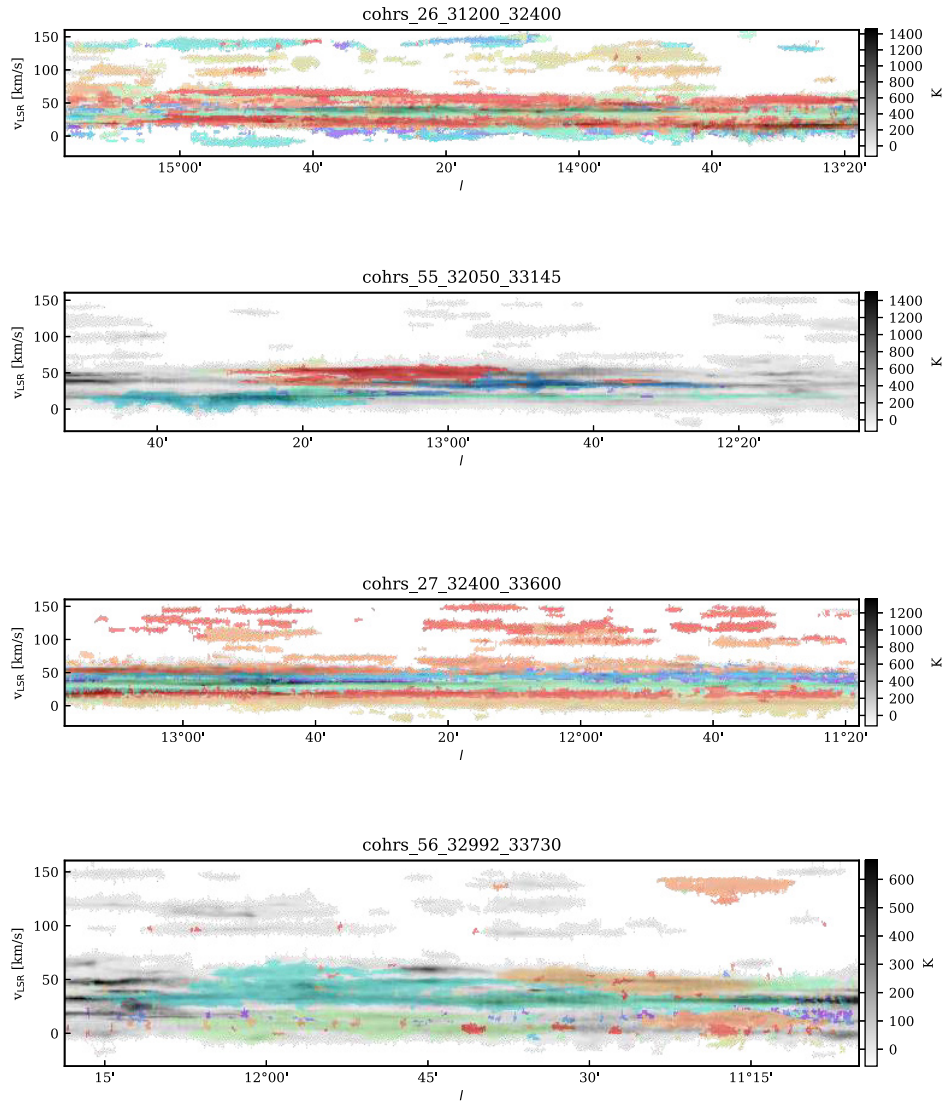


Figure E29. Longitude–velocity integrated maps of COHRS given sub-cubes masked as explained in Section 4. In colour the identified clouds are indicated.

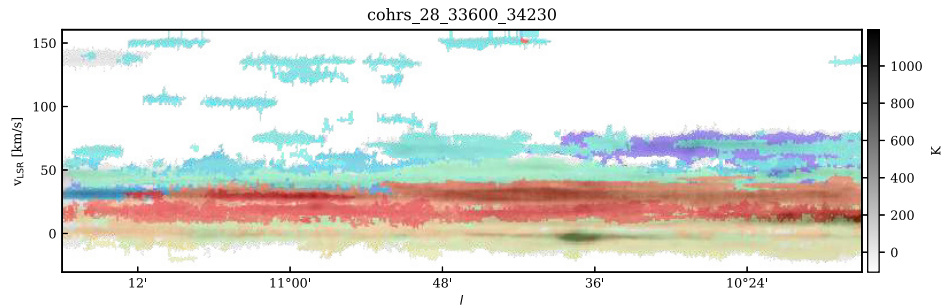


Figure E30. Longitude–velocity integrated map of COHRS given sub-cube masked as explained in Section 4. In colour the identified clouds are indicated.

This paper has been typeset from a \LaTeX file prepared by the author.