A Catalog of 13CO Clumps from the MWISP in l = 10°–20° (Postprint)

Authors: Xiaoyu Luo, Sheng Zheng, Zhibo Jiang, Zhiwei Chen, Yao Huang, Shuguang Zeng, Xiangyun Zeng, Rui Zhang, Chen Long, Guangrong Zhou et al., Sheng Zheng, Zhibo Jiang, Zhiwei Chen, Yao Huang, Shuguang Zeng, Xiangyun Zeng, Rui Zhang, Chen Long, Guangrong Zhou, Jinbo Hu

Date: 2024-05-24T00:00:00+00:00

Abstract

In this study, we present a catalog of molecular clumps extracted from 13CO (J = 1 − 0) emission data of the Milky Way Imaging Scroll Painting (MWISP) project. The data covers the inner Milky Way within the longitude range $10° \leq l \leq 20°$ and the latitude strip of . The workflow for the extraction of clumps, namely Facet-SS-3D-Clump, consists of two parts: the identification of clump candidates and their verification. First, Facet-SS-3D-Clump employs FacetClumps to identify clump candidates. Subsequently, high-confidence clumps are obtained by cross-matching with the clumps detected by other algorithms, such as dendrogram. Second, these high-confidence clumps are used as prior knowledge to train a semi-supervised deep clustering approach, SS-3D-Clump, which is applied to verify clump candidates detected by FacetClumps, providing confidence levels for the molecular clumps. Finally, the catalog comprising 18,757 molecular clumps was obtained using Facet-SS-3D-Clump, and the catalog is 90% complete above 37 K km s−1. We observe a significant deviation of the mean Galactic latitude for clumps within b ≤ 2° from the midplane, with . We found that 82.3% of the dust clumps correspond to 13CO clumps by matching with Herschel infrared dust clumps. In the future, Facet-SS-3D-Clump will be applied to detect 13CO clumps in the entire MWISP data.

Full Text

Preamble

Research in Astronomy and Astrophysics, 24:055018 (17pp), 2024 May https://doi.org/10.1088/1674-4527/ad3d12

A Catalog of 13CO Clumps from the MWISP in l = 10°–20°

Xiaoyu Luo[1,2], Sheng Zheng[1,3], Zhibo Jiang[4], Zhiwei Chen[4], Yao Huang[1,3], Shuguang Zeng[1,3], Xiangyun Zeng[1,3], Rui Zhang[1,3], Chen Long[1,3], Guangrong Zhou[1,3], and Jinbo Hu[1]

[1] Center for Astronomy and Space Sciences, China Three Gorges University, Yichang 443000, China; zsh@ctgu.edu.cn [2] College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443000, China [4] Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China; zbjiang@pmo.ac.cn

Received 2024 February 4; revised 2024 March 18; accepted 2024 April 2; published 2024 May 13

Abstract

In this study, we present a catalog of molecular clumps extracted from 13CO (J = $1-0$) emission data from the Milky Way Imaging Scroll Painting (MWISP) project, covering the inner Milky Way within the longitude range $10° \leq l \leq 20°$ and latitude strip $|b| \leq 5°.25$. The workflow for clump extraction, termed Facet-SS-3D-Clump, consists of two components: identification of clump candidates and their verification. First, Facet-SS-3D-Clump employs FacetClumps to identify clump candidates, obtaining high-confidence clumps through cross-matching with clumps detected by other algorithms such as dendrogram. Second, these high-confidence clumps serve as prior knowledge to train a semi-supervised deep clustering approach, SS-3D-Clump, which verifies clump candidates detected by FacetClumps and provides confidence levels for the molecular clumps. Using Facet-SS-3D-Clump, we obtained a catalog comprising 18,757 molecular clumps that is 90% complete above 37 K km s^{-1}. We observe a significant deviation of the mean Galactic latitude for clumps within $|b| \leq 2°$ from the midplane, with $b = -0°.110$. By matching with Herschel infrared dust clumps, we found that 82.3% of the dust clumps correspond to 13CO clumps. In the future, Facet-SS-3D-Clump will be applied to detect 13CO clumps across the entire MWISP dataset.

Key words: ISM: molecules – methods: data analysis – stars: formation

1. Introduction

It is widely accepted that stars form in molecular clouds (MCs; e.g., Shu et al. 1987; Krumholz & McKee 2005; Zinnecker & Yorke 2007; Krumholz et al. 2009; Nejad-Asghar 2011; Könyves et al. 2015; Basu 2016), while molecular clumps that condense from the ambient cloud represent a critical step in the star formation process (Hacar et al. 2013). Molecular clump formation also appears to play an essential role in determining the final mass of stars, since the distribution of masses among starless cores in a cloud mimics the stellar initial mass function (IMF; e.g., Motte et al. 1998; Alves et al. 2007; Nutter

& Ward-Thompson 2007; Könyves et al. 2010; Benedettini et al. 2018; Bresnahan et al. 2018; Arzoumanian et al. 2019). Although great progress has been made in determining the form of the IMF (e.g., Chabrier 2003; Nutter & Ward-Thompson 2007; Olmi et al. 2009; Kroupa et al. 2013; Könyves et al. 2015), a detailed explanation of its form and possible environmental dependence requires understanding the nature and evolution of molecular clumps (Marsh et al. 2016), particularly dense molecular clumps. Thus, surveys of molecular clumps in the interstellar medium (ISM) play an important role in understanding the location and mode of star formation in the Milky Way.

Since the discovery of CO molecules in the 1970s, numerous survey projects have been carried out successively, including the Galactic Ring Survey (GRS; Jackson et al. 2006), the Exeter-FCRAO CO Galactic Plane Survey (Mottram & Brunt 2010), the Milky Way Imaging Scroll Painting (MWISP; Jiang & Li 2013), the Three-mm Ultimate Mopra Milky Way Survey (ThrUMMS; Barnes et al. 2015), the CO Heterodyne Inner Milky Way Plane Survey (CHIMPS; Rigby et al. 2016), the Structure, Excitation, and Dynamics of the Inner Galactic Interstellar Medium (SEDIGISM; Schuller et al. 2017) survey, and the Forgotten Quadrant Survey (FQS; Benedettini et al. 2017).

These surveys, with different sensitivity and spatial resolution, help us detect the distribution of molecular components across various scales (Benedettini et al. 2021): from dense clouds and filamentary structures (André et al. 2014) to pre-stellar clumps and young stellar objects (YSOs). The MWISP project, implemented with the Purple Mountain Observatory Delingha (PMODLH) 13.7 m telescope (Zuo et al. 2011), is dedicated to a new large-scale survey of molecular gas in 12CO (J = 1 − 0), 13CO (J = 1 − 0), and C18O (J = 1 − 0), targeting the northern Galactic Plane within $-10°.25 \leq l \leq 250°.25$ and $-5°.25 \leq b \leq 5°.25$, along with several other regions of interest. The goal is to advance our knowledge of the precise content and distribution of molecular gas in the Milky Way, the physical and chemical processes converting tenuous interstellar gas to dense molecular gas, and the rules governing star formation from MCs.

With the progress of sky survey projects, unsupervised machine learning-based algorithms for detecting molecular clumps have also emerged (e.g., Rosolowsky et al. 2008; Berry 2015; Luo et al. 2022; Jiang et al. 2023). The dendrogram algorithm, developed by Rosolowsky et al. (2008), is well-suited for illustrating changes in the hierarchical structure of isosurfaces within molecular line data cubes as contour levels vary (Rani et al. 2023). It has been widely used in continuum, atomic hydrogen, and molecular line data (Cheng et al. 2018; Takekoshi et al. 2019; Nakanishi et al. 2020; Zhang et al. 2021). Luo et al. (2022) utilized a local density clustering method, inspired by the concept from Alex Rodriguez (2014), to identify clumps as local dense regions embedded within molecular gas with lower average bulk density, as suggested in Blitz & Stark (1986); Lada (1992); Bergin & Tafalla (2007). Additionally, a Multiple Gaussian Model (MGM) is utilized to simultaneously fit overlapping clumps, allowing for the extraction of clump parameters. Jiang et al. (2023) presented FacetClumps, a

Gaussian facet-based model designed for fitting local surfaces. The extremum determination theorem of multivariate functions is applied to determine clump centers, and local regions around the centers are then clustered to identify clumps by considering connectivity and minimum distance. Experiments by Jiang et al. (2023) show that FacetClumps exhibits excellent recall and precision rates.

The accumulation of survey data and development of automatic detection algorithms for molecular clumps have made it possible to conduct a comprehensive census of these clumps. For example, Wu et al. (2012) expanded the horizon of cold astronomy by studying Planck cold clumps, while Marsh et al. (2016) published a catalog of dense cores in the Taurus star-forming region derived from Herschel SPIRE and PACS observations. Benedettini et al. (2020) presented a catalog of MCs extracted from FQS 12CO ($J = 1 - 0$) spectral cubes. Typically, researchers mitigate false positives through visual inspection; for instance, Rigby et al. (2019) visually inspected each CHIMPS clump with three independent reviewers and assigned a reliability flag, while Rojas et al. (2022) visually inspected and classified lens candidates into two catalogs using high-resolution imaging and spectroscopy. However, researchers must devise meticulous strategies to eliminate potential subjectivity introduced during the visual inspection process, which requires significant work and time. Therefore, an automated verification method as a substitute for manual verification becomes increasingly necessary in the era of extensive data. Luo et al. (2024) proposed a semi-supervised deep clustering method for molecular clump verification, namely SS-3D-Clump, which extracts deep features of clumps and classifies these features through a clustering algorithm to obtain pseudo-labels. Subsequently, these pseudo-labels are used as supervision to update the weights of the entire network. This approach can leverage unlabeled samples through semi-supervised learning to enhance the generalization ability of SS-3D-Clump and has shown remarkable performance.

In this paper, we present a comprehensive workflow for detecting and verifying 13CO clumps in the MWISP project. The workflow, termed Facet-SS-3D-Clump, first employs FacetClumps to obtain candidates for molecular clumps and then uses SS-3D-Clump for their verification. To validate the reliability of the workflow, we selected a subregion within $10° \leq l \leq 20°$ & $-5°.25 \leq b \leq 5°.25$ from MWISP as experimental data, which includes several active high-mass star-forming regions (hereafter HSRs), such as M16 (Hill et al. 2012; Tremblin et al. 2014), M17 (Felli et al. 1984; Chen et al. 2021; Yin et al. 2022), W31 (Beuther et al. 2011; Gama et al. 2016; Maity et al. 2022), W33 (Messineo et al. 2015; Tursun et al. 2022), and W39 (Kerton et al. 2013). Additionally, the infrared dust clumps from Herschel (Elia et al. 2017) also cover these HSRs, providing an opportunity to validate Facet-SS-3D-Clump using data from different wavelength bands. Dense molecular cores and clumps are empirically defined as compact (0.1 and 1 pc, respectively) and dense (10^4–10^5 H_2 cm^{-3}) structures (e.g., Williams et al. 2000; Zhang et al. 2009; Ohashi et al. 2016; Motte et al. 2018). Here, we refer to the compact regions traced by 13CO as "clumps." Using Facet-SS-3D-Clump, we obtained a catalog of clumps in the HSR.

The structure of this paper is as follows: Section 2 introduces the MWISP observations. Section 3 describes the details of Facet-SS-3D-Clump, including the molecular clump extraction algorithm, parameter estimation algorithm, completeness experiments, and SS-3D-Clump verification of clump candidates. In Section 4, we present a catalog containing 18,757 13CO clumps extracted from the HSR and provide a brief statistical analysis of clumps in terms of their spatial distribution and matching results with Herschel dust clumps. The conclusion is provided in Section 5.

2. Data Introduction

The MWISP observations utilize a nine-beam superconducting spectroscopic array receiver, operating in sideband separation mode and employing a fast Fourier transform spectrometer (Shan et al. 2012; Su et al. 2019). The three CO isotopologue lines, including 12CO, 13CO, and C18O, can be observed simultaneously. All observations are taken in position-switch On-The-Fly (OTF; see Sun et al. 2018) mode. The observing strategy and data reduction adopted for the MWISP Survey are described in detail in Su et al. (2019), and details on the telescope can be found at http://www.radioast.nsdc.cn/mwisp.php. The 13CO line was observed in the upper sideband with a main beamwidth of 48 , while the 13CO and C18O lines were observed in the lower sideband with a main beamwidth of 50 . Both bandwidths are 1000 MHz wide with 16,384 channels, resulting in velocity separations of about 0.159 km s^{-1} for 12CO and 0.166 km s^{-1} for 13CO and C18O. The typical noise temperatures, including the atmosphere, are 140 K and 250 K at 110 GHz and 115 GHz, respectively. A detailed description of the instrument was given by Shan et al. (2012).

The sky coverage of the MWISP project is divided into 10,941 cells. Each cell, being 30 × 30 , is scanned along Galactic longitude (l) and Galactic latitude (b) at least twice to reduce noise fluctuations. The reasons for choosing a cell size of 30 × 30 are detailed in Su et al. (2019). The sample spacing between adjacent scans is 15 , ensuring that the MWISP survey fully samples the mapped area. The OTF raw data were regridded into 30 × 30 pixels. Data reduction was carried out using the GILDAS software described in detail in Pety (2005). Finally, three-dimensional (3D) FITS data cubes of each cell were produced with a grid spacing of 30 for the 12CO, 13CO, and C18O ($J = 1 - 0$) lines. The rms distributions are presented in Figure 1 [Figure 1: see original paper]. The typical rms noise levels of the spectra are 0.47, 0.22, and 0.21 K for 12CO ($J = 1 - 0$), 13CO ($J = 1 - 0$), and C18O ($J = 1 - 0$), respectively.

The 13CO integrated $l - b$ map of the HSR is shown in Figure 2 [Figure 2: see original paper], integrated from v_{LSR} = -5 to 80 km s^{-1}. The integrated intensity threshold is $2 \times \sigma\{rms\}, where\ \sigma$\{rms\} is the rms noise per velocity channel. Several renowned star-forming regions are present in the HSR. Surveying molecular clumps in the HSR provides a substantial sample for studying different evolutionary stages of molecular clumps and represents a crucial step in understanding the early stages of stellar evolution.

3. Generation of Clump Catalogs from MWISP

Star-forming regions are messy and chaotic environments with structures on many scales (Wurster & Rowan 2023). The entire region is typically referred to as a cloud, dense regions embedded within the cloud are clumps, and the very dense regions within clumps are cores (Bergin & Tafalla 2007). While there is general agreement on these three terms, there is ambiguity regarding specific definitions and divisions between the levels. For example, Rathborne et al. (2009) identified MCs and clumps using the ClumpFind algorithm with 13CO (J = 1 − 0) emission line data from GRS (Jackson et al. 2006). Takekoshi et al. (2019) analyzed the statistical properties of C18O (J = 1 − 0) clumps in the Cygnus X cluster-forming region using data from the Nobeyama 45 m radio telescope, with clumps identified through the dendrogram algorithm (Rosolowsky et al. 2008). Liu et al. (2022) refer to centrally concentrated structures as clumps and treat dendrogram-defined leaves accordingly.

The main characteristics of molecular clumps are local intensity enhancement and varied shapes, while they are embedded in molecular gas with lower average bulk density (Blitz & Stark 1986; Lada 1992). Clumps (or cores) are empirically defined as regions with concentrated, enhanced intensity in a data cube. It is worth noting that 13CO can only trace structures of specific density, which falls within the range of 10^3–10^4 cm^{-3}. This limitation means that 13CO reveals only local structures within MCs. Therefore, in this study, we define structures traced by 13CO data as clumps.

The workflow of Facet-SS-3D-Clump for obtaining the molecular clump catalog is shown in Figure 3 [Figure 3: see original paper], in which we utilize the dendrogram (Rosolowsky et al. 2008) and FacetClumps (Jiang et al. 2023) algorithms to detect 13CO data. Subsequently, we perform cross-matching to obtain high-confidence clumps. Simultaneously, these high-confidence clumps are used as prior knowledge to train an SS-3D-Clump model, which verifies the remaining clump candidates and provides confidence levels for molecular clumps. At the same time, we use an MGM described in Luo et al. (2022) to simultaneously fit overlapping molecular clumps and derive their geometric morphological parameters. Finally, we obtain the catalog of molecular clumps, which includes positions, morphological parameters, and confidence levels associated with the clumps.

To avoid the computational expense of detecting clumps from the entire MWISP data, we employ a strategy that processes sub-cubes and performs global stitching (see Figure 4 [Figure 4: see original paper]). A sub-cube spans 120 pixels in longitude and latitude, corresponding to 1°, given the MWISP data pixel size of 30 . Candidates whose centers fall outside the local region of the sub-cube are rejected from the first catalog. Then, clumps belonging to the entire region are obtained by retaining those whose centers lie within the local regions of all sub-cubes. As shown in Figure 4, clumps labeled a, b, and c are in sub-cube 1, while the center of clump c (marked by a red dot) lies outside the local region of sub-cube 1. Thus, clump c (colored gray) is removed from sub-cube 1, while clumps a and b (colored blue) are retained in the catalog of sub-cube 1. In sub-cube 2, the center of clump b lies outside the local region, while that of clump c falls within the local region, so clump c is retained in the catalog of sub-cube 2.

3.1. Clump Extraction Algorithm

Dendrogram (Rosolowsky et al. 2008), implemented in astrodendro, is an abstraction of the changing topology of isosurfaces as a function of contour level, using a tree diagram to describe hierarchical structures over a range of scales in a two-dimensional (2D) or 3D datacube (Zhang et al. 2021). Two types of structures are returned in the results: leaves, which have no substructure, and branches, which can split into multiple branches or leaves. The algorithm has two main parameters: T_{min} and ΔT. T_{min} is the minimum value to be considered in the dataset; in the fiducial case, we adopt T_{min} = 2σ{rms}, *where* σ{rms} is the local noise level. ΔT describes how significant a leaf must be to be considered an independent entity. We adopt a fiducial value of ΔT = 2σ{rms}, *meaning a clump must have a peak flux reaching* 5σ{rms} above the noise.

FacetClumps was developed by Jiang et al. (2023), which uses a Gaussian facet model to fit local surfaces and employs the extremum determination theorem of multivariate functions to determine clump centers. Based on the identified clump centers, FacetClumps clusters regions near the centers by considering connectivity and minimum distance, thereby detecting molecular clumps. The FacetClumps algorithm uses default parameters in this paper (see Appendix C in Jiang et al. 2023). By cross-matching the detection results of the two algorithms, "extremely" high-confidence clumps can be obtained. As shown in Figure 5 [Figure 5: see original paper], drawn using the Cube Analysis and Rendering Tool for Astronomy (CARTA; Comrie et al. 2021), the green circles and blue plus signs represent the centroids of clumps detected by FacetClumps and dendrogram, respectively. When the deviations of peak positions between two clumps detected by the two algorithms are less than 2 pixels in all three directions—corresponding to deviations less than 1 in Galactic longitude and latitude and less than 0.32 km s^{-1} in the velocity axis in the actual data—they are considered to be the same clump. In the HSR, FacetClumps and dendrogram detected 24,367 and 23,019 clumps, respectively, with 10,754 clumps matched between the two.

3.2. Verification by SS-3D-Clump

To minimize false positives, both Rigby et al. (2019) and Rojas et al. (2022) employed visual inspection by independent reviewers, assigning reliability flags for CHIMPS clumps and classifying lens candidates into two catalogs using high-resolution imaging and spectroscopy, respectively. To minimize subjectivity in manual verification, researchers make every effort to control this uncertainty during the verification process. For example, Rojas et al. (2022) re-displayed previously classified ring galaxies during the verification process as a consistency check (users should re-classify them as rings, or at least not classify them as lenses). Therefore, manually verifying candidates requires significant effort and time, especially for large datasets. Consequently, an automated verification method as a substitute for manual verification becomes increasingly necessary in the era of big data.

Inspired by the success of supervised deep learning in galaxy classification (Zhu et al. 2019; Cheng et al. 2020; Lukic et al. 2019; He et al. 2021; Gupta et al. 2022), we integrate limited labeled data with a deep learning approach. Utilizing semi-supervised deep learning, namely SS-3D-Clump (Luo et al. 2024), we verify the clump candidates. As shown in Figure 6 [Figure 6: see original paper], the deep features of clump candidates are extracted by the feature extraction component of SS-3D-Clump (Luo et al. 2024). Using these deep features as a foundation, the Constrained-KMeans (Basu et al. 2002) algorithm efficiently generates pseudo-labels for candidates using small labeled samples as seeds. The SS-3D-Clump classifier also assigns predict-labels to candidates based on their deep features. The difference between these labels is used to optimize the parameters of the SS-3D-Clump model. SS-3D-Clump iteratively groups deep features with a standard clustering algorithm and uses the subsequent assignments as supervision to update the network weights.

Figure 7 [Figure 7: see original paper] illustrates the curve of Normalized Mutual Information (NMI) changing with epochs during the training process of SS-3D-Clump. NMI is a metric used to assess the similarity between two datasets or the performance of clustering algorithms. The range of NMI typically falls between 0 and 1, where 1 indicates complete similarity between two clustering results and 0 signifies no resemblance. During training, NMI curves can be utilized to observe how the algorithm's performance changes with increasing training epochs, facilitating optimization of the model training process. Measuring NMI between clusters at epoch $t-1$ and t provides insights into the actual stability of SS-3D-Clump. Figure 7 shows that by the 15th training epoch, NMI increases to around 0.97 and remains stable, indicating that SS-3D-Clump experiences fewer reassignments and the clusters are stabilizing.

Since SS-3D-Clump belongs to semi-supervised deep learning, verification of molecular clumps is completed simultaneously with model training. Using the stabilized SS-3D-Clump after training, we conducted verification on candidates detected by FacetClumps, excluding those with confidence levels below 0.8. Ultimately, SS-3D-Clump excluded 5,792 out of the 24,367 molecular clump candidates detected by FacetClumps, resulting in a catalog containing 18,575 molecular clumps. The histograms of peak intensities of the molecular clumps are illustrated in Figure 8 [Figure 8: see original paper]. The green histogram depicts the peak value distribution of clumps retained by SS-3D-Clump, while the red histogram shows the cases excluded by SS-3D-Clump. Figure 8 reveals that the distribution of retained clumps has a sharp rise and spike at approximately 1.5 K. After passing the peak position, the distribution descends rapidly. Appendix A presents a subset of results obtained through SS-3D-Clump verification, containing integrated maps of molecular clumps on the $l - b$ plane along with their associated confidence levels.

3.3. Morphological Parameter Estimation

For blended source pairs, Gaussian fitting results in larger size estimates, whereas when two or multiple sources are resolved by the detection algorithm, simultaneous fitting of multiple components enables better separation (Molinari et al. 2016). Therefore, we use the MGM described in Luo et al. (2022) to simultaneously fit overlapping molecular clumps and derive their geometric morphological parameters. The integrated map shown in Figure 9 [Figure 9: see original paper] represents a local region in the HSR with an integrated velocity range of 25.6–28.1 km s^{-1}. For molecular clumps detected by FacetClumps, morphological parameters are estimated using MGM. The green dots signify clump centroids, and the white ellipses are drawn based on the clumps' major axis, minor axis, and rotation angle.

3.4. Completeness

Telescope sensitivity limitations cause low-quality clumps to be missed. Key indicators of algorithm performance are completeness and the detection rate above this limitation. The "completeness limit" here refers to the total flux or mass above which a clump can be detected at a certain level with an algorithm. Smaller and weaker molecular clumps are less likely to be detected. To quantify the completeness of the extracted clumps, we carried out extensive synthetic data experiments by injecting simulated clumps into the 13CO ($J = 1 - 0$) data in the HSR. Completeness experiments were conducted by partitioning the entire HSR into 16 subregions that collectively cover the whole area.

Through multiple injections, we added 36,000 simulated clumps modeled as 3D ellipsoidal Gaussians for each subregion, with their axis sizes and intensities randomly distributed within specific ranges. The peak intensity values of these simulated clumps range from 0.7 to 15, while the size along the velocity axis ranges from 2 to 5, and the sizes in the Galactic longitude and latitude axes range from 1.5 to 4. This approach allowed us to assess the capability of recovering a statistically comparable population of clumps from the designated region. These simulated clumps were randomly spread across the map, with the only constraint being to avoid positional overlap.

The synthetic data were processed with FacetClumps using the same parameters as for the real HSR data, and the outputs were compared with the truth table of simulated clumps. The number of clumps within each total flux interval was counted to obtain completeness. An example of the recovery fraction as a function of integrated flux density of simulated clumps in different latitude intervals within the HSR is shown in Figure 10 [Figure 10: see original paper]. The figure shows that completeness curves decrease significantly when clump flux ranges from 40 to 20 K km s^{-1} within latitude intervals. When the completeness reaches 90% for the entire region, the total flux of clumps is approximately 37 K km s^{-1}. In Figure 11 [Figure 11: see original paper], the dashed line represents the estimated completeness limit as a function of Galactic latitude, while the histogram shows the distribution of missed simulated clumps in Galactic latitude during the synthetic data experiment. Around 0°, specifically in $-1°.5 \leq b \leq 0°.5$, the dashed line exhibits a noticeable increase, corresponding to a peak in the histogram at this position. This is explained by brighter emission at lower latitudes, making detection of fainter objects more challenging.

The completeness limits reported in Figure 11 should be considered conservative because they are determined by spreading synthetic clumps randomly across each entire region. However, the background has non-uniform characteristics in each region, decreasing toward the north and south Galactic directions as Galactic latitude increases. Figure 12 [Figure 12: see original paper] shows a typical example, with the upper panel displaying the integrated 13CO map of MWISP centered at l = 17°. Superimposed are the extracted clumps with integrated fluxes above (yellow crosses) and below (blue crosses) the completeness limit. The lower panel further depicts the latitude distribution of the two groups of clumps, signified with solid and dashed lines for sources above and below the confusion limit, respectively. The two groups exhibit subtle differences in spatial distribution, with clumps brighter than the completeness limit mainly concentrated in the Galactic latitude range of −1° to 0°.6, while the distribution of the other group is more uniform.

4.1. Detailed Contents of Clump Catalog

A total of 18,575 13CO clumps were extracted in the HSR. Table B1 in Appendix B presents a portion of the clump catalog, while the full catalog can be obtained online (https://www.scidb.cn/en/s/qEfe2m). In the table, columns (2) and (3) provide the positions of clumps in Galactic coordinates, and column (4) gives the central radial velocity with respect to the local standard of rest (LSR). Columns (5) and (6) represent the major and minor sizes of the fitted ellipse on the sky plane, respectively. Column (7) represents the FWHM size in the velocity direction. Column (8) is the tilt angle on the sky plane, calculated from the north Galactic pole to the major axis of the clump in the counterclockwise direction. Columns (9) and (10) list the clump's peak intensity and total integrated intensity, respectively. Column (11) represents the peak signal-to-noise ratio (S/N) of the clump. Column (12) is the fitting error, calculated by the rms error between the fitted and actual values. Column (13) contains the confidence levels associated with the clumps, verified by SS-3D-Clump. Names, units, symbols, and detailed descriptions of the columns in the molecular clump catalog are provided in Table 1 .

4.2. Cross Match with Hi-GAL Sources

The Herschel infrared Galactic Plane Survey (Hi-GAL) is a large-scale survey of the Galactic plane (Molinari et al. 2010) in five infrared continuum bands performed with Herschel between 70 and 500 m using the Herschel Space Observatory (Pilbratt et al. 2010). The first public release of high-quality products from the Hi-GAL survey was presented by Molinari et al. (2016), in which 10^5 compact sources are identified in each band. Elia et al. (2017) merged the sources from the five bands to obtain a band-merged catalog containing 100,922 sources with regular spectral energy distributions (SEDs), 24,584 of which show a 70 m counterpart and are considered protostellar.

Within the Galactic longitude range $10° \leq l \leq 20°$, there are 8,517 band-merged sources. Through matching with 13CO clumps, 7,154 sources have been identified as corresponding to 13CO clumps. Among them, 2,200 sources have one-to-one matches with 13CO clumps, while 4,954 sources match multiple 13CO clumps at different velocities. Figure 13 [Figure 13: see original paper] illustrates the one-to-one matching results of a Herschel 350 m dust clump and a 13CO clump. The upper-left subplot shows the channel map with the maximum intensity of the 13CO clump, with the corresponding Herschel dust clump located in the upper-right subplot. The lower subplot illustrates the average spectrum of the 13CO clump, with the averaging range delineated by the black circle in the upper-left subplot. The blue contours of 13CO in Figure 13 and their overlay onto the dust clumps show that the positions correspond very well. Figure 14 [Figure 14: see original paper] displays similar results, but showcases a Herschel dust clump matched to multiple 13CO clumps. The figure reveals similarities between the 13CO data and Herschel dust emission. In the one-to-many matching results, the average spectrum of 13CO data at the positions of Herschel dust clumps exhibits multiple peaks, explaining why the number of 13CO clumps exceeds that of Herschel dust clumps. The upper-left subplot of Figure 14 displays the channel map of maximum intensity in the average spectrum.

4.3. Spatial Distribution

A comparison with Hi-GAL dust clumps (Elia et al. 2017), covering the inner Milky Way in the longitude range $10° \leq l \leq 20°$, shows substantial similarities in source count distributions. The latitude distribution of HSR clumps and Hi-GAL compact sources is shown in Figure 15 [Figure 15: see original paper]. The blue histogram represents the distribution of molecular clumps obtained by Facet-SS-3D-Clump, while the orange histogram represents the distribution of dust clumps in Hi-GAL (Elia et al. 2017). Both histograms peak at slightly negative values, with the median HSR clump latitude being $-0°.118$ below the nominal midplane. The mean Galactic latitude for clumps within $|b| \leq 2°$ is also significantly below the midplane: $b = -0°.110$. The histograms also reveal that the distribution of molecular clumps exhibits an asymmetrical pattern. However, we cannot rule out an asymmetric distribution with respect to the central plane in the studied field. As Molinari et al. (2016) pointed out, this may be attributed to an incorrect assumption about the vertical position of the Sun in the Milky Way.

The longitude distribution of 13CO clumps is shown in Figure 16 [Figure 16: see original paper], with the dashed line representing the star-forming region in the HSR. The figure shows a local clustering of 13CO clumps near the longitude of the star-forming region, reflecting the consistency of clumps obtained by Facet-SS-3D-Clump. The top panel of Figure 17 [Figure 17: see original paper] shows clump positions in the $l - b$ plane. The bottom panel presents clump centers in the $l - v$ plane. The yellow, red, green, and purple lines correspond to the spiral arms of Aql Rift, Sagittarius, Scutum, and Norma, respectively. Dashed lines indicate the near side of the arms, while solid lines represent the far side.

5. Conclusion

We employed Facet-SS-3D-Clump to identify and verify molecular clumps in the 13CO data of MWISP, covering the longitude strip $10° \leq l \leq 20°$ and latitude range $-5°.25 \leq b \leq 5°.25$. Based on analysis of the results, we conclude that: (1) We obtained a catalog containing 18,575 13CO clumps from the MWISP data using Facet-SS-3D-Clump. (2) The flux completeness limits show that the catalog is 90% complete above 37 K km s^{-1}. (3) Matching results with Herschel infrared dust clumps reveal that 82.3% of dust clumps correspond to 13CO clumps. Facet-SS-3D-Clump will be applied to survey 13CO clumps across the entire MWISP dataset.

This work was supported by the National Natural Science Foundation of China (NSFC; Grant Nos. U2031202, 11903083, 11873093, and 12203029). This research made use of data from the Milky Way Imaging Scroll Painting (MWISP) project, a multi-line survey in 12CO/13CO/C18O along the northern Galactic plane with the PMO-13.7 m telescope. We are grateful to all members of the MWISP working group, particularly the staff at the PMO-13.7 m telescope, for their long-term support. MWISP was sponsored by the National Key R&D Program of China with grant 2017YFA0402701 and the CAS Key Research Program of Frontier Sciences with grant QYZDJ-SSW-SLH047. We thank Hongli Liu for useful suggestions on the manuscript.

Software: CARTA (Comrie et al. 2021), Astropy (Astropy Collaboration et al. 2018), TensorFlow (Abadi et al. 2015), and Scikit-Learn (Pedregosa et al. 2011).

Appendix A: Examples of Clumps Verified by SS-3D-Clump

To demonstrate the efficacy of SS-3D-Clump in verifying molecular clump candidates, a subset of samples was visualized. Figure A1 shows molecular clumps on the l–b plane with confidence levels obtained through SS-3D-Clump verification.

Appendix B: Partial Molecular Clumps in the Catalog

Table B1 presents a partial catalog of clumps, providing information about their positions, sizes, intensities, and fluxes.

References

Abadi, M., Agarwal, A., Barham, P., et al., 2015 v2.14.0 TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Zenodo, Alex Rodriguez, A. L. 2014, Science, 344, 1492 Alves, J., Lombardi, M., & Lada, C. J. 2007, A&A, 462, L17 André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. of Arizona Press), 27, Protostars and Planets VI Arzoumanian, D., André, P., Könyves, V., et al. 2019, A&A, 621, A42 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123 Barnes, P. J., Muller, E., Indermuehle, B., et al. 2015, ApJ, 812, 6 Basu, S. 2016, in IAU Symposium, Vol. 315, From Interstellar Clouds to Star-Forming Galaxies: Universal Processes?, ed. P. Jablonka, P. André, & F. van der Tak (Cambridge: Cambridge University Press), 85 Basu, S., Banerjee, A., & Mooney, R. J. 2002, Machine Learning, Proc. of the Nineteenth Int. Conf. (ICML 2002), Univ. of New South Wales, Sydney, Australia, July 8-12, 2002 Benedettini, M., Molinari, S., Baldeschi, A., et al. 2017, MMSAI, 88, 722 Benedettini, M., Molinari, S., Baldeschi, A., et al. 2020, A&A, 633, A147 Benedettini, M., Pezzuto, S., Schisano, E., et al. 2018, A&A, 619, A52 Benedettini, M., Traficante, A., Olmi, L., et al. 2021, A&A, 654, A144 Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 339 Berry, D. S. 2015, A&C, 10, 22 Beuther, H., Linz, H., Henning, T., et al. 2011, A&A, 531, A26 Blitz, L., & Stark, A. A. 1986, ApJL, 300, L89 Bresnahan, D., Ward-Thompson, D., Kirk, J. M., et al. 2018, A&A, 615, A125 Chabrier, G. 2003, PASP, 115, 763 Chen, Z., Sun, W., Chini, R., et al. 2021, ApJ, 922, 90 Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2020, MNRAS, 493, 4209 Cheng, Y., Tan, J. C., Liu, M., et al. 2018, ApJ, 853, 160 Comrie, A., Wang, K.-S., Hsu, S.-C., et al., 2021 v2.0.0 CARTA: The Cube Analysis and Rendering Tool for Astronomy, Zenodo, Du, X., Xu, Y., Yang, J., & Sun, Y. 2017, ApJS, 229, 24 Elia, D., Molinari, S., Schisano, E., et al. 2017, MNRAS, 471, 100 Felli, M., Churchwell, E., & Massi, M. 1984, A&A, 136, 53 Gama, D. R. G., Lepine, J. R. D., Mendoza, E., Wu, Y., & Yuan, J. 2016, ApJ, 830, 57 Gupta, R., Srijith, P., & Desai, S. 2022, A&C, 38, 100543 Hacar, A., Tafalla, M., Kauffmann, J., & Kovács, A. 2013, A&A, 554, A55 He, Z., Qiu, B., Luo, A.-L., et al. 2021, MNRAS, 508, 2039 Hill, T., Motte, F., Didelon, P., et al. 2012, A&A, 542, A114 Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, ApJS, 163, 145 Jiang, Y., Chen, Z., Zheng, S., et al. 2023, ApJS, 267, 32 Jiang, Z., & Li, J. 2013, Protostars and Planets VI Posters Kerton, C. R., Arvidsson, K., & Alexander, M. J. 2013, AJ, 145, 78 Könyves, V., André, P., Men'shchikov, A., et al. 2010, A&A, 518, L106 Könyves, V., André, P., Men'shchikov, A., et al. 2015, A&A, 584, A91 Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. 2013, in Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations, ed. T. D. Oswalt & G. Gilmore, 5 (Dordrecht: Springer), 115 Krumholz, M. R., & McKee, C. F. 2005, ApJ, 630, 250 Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 699, 850 Lada, E. A. 1992, ApJL, 393, L25 Liu, L., Bureau, M., Li, G.-X., et al. 2022, MNRAS, 517, 632 Lukic, V., de Gasperin, F., & Brüggen, M. 2019, Galax, 8, 8, Galaxies Luo, X., Zheng, S., Huang, Y., et al. 2022, RAA, 22, 015003 Luo, X., Zheng, S., Jiang, Z., et al. 2024, A&A, 683, A104 Maity, A. K., Dewangan, L. K., Sano, H., et al. 2022, ApJ, 934, 2 Marsh, K. A., Kirk, J. M., André, P., et al. 2016, MNRAS, 459, 342 Messineo, M., Clark, J. S., Figer, D. F., et al. 2015, ApJ, 805, 110 Molinari, S., Schisano, E., Elia, D., et al. 2016, A&A, 591, A149 Molinari, S., Swinyard, B., Bally, J., et al. 2010, PASP, 122, 314 Motte, F., Andre, P., & Neri, R. 1998, A&A, 336, 150 Motte, F., Bontemps, S., & Louvet, F. 2018, ARA&A, 56, 41 Mottram, J. C., & Brunt, C. M. 2010, in ASP Conf. Ser. 438, The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey, ed. R. Kothes, T. L. Landecker, & A. G. Willis (San Francisco, CA: ASP), 98 Nakanishi, H., Fujita, S., Tachihara, K., et al. 2020, PASJ, 72, 43 Nejad-Asghar, M. 2011, Ap&SS, 334, 27 Nutter, D., & Ward-Thompson, D. 2007, MNRAS, 374, 1413 Ohashi, S., Sanhueza, P., Chen, H.-R. V., et al. 2016, ApJ, 833, 209 Olmi, L., Ade, P. A. R., Anglés-Alcázar, D., et al. 2009, ApJ, 707, 1836 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825 Pety, J. 2005, SF2A-2005: Semaine de l'Astrophysique Francaise ed. F. Casoli et al. (Paris: EdP Sciences), 721 Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1 Rani, R., Moore, T. J. T., Eden, D. J., et al. 2023, MNRAS, 523, 1832 Rathborne, J. M., Lada, C. J., Muench, A. A., et al. 2009, ApJ, 699, 742 Rigby, A. J., Moore, T. J. T., Eden, D. J., et al. 2019, A&A, 632, A58 Rigby, A. J., Moore, T. J. T., Plume, R., et al. 2016, MNRAS, 456, 2885 Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73 Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338 Schuller, F., Csengeri, T., Urquhart, J. S., et al. 2017, A&A, 601, A124 Shan, W., Yang, J., Shi, S., et al. 2012, ITTST, 2, 593 Shu, F. H., Adams, F. C., & Lizano, S. 1987, ARA&A, 25, 23 Su, Y., Yang, J., Zhang, S., et al. 2019, ApJS, 240, 9 Sun, J. X., Lu, D. R., Yang, J., et al. 2018, AcASn, 59, 3 Sun, Y., Su, Y., Zhang, S.-B., et al. 2017, ApJS, 230, 17 Sun, Y., Xu, Y., Yang, J., et al. 2015, ApJL, 798, L27 Takekoshi, T., Fujita, S., Nishimura, A., et al. 2019, ApJ, 883, 156 Tremblin, P., Schneider, N., Minier, V., et al. 2014, A&A, 564, A106 Tursun, K., Henkel, C., Esimbek, J., et al. 2022, A&A, 658, A34 Williams, J. P., Blitz, L., & McKee, C. F. 2000, in Protostars and Planets IV, ed. V. Mannings, A. P. Boss, & S. S. Russell (Cambridge: Cambridge University Press), 97 Wu, Y., Liu, T., Meng, F., et al. 2012, ApJ, 756, 76 Wurster, J., & Rowan, C. 2023, MNRAS, 523, 3025 Yin, J., Chen, Z., Yao, Y., et al. 2022, RAA, 22, 035021 Zhang, Q., Wang, Y., Pillai, T., & Rathborne, J. 2009, ApJ, 696, 268 Zhang, S., Xu, Y., & Yang, J. 2014, AJ, 147, 46 Zhang, S., Zavagno, A., López-Sepulcre, A., et al. 2021, A&A, 646, A25 Zhu, X.-P., Dai, J.-M., Bian, C.-J., et al. 2019, Ap&SS, 364, 55 Zinnecker, H., & Yorke, H. W. 2007, ARA&A, 45, 481 Zuo, Y.-X., Li, Y., Sun, J.-X., et al. 2011, ChA&A, 35, 439