Searching for Molecular Outflows with Support Vector Machines: The Dark Cloud Complex in Cygnus

Shaobo Zhang, Ji Yang, Ye Xu, Xuepeng Chen, Yang Su, Yan Sun, Xin Zhou, Yingjie Li, and Dengrong Lu
Purple Mountain Observatory & Key Laboratory for Radio Astronomy, Chinese Academy of Sciences, Nanjing 210023, People’s Republic of China

Received 2019 July 19; revised 2020 March 26; accepted 2020 April 6; published 2020 May 5

Abstract

We present a survey of molecular outflows across the dark cloud complex in the Cygnus region, based on a 46.75 deg² field of CO isotopologue data from the Milky Way Imaging Scroll Painting survey. A supervised machine-learning algorithm, the support vector machine, is introduced to accelerate our visual assessment of outflow features in the data cube of 12CO and 13CO J = 1−0 emission. A total of 130 outflow candidates are identified, 77 of which show bipolar structures and 118 are new detections. Spatially, these outflows are located inside dense molecular clouds, and some of them are found in clusters or in elongated linear structures tracing the underlying gas filament morphology. Along the line of sight, 97, 31, and 2 candidates reside in the Local, Perseus, and Outer Arms, respectively. Young stellar objects as outflow drivers are found near most outflows, while 36 candidates show no associated source. The clusters of outflows that we detect are inhomogeneous in their properties; nevertheless, we show that the outflows cannot inject turbulent energy on cloud scales. Instead, at best, they are restricted to affecting the so-called “clump” and “core” scales, and only on short (~0.3 Myr) estimated timescales. Combined with outflow samples in the literature, our work shows a tight outflow mass–size correlation.

Unified Astronomy Thesaurus concepts: Stellar jets (1607); Astronomical object identification (87); Molecular clouds (1072)

Supporting material: figure set, machine-readable tables

1. Introduction

Molecular outflows are ubiquitous in star-forming regions, as they are thought to be an indispensable process that every protostar undergoes (Bally & Lada 1983; Shu et al. 1987; Frank et al. 2014; Bally 2016). Since the discovery of the first bipolar structure by Snell et al. (1980), numerous observational and theoretical studies have been carried out that reveal outflows as a key link in the star-forming process. Outflows not only record the histories of mass ejection from young stars that are forming, but they can also be related to mass accretion onto protostars (Bally 2016), as well as feedback to the environment (Elmegreen & Scalo 2004; Scalo & Elmegreen 2004; Matzner 2007).

Molecular outflows are usually diagnosed in observations of high-velocity gas on molecular line wings, which show monopolar, bipolar, or multipolar localized structures (Wu et al. 2004). These molecular outflows have dimensions of up to several parsecs, wide mass ranges from $10^{-3}$ to $10^3 M_\odot$, and a typical kinetic energy of $10^{45}$ erg (Bally & Lada 1983; Wu et al. 2004). On the other hand, theoretical studies have proposed a wide range of outflow models concerning their launching mechanism (Shu et al. 2000; Pudritz et al. 2007), the formation of molecular outflows (summarized by Arce et al. 2007), and their feedback in self-regulating star formation (Nakamura & Li 2007; Federrath et al. 2014). However, the models have been debated for a long time, as no single model could explain all the features observed, especially for the outflows formed from clusters and massive stars, toward which there are a limited number of observations. A large-scale unbiased survey of outflows could move us forward in resolving the arguments.

The traditional methods to search for outflows are usually based on reference sources of known star formation activity obtained from other observations (Bally & Lada 1983; Dobashi et al. 2001; Hatchell et al. 2007; Li et al. 2015). Either infrared sources or dust cores were used as indicators for molecular outflows, which leads to biases during source selection and may limit the outflow study to cloud scales. Unbiased large-scale searches for molecular outflow became feasible in the past decade on account of mass data acquired from a number of molecular line surveys with (sub-)arcminute angular resolutions. CO surveys toward the Perseus, Taurus, and other regions were carried out using the Five College Radio Astronomy Observatory (FCRAO) telescope and the James Clerk Maxwell Telescope. Molecular outflows were manually searched for in these regions, and hundreds of new outflows were detected (Hatchell & Dunham 2009; Arce et al. 2010; Curtis et al. 2010; Narayanan et al. 2012; Li et al. 2015; Li et al. 2018). However, manual identification is time consuming, nonrepeatable, and involve subjective factors such as people’s perception of a image.

There have been a limited number of attempts to introduce computer-assisted methods in outflow searches. Arce et al. (2010) proposed a semiautomated procedure to visualize outflow candidates as spikes on a three-dimensional isosurface. Based on this idea, Li et al. (2018) combined the isosurface with optically thin lines and adopted a clump-finding algorithm to search for the spikes. However, these approaches are still based on empirical criteria, like traditional methods, in which a number of fine-tuned thresholds are necessary to find matched candidates. In recent years, machine-learning algorithms, with their simplicity and preciseness, have been widely involved in the field of pattern recognition. Relying on a set of prelabeled samples rather than fixed criteria, they are ideal tools to identify structures like outflows, which are difficult to define explicitly.
In this paper, we present a survey of molecular outflows toward a complex of dark clouds in Cygnus using a machine-learning algorithm. A pixel-by-pixel searching procedure is established to identify the gas emissions of outflows in CO data cubes. The structure of this paper is as follows. In Section 2, we describe the observations we made and the distributions of molecular clouds in the data. The method for searching outflow features is given in Section 3. In Section 4, we present the outflow catalog and analyze physical properties of the detected samples. These properties are compared with those in the literature, and feedback to the parent clouds is discussed in Section 5. A summary is then given in Section 6.

2. The Data
2.1. Observations

As part of the Milky Way Imaging Scroll Painting (MWISP; Su et al. 2019) project,1 we observed a series of Lynds dark clouds in the Cygnus region connected to L935 to IC 5146 with the Purple Mountain Observatory Delingha (PMODLH) 13.7 m telescope. $^{12}$CO (1−0), $^{13}$CO (1−0), and C$^{18}$O (1−0) lines were observed simultaneously with the 3 × 3 beam superconducting array receiver working in sideband separation mode, with a set of fast Fourier transform spectrometers employed (Shan et al. 2012). The observed region was separated into 187 cells of dimension 30′ × 30′ and covered a total area of 46.75 deg$^2$. Each cell was mapped using the on-the-fly (OTF) observation mode, with at least two scans along the Galactic longitude and latitude directions to reduce the scanning effects. A total of 600 hr were used for the observations from 2012 to 2016. The half-power beam width is 52″ at 115 GHz, and the pointing accuracy is better than 5″.

The standard chopper wheel method (Ulich & Haas 1976) was used to calibrate the antenna temperature. We divided the antenna temperature ($T_A$) by the main beam efficiency ($B_{\text{eff}}$) to obtain the main beam temperature ($T_{\text{mb}}$). Different efficiencies were adopted, as $B_{\text{eff}}$ varied within 44%–52% at 115 GHz and 48%–56% at 110 GHz during the five years of observations according to the annual status report of PMODLH. The calibration errors are estimated to be within 10%. The OTF data were regridded to 30″ pixels and then mosaicked to a FITS cube using the GILDAS software package (Pety 2005; Gildas Team 2013). The L935 region was mapped once in 2011 as a pilot survey (see Zhang et al. 2014 for details). We reprocessed these additional data following the same procedure and included them while mosaicking. The resulting rms noise is 0.44 K for $^{12}$CO at the resolution of 0.16 km s$^{-1}$, and 0.25 K for $^{13}$CO and C$^{18}$O at 0.17 km s$^{-1}$. The L935 region presents a lower noise level of 0.23 K and 0.15 K, respectively.

2.2. Distributions of Molecular Clouds

The spectra in our mapping region present three main velocity components, which correspond to the Local Arm ($−30 \sim 30$ km s$^{-1}$), the Perseus Arm ($−60 \sim −30$ km s$^{-1}$), and the Outer Arm ($−85 \sim −65$ km s$^{-1}$). The distributions of $^{12}$CO and $^{13}$CO in the three components are shown in Figure 1. $^{12}$CO emission is bright and presents filamentary or extended structures of different spatial scales in three components throughout the mosaic, while $^{13}$CO presents condensations that follow the distribution of $^{12}$CO. C$^{18}$O is not presented in the map as its emission only appears around a few peak positions of $^{13}$CO.

Most of the molecular clouds in the Local Arm are associated with dark nebulae (Lynds 1962; Dobashi et al. 2005; Dobashi 2011). The brightest portions among the clouds are the L933, L935, L936, L941, L975, and L1055 regions. A large H II region, Sh 2-117, is located in the west of our mapping region, where associated cavity structures are found surrounding the ionized gas (Bally & Scoville 1980; Zhang et al. 2014). Several Planck cold clumps (Planck Collaboration et al. 2016) are located within the filamentary clouds extending from L935 to IC 5146, which suggests a relatively quiescent environment. We note that there is a large shell structure appearing to the north of IC 5146 centered at L1012 with a diameter of ~4°, which closely matches the far-infrared loop identified by Kiss et al. (2004). Cometary and finger-like molecular clouds could be found on the loop near L1048, L1008, and L1001. The integrated intensity ratio of $^{12}$CO to $^{13}$CO is 6.8 in regions with $^{13}$CO intensity over 2.5 K km s$^{-1}$. Such ratio is similar to the result in Dobashi et al. (1994).

Different from the distributions in the Local Arm, clouds in the Perseus Arm are mainly located at $b > 2°$. Clouds in the Perseus Arm seem to be less extended than those in the Local Arm. Most clouds with strong $^{12}$CO emission are actively forming stars because there are massive young stellar objects (YSOs) and H II regions embedded in them, as reported by Urquhart et al. (2009). Two linked ring-like structures with consistent velocities can be spotted on the westernmost part of the mapping region (hereafter G084.9−0.4). One is centered at $l = 84°8$, $b = −0°6$ with a radius of ~12′, and the larger one to the north is peaked at a filamentary arc near $l = 85°1$, $b = 0°5$. Molecular clouds in the Outer Arm seem to be more compact because diffuse molecular gas far away from us (~9 kpc) is difficult to detect due to beam dilution. Still, filamentary and shell structures could be revealed in the map.

3. Searching Procedures

Our goal is to locate localized high-velocity wings (LHWs) in the data cube as outflow candidates. Considering the large amount of unbiased survey data, we require an automatic detector to blindly search outflow features based on a set of manually identified samples. In this way, our task becomes a supervised learning problem and involves the following steps.

1. Manually identify LHWs in a representative subset of data as training samples. The LHWs checked by eye should meet the criteria for outflows (Section 3.2).
2. Extract numerical feature vectors from all data. These features should be able to summarize the differences between the LHW and other emission, based on the line profile and spatial morphology (Section 3.3).
3. Train and optimize the learning model with training data using cross-validation. The model trained using a portion of the training samples should perform well in picking out LHWs from the rest of the samples (Section 3.4).
4. Apply the model to the rest of the test data set to obtain LHW candidates. The candidates should then be graded based on criteria in step 1 and paired into bipolar outflows if possible (Section 3.5).

---

1 http://www.radioast.nsdc.cn/mwisp.php
3.1. Support Vector Machines

We adopted the support vector machine (SVM; Vapnik 1995) algorithm over several other statistical methods owing to its excellent performance in multidimensional data. The SVM algorithm is a supervised learning algorithm used for classification and regression analysis. The main objective of SVM is to maximize the margin between different classes. This method has been widely used and proved to be efficient in different kinds of astronomy topics, such as galaxy classification (Huertas-Company et al. 2011), shocked gas search in supernova remnants (Beaumont et al. 2011), YSO classification (Marton et al. 2016), etc. In our study, we adopted the SVM with a Gaussian radial basis function kernel as a classifier, which calculated the decision boundaries between manually selected LHWs and non-LHWs, and applied the boundaries to samples of unknown classes. SVMlight (Joachims 1999) provides implementations of the SVM algorithm in a wide range of applications.

Figure 1. Integrated intensity map over three velocity ranges: $-30$ to $\sim 30$ km s$^{-1}$ (top), $-60$ to $\sim 30$ km s$^{-1}$ (middle), and $-85$ to $\sim 65$ km s$^{-1}$ (bottom). The background grayscale image is the integrated intensity map of $^{13}$CO, while the green contours represent the integrated intensity map of $^{13}$CO from $3\sigma$ at intervals of $5\sigma$. Dark nebulae associated with the molecular clouds are labeled on the map (Lynds 1962; Dobashi et al. 2005; Dobashi 2011). Circles indicate H II regions (dashed circles) and supernova remnants (solid circles), whose positions and sizes are cataloged in Sharpless (1959), Kothes et al. (2001), and Green (2014). Outflows identified in this work are marked with blue dots and red crosses, which represent the blueshifted and redshifted lobes, respectively. The dark and light symbols indicate outflows of grade A and B, respectively. Two boxes in the upper panel show the training region, while the dashed box outlines the validation set for cross-validation.
3. We plotted four position–velocity (P–V) diagrams slicing through the tip of each bump at position angles of 0°, 45°, 90°, and 135°, respectively. The LHWs are expected to present a bump on contours with velocity shift over 2 km s⁻¹ in all four P–V maps.

Applying these criteria resulted in 35 LHWs, 18 of which are blueshifted lobes. As we intended to do a pixel-by-pixel classification of the data, the LHWs were then disassembled into pixels. We extracted a total of 7702 pixels in the line wings as the LHW class training set.

For a successful learning process, it is vital to choose a set of appropriate features which could show the difference between the pixels of the two classes. An advantage of SVM is that the algorithm performs stably even in a high-dimensional feature space. For our data, the ¹²CO emissions are usually optically thick; thus, we measured the velocity and line width of each component from ¹³CO lines. Five features were then extracted at each pixel with intensity over 3σ in ¹²CO data according to the following methods.

1. \( T_{\text{inner}}/T_{\text{peak}} \), where \( T_{\text{inner}} \) is the intensity averaged within a \( 3 \times 3 \times 7 \) pixel subcube in position–position–velocity space, and \( T_{\text{peak}} \) is the peak intensity of ¹²CO spectra at the same spatial position. The subcube is centered at the pixel waiting to be classified and corresponds to a \( 2/5 \times 2/5 \times 1 \) km s⁻¹ region, which guarantees that LHWs larger than this size are not missing. This feature is expected to be lower in the line wings than near the line peak.

2. \( |v_{\text{pixel}} - v_0|/\Delta v \), where \( v_{\text{pixel}} \) is the velocity of the pixel to be classified, and \( v_0 \) and \( \Delta v \) are the systemic velocity and line width estimated from the ¹³CO component which the pixel belongs to. This feature represents the relative position of the pixel on the line profile and is expected to be high in the line wings.

3. \( s \times \Delta v/T_{\text{peak}} \), where \( s \) is the slope of a linear fitting to the averaged spectrum in the subcube, and \( \Delta v \) and \( T_{\text{peak}} \) are added to scale the slope. We use this feature to prevent the emission between two adjacent components from being classified as an LHW class. This feature is expected to be positive in most cases, but negative if another component appears.

4. \( \Gamma_{\sigma_{\text{LHW}}} \), the difference of the Gaussians of the image, involves the image convolved with the difference of the two Gaussian kernels of the standard deviation \( \sigma_1 \) and \( \sigma_2 \) (Gonzalez & Woods 2006). It is a commonly used algorithm in image processing for blob detection. We choose this feature to describe the localization of LHWs, as the integration intensities of the line wings are higher than those in the surrounding regions. The integration intensity map is derived within the same seven channels as the subcube. We adopt \( \sigma_1 = 2 \) and \( \sigma_2 = 4 \), which means the feature is more sensitive in detecting structures with spatial dimensions of \~2′–4′.

3.3. Feature Extraction

For a successful learning process, it is vital to choose a set of appropriate features which could show the difference between the pixels of the two classes. An advantage of SVM is that the algorithm performs stably even in a high-dimensional feature space. For our data, the ¹²CO emissions are usually optically thick; thus, we measured the velocity and line width of each component from ¹³CO lines. Five features were then extracted at each pixel with intensity over 3σ in ¹²CO data according to the following methods.

1. \( T_{\text{inner}}/T_{\text{peak}} \), where \( T_{\text{inner}} \) is the intensity averaged within a \( 3 \times 3 \times 7 \) pixel subcube in position–position–velocity space, and \( T_{\text{peak}} \) is the peak intensity of ¹²CO spectra at the same spatial position. The subcube is centered at the pixel waiting to be classified and corresponds to a \( 2/5 \times 2/5 \times 1 \) km s⁻¹ region, which guarantees that LHWs larger than this size are not missing. This feature is expected to be lower in the line wings than near the line peak.

2. \( |v_{\text{pixel}} - v_0|/\Delta v \), where \( v_{\text{pixel}} \) is the velocity of the pixel to be classified, and \( v_0 \) and \( \Delta v \) are the systemic velocity and line width estimated from the ¹³CO component which the pixel belongs to. This feature represents the relative position of the pixel on the line profile and is expected to be high in the line wings.

3. \( s \times \Delta v/T_{\text{peak}} \), where \( s \) is the slope of a linear fitting to the averaged spectrum in the subcube, and \( \Delta v \) and \( T_{\text{peak}} \) are added to scale the slope. We use this feature to prevent the emission between two adjacent components from being classified as an LHW class. This feature is expected to be positive in most cases, but negative if another component appears.

4. \( \Gamma_{\sigma_{\text{LHW}}} \), the difference of the Gaussians of the image, involves the image convolved with the difference of the two Gaussian kernels of the standard deviation \( \sigma_1 \) and \( \sigma_2 \) (Gonzalez & Woods 2006). It is a commonly used algorithm in image processing for blob detection. We choose this feature to describe the localization of LHWs, as the integration intensities of the line wings are higher than those in the surrounding regions. The integration intensity map is derived within the same seven channels as the subcube. We adopt \( \sigma_1 = 2 \) and \( \sigma_2 = 4 \), which means the feature is more sensitive in detecting structures with spatial dimensions of \~2′–4′.

2. A package in C was adopted in this work. Other versions include Pysvmlight and PySVMLight in Python, the CRAN package klaR in the R statistical software environment, and mex-svm in Matlab, etc., which are available at http://svmlight.joachims.org.
5. $T_{\text{outer}}/T_{\text{peak}}$, where $T_{\text{outer}}$ is the intensity averaged in the surrounding region with a radius of 4', where the subcube is masked in the calculations. The feature further constrains the emission around the LHWs and can help us locate the spatial peak position of an LHW, where this feature is expected to be lower than the first feature.

All these features were extracted from each pixel of both the prelabeled training set and unlabeled test set. The distributions of the features in the training set are shown in Figure 3. These features show different distributions and scales between classes. However, the different scales in the feature space may mislead the classifier to pay too much attention to a certain feature. To avoid this situation, we normalized the features according to the scales of the LHW class to make sure the features are equally weighted.

3.4. Optimizing with Cross-validation

The performance of a trained SVM model mainly depends on two free parameters: the trade-off factor $C$ and kernel parameter $\gamma$. Beaumont et al. (2011) illustrated the impacts of these parameters on the boundary in a two-dimensional feature space in detail. In our cases, $C$ determines the trade-off between the LHW and non-LHW class when misclassification occurs. A model trained with a high $C$ is more likely to find every pixel of the LHW class but meanwhile bring in more pixels of the non-LHW class because an LHW pixel misclassification introduces a higher penalty to the model than a non-LHW one. According to the study of Morik et al. (1999), the trade-off factor for imbalanced classes could be chosen to be near the ratio of the sample numbers of the non-LHW and LHW classes. After excluding the non-LHW pixels that touch the edge of feature space, we derived a ratio around 1.5 and tested three factors in our training process: 1.0, 1.5, and 2.0. The other parameter, $\gamma$, controls the curvature of the decision boundaries. A model trained with a higher $\gamma$ focuses on small scales in feature space and gives complicated boundaries to separate as many samples as possible, even if some of them are noise. As our feature space was normalized, we tested $\gamma$ of different magnitudes, ranging from 0.01 to 100.

Training models with cross-validation can provide the appropriate parameters to avoid underfitting and overfitting, which are common issues in machine learning. A small portion of the training set ($\sim$38%) is used as the validation set (dashed line box in Figure 1) in this controlled experiment, while the rest are regarded to be the training set. While traversing all parameter combinations, we trained models using the training set. The models were then applied to check whether LHWs could be retrieved in the training set and validation set. When underfitting occurs, a model that fails to catch the trend in the training set performs almost equally bad in the validation set. On the other end, an overfitted model performs well in the training set by capturing the noise in it, which is not generalizable to the validation set. By comparing all models, we eventually chose a relatively loose model with $C = 1.0$ and $\gamma = 10$ as it performed similarly in both sets, recovered all the LHWs in the validation set, and included an acceptable quantity of contaminations.

Learning curves are widely used to diagnose the generalizability of models from a training set incrementally. Figure 4 shows a form of learning curve in which we use the accuracy and F1 score (for more details, see the Appendix) to evaluate the performance of a model in a certain set. As an increasing number of random pixels is used for training, the two curves converge toward a relatively high accuracy and F1 score, which suggest neither underfitting nor overfitting is significant in our model. Furthermore, though our model may be improved by collecting more training data, such approach is likely to be inefficient. The robustness of our model could also be tested by repeating the cross-validation procedure using an incomplete training set. The feature distributions were redefined according to the data used for training. Ten experiments, in which 37% of the date are randomly removed from the training set, show that on average the same number of LHW candidates could be reproduced as when using a full set. Of the candidates, $\sim$93% overlap, which include all LHWs found by eye.

3.5. Prediction and Grading

The trained classifier was applied to predict all the pixels in our mapping region, including the training region. Because

![Figure 3. Feature distributions of the two classes in the training data set.](image-url)
each pixel was calculated independently, a large number of isolated pixels were accidentally classified as LHW. We prepared a filter to only retain the contiguous pixels occupying more than three pixels in the spatial dimensions (∼beam size) and spanning longer than 12 velocity channels (∼2 K km s⁻¹). These pixels were labeled and grouped into 960 LHW candidates in the region.

Then, we followed the method mentioned in Section 3.2 to examine each candidate carefully. In total, 2880 diagrams were checked. As a result, 127 candidates fulfilled all three criteria and were classified as grade A. A total of 290 candidates were classified as grade B, satisfying the first two criteria but either showing a smaller velocity shift (1 < Δv < 2 km s⁻¹) or having other nearby components with the same velocity in the P−V maps. The remaining 543 candidates, which are basically contaminations in our searching process, were classified as grade C and discarded. Blueshifted and redshifted LHWs were then paired according to their positions and velocities to come up with bipolar candidates. The paired LHWs should be close to a gas condensation and extend from the same velocity component of ¹³CO. To derive a reliable outflow catalog from the candidates, we only kept the bipolar candidates and monopolar ones of grade A, which will result in a higher bipolar ratio in our catalog. Monopolar candidates in the Outer Arm were further removed. Eventually, we identify 130 outflow candidates.

4. Result

4.1. List of Outflows

Among the 130 outflow candidates, 77 (59.2%) samples show bipolar structures; 29 (22.3%) and 24 (18.5%) samples are mono-blue and mono-red polar, respectively. In Table 1, we list all outflows detected in this work, with brief comments. Each outflow is referred to as a “Delingha Molecular Outflow Candidate” (DMOC). The position and velocity of the blueshifted and redshifted lobes are cataloged separately, with the coordinates representing the peak positions of the lobes. Some bipolar outflows contain more than two lobes, which indicate that there are multiple peaks in its blueshifted or redshifted lobe. The velocity range and grade of the outflows are the direct results of the outflow searching procedure. Because studies toward molecular clouds in our mapping region are limited, 118 outflows are detected among all detections using the molecular spectra line for the first time. We also detect the counterparts of two known monopolar outflows and classify them as bipolar outflows in our catalog. The spectra, integrated intensity map, and P−V diagram of two new detected outflows are shown in Figure 5 as examples.

In Figure 1, we map the locations of the identified outflows in each velocity component, where the blueshifted and redshifted lobes are shown by color-coded symbols. Most outflows are concentrated in the clouds with high column density. Outflows mainly cluster in the Gulf of Mexico (L935), Pelican’s Head (L933, L936), IC 5146 (L1040/1042/1055), G084.9−0.4, and Sh 2-24 region. Groups of outflows are arranged on the filamentary molecular cloud, such as L931, IC 5146, G084.9−0.4. There are 97, 31, and 2 outflows identified in the Local, Perseus, and Outer Arms, respectively.

As a potential driving source of outflows, YSOs are found to be associated with some of our molecular outflows. We list the class of YSOs within 3′ around each outflow in the 9th column of Table 1 where two YSO catalogs are included. We identified YSOs in our mapping region following the detection and classification scheme given by Koenig & Leisawitz (2014). Such scheme identifies Class I and II stars based on the near- and mid-infrared colors of WISE and 2MASS photometry. We further included the YSO catalog identified by Rebull et al. (2011) in the North American and Pelican Nebulae region to complement our sample. In our sample, 61 (~46.9%) are associated with Class I/flat YSOs, while 82 outflows (~63.1%) are located near Class II YSOs. Some outflows are located close to several YSOs, which may be of different classes. Observations with higher resolution are required to further distinguish the driving sources. Furthermore, 36 (~27.7%) outflows have no YSO associated, half of which are located in the Local Arm. It is evident that the sensitivities of infrared observations limit the detection of distant or deeply embedded YSOs as outflow-driving sources. We should also consider the possibility that some of the outflows are driven by protostars at an earlier stage than Class I. As pointed out by Mottram et al. (2017), Class 0 objects tend to present a higher outflow mass and momentum injection rate than those of Class I.

We list the associated infrared sources and outflow indicators from other bands in the last column of Table 1. There are 29 samples associated with IRAS point sources. Different types of shock indicators such as Herbig–Haro (HH) objects, molecular hydrogen emission-line objects (MHOs), and water masers are found in the lobes, improving the reliability of our detections (Bally & Reipurth 2003; Sunada et al. 2007; Armond et al. 2011; Toujima et al. 2011; Urquhart et al. 2011; Bally et al. 2014).

4.2. Physical Properties

It is important to derive accurate physical properties such as mass, momentum, and mechanical energy of the outflow, because they will help us to understand the feedback activities of forming stars to their immediate environment. Prior to the calculation, the distance of the outflows need to be settled. We use the distance estimator given by Reid et al. (2016), which provides the combined result of parallax and kinematic distance using a Bayesian approach. The position and velocity of ¹³CO for each outflow are extracted and taken as input parameters, and the output distance with the highest probability is designated. The calculated distance of Sh 2-124 agrees with
that reported by Foster & Brunt (2015). However, the calculated accuracy is limited for nearby clouds with velocity around 0 km s\(^{-1}\), and such method fails to converge at relatively high latitudes such as in IC 5146, L1042, and L1045. Green et al. (2015) provide a method based on Pan-STARRS 1 and 2MASS photometry to derive more accurate moduli if a cloud is associated with an optical or near-infrared dark cloud. Such method gives a distance of 0.9 kpc for those high-latitude clouds, which is similar to the distance of 0.95 kpc given by Harvey et al. (2008). However, Herbig & Dahm (2002) and Nunes et al. (2016) suggested a larger measurement of 1.1–1.4 kpc, which will multiply the calculated mass by a factor of 1.3–2.2.

We list the properties of the outflows in Table 2, in which the \( v_{\text{avg}} \) column presents the average velocity of the \(^{12}\text{CO} \) lobe relative to the systemic velocity of the corresponding \(^{13}\text{CO} \) emission. The average velocity ranges from 2.1 to 10.5 km s\(^{-1}\). The size of a lobe is given by measuring the extent along the Galactic longitude and latitude of the contour at half-maximum, while its length is measured from the 3\( \sigma \) contour to the nearest \(^{13}\text{CO} \) clump or YSO. While measuring the scale and velocity of the outflows, their inclinations are an intractable problem, for which several correction methods have been proposed in the literature (Dunham et al. 2014). Proper motion studies of HH objects provide the most reliable measurement of the inclination. However, such correction is impossible for our sample, as only a few lobes are associated with an HH object, and no proper motion study has been done. Therefore, no inclination correction is applied to the parameters listed in Table 2. If a mean inclination of 57°/3 is adopted (assuming all orientations are uniformly distributed in space), the average velocity and length will be scaled up by a factor of 1.85 and 1.19, respectively.

We further calculate the column density, mass, momentum, and energy of each lobe under the assumption of local thermodynamic equilibrium (LTE) by following the standard method given by Wilson et al. (2009) and Curtis et al. (2010). Without opacity correction, the column density can be calculated with

\[
\frac{N(^{12}\text{CO})}{\text{cm}^{-2}} = 4.17 \times 10^{13} \frac{T_{\text{ex}}}{K} \exp\left(-5.53 K/T_{\text{ex}}\right) \int T_{\text{mb}}(\nu) d\nu, \]

where \( T_{\text{ex}} = 10^{-4} \) (Ferking et al. 1982) and a mean molecular weight of 2.72 (Brunt 2010) are adopted to derive the column density of molecular hydrogen. The column densities in each velocity slice are then summed over each lobe to derive its mass, momentum, and energy, which are defined as

\[
M = \int m(\nu) d\nu; \]
\[
P = \int m(\nu) [v - v_0] d\nu; \]
\[
E = \frac{1}{2} \int m(\nu) [v - v_0]^2 d\nu. \]

For most of our candidates, the \(^{12}\text{CO} \) line wing emissions are assumed to be optically thin, as no \(^{13}\text{CO} \) line wing is detected within the velocity range of \(^{12}\text{CO} \) lobes except for a few. There are weak \(^{13}\text{CO} \) line wing features detected in the blueshifted lobe of DMOC-0025, 0072, 0123, and 0128 and the redshifted lobe of DMOC-0016 and 0072. Their optical depths could be estimated following the relation given by Hatchell et al. (1999):

\[
\frac{T_{\text{ex}}^{^{12}\text{CO}}}{T_{\text{ex}}^{^{13}\text{CO}}} = \left( \frac{\nu_{12}}{\nu_{13}} \right)^2 \frac{[^{12}\text{CO}]}{[^{13}\text{CO}]} \frac{1 - e^{-\tau_2}}{1 - e^{-\tau_1}},
\]

where \([^{12}\text{CO}] / [^{13}\text{CO}] = 62\) is the relative abundance between the two species (Langer & Penzias 1993). The resulting optical depths range from 6 to 10, which will bring a correction factor of \(\tau / (1 - e^{-\tau})\) to the parameters concerning mass. In the above calculations, we use a uniform excitation temperature of 25 K, which could be different from the real situation. A large range of excitation temperatures has been discussed in several

### Table 1
List of Detected Outflows

| Name     | Lobe | \( l \) (°) | \( b \) (°) | Velocity Range (km s\(^{-1}\)) | \( v_{\text{avg}} \) (km s\(^{-1}\)) | Grade | New Detection | YSO Class | Comments |
|----------|------|-------------|-------------|-----------------------------|---------------------------------|-------|---------------|-----------|---------|
| DMOC-0001 | Red  | 83.767      | +0.125      | (2.7, 6.5)                  | 0.3                             | A     | Y             | I/II      |         |
| DMOC-0002 | Blue | 83.850      | −2.092      | (−5.1, −2.9)                | 1.7                             | B     | Y             | II        |         |
| DMOC-0003 | Red  | 83.850      | −2.100      | (6.2, 9.5)                  | 1.7                             | B     | Y             | II        |         |
| DMOC-0004 | Red  | 83.875      | +0.092      | (3.3, 7.5)                  | 1.0                             | A     | Y             | ...       |         |
| DMOC-0005 | Blue | 83.942      | +0.783      | (−15.1, −9.4)               | −5.4                            | A     | Y             | ...       | 3C 423  |
| DMOC-0006 | Blue | 83.967      | +0.033      | (−5.1, −1.9)                | 0.6                             | A     | Y             | I/flat/II | IRAS 20472+4338 |
| DMOC-0007 | Blue | 83.967      | +0.050      | (3.2, 6.2)                  | 0.6                             | A     |               |           |         |
| DMOC-0008 | Blue | 84.017      | +0.008      | (−5.1, −1.4)                | 0.5                             | B     | Y             | I/II      |         |
| DMOC-0009 | Blue | 84.008      | +0.008      | (3.5, 6.2)                  | 0.5                             | B     |               |           |         |
| DMOC-0010 | Blue | 84.108      | −0.550      | (−7.5, −2.7)                | 1.3                             | A     | Y             | ...       |         |
| DMOC-0011 | Blue | 84.167      | −1.733      | (−3.5, −0.2)                | 2.1                             | A     | Y             | ...       |         |
| DMOC-0012 | Blue | 84.292      | +0.900      | (−10.3, −8.6)               | −3.8                            | A     | Y             | ...       | IRAS 20444+4425 |

Note. List of the identified molecular outflow candidates in this work. Columns are outflow name, lobe, position (Galactic longitude and latitude), velocity range, systemic velocity, grade assigned in searching procedure, newly detected outflow in this work or not, class of associated YSOs, and comments.

References. (1) Armond et al. (2011); (2) Bally & Reipurth (2003); (3) Bally et al. (2014); (4) Dunham et al. (2012); (5) Li et al. (2016); (6) Sunada et al. (2007); (7) Toujima et al. (2011); (8) Urquhart et al. (2011); (9) Wu et al. (2004).

(This table is available in its entirety in machine-readable form.)
studies such as Hirano & Taniguchi (2001) and Stojimirović et al. (2006). Assuming a lower temperature of 10 K, as in Downes & Cabrit (2007), will decrease the mass, momentum, and energy by a factor of 1.8, while a higher value of 50–100 K, as in Hirano & Taniguchi (2001), will increase these parameters by a factor of 1.8–3.4. It is also noteworthy that outflow emission at low velocity, which may mix with ambient gas, are omitted. At least a factor of 2 will be introduced to take such correction into consideration (Margulis & Lada 1985). Dunham et al. (2014) discussed the correction factors for the parameters in detail based on a line ratio analysis of low-J CO outflows and gave a higher mean factor of 7. As a result, the calculated mass, momentum, and energy are lower limits.

Figure 6 shows the distributions of physical properties, including average velocity, size, mass, momentum, energy, and dynamical timescale. Double-peak distributions are shown in all the diagrams with the exception of the average velocity, which is measured along the line of sight and unrelated to distance. By simply dividing the samples into groups with distance greater and less than 2 kpc, we note that those peaks represent outflows at different distances, and the parameters of nearby outflows seem to be lower than those of distant ones. Such divergence is illustrated more clearly by tendency of the lobe size and mass to vary with distance as shown in Figure 7. The double-peak distribution in Figure 6 could be the result of the uneven distribution of star-forming regions along the distance axis due to the spiral arms, because we only cover a short range of Galactic longitude. This trend can be explained as an observational effect wherein low-mass outflows in distant molecular clouds fall below the detection limit. Furthermore, the resolution of our observations is insufficient to resolve each outflow even in some nearby clouds, which indicates that some of our candidates might be stacking unresolved outflows from clusters or small groups of YSOs. Figure 7 also shows that the lower limit of mass for the identified outflows is much higher than the limit of our searching procedure, while most of the grade C candidates fall between the two limits. We could further estimate an outflow detection limit of \( m(M_{\odot}) > 0.16 \times D(\text{kpc})^2 \) for the MWISP survey.

The dynamical timescale calculated by dividing the lobe length with average velocity will usually underestimate the outflow age, as the low-J CO intensity of the outflow drops increasingly rapidly at high velocity (Downes & Cabrit 2007). Still, we could estimate the mean mass entrainment rate \( \dot{M}_{\text{out}} \) and momentum injection rate or outflow force \( F_{\text{CO}} \) to be \( 1.1 \times 10^{-5} M_{\odot} \text{ yr}^{-1} \) and \( 4.7 \times 10^{-5} M_{\odot} \text{ km s}^{-1} \text{ yr}^{-1} \), respectively, both of which give an uncertainty of ±0.5 dex. The obtained rates represent time-averaged results over a bulk of the shocked gas including a wide range of conditions, whereas the rates match the results given by Mottram et al. (2017).
4.3. Outflow Clusters and Individuals

4.3.1. Gulf of Mexico

The Gulf of Mexico region, located in the south of L935, is the most active star-forming region in the North America and Pelican Nebulae. A large number of shock features such as HH objects (Armond et al. 2011), MHOs (Bally et al. 2014; Makin & Froebrich 2018) and their possible drivers such as Class I, II, and III YSOs (Guieu et al. 2009; Rebull et al. 2011), and highly embedded Class 0 YSOs have been found in this region. Together with CO outflow candidates, we show the distributions of molecular clouds and outflow features in Figure 8. As in the labels on the figure, Bally et al. (2014) further subdivided the region into three subregions, which are the elongated “Gulf core” in the east, the compact “Gulf SW1” in the middle, and the gas-rich filaments “Gulf SW2” in the west.

The clumps in the “SW2” region exhibit large and strong C$^{18}$O and 1.1 mm emission as shown in Figure 8, and are therefore confirmed to be massive ones (Zhang et al. 2011). Although a great number of YSOs, as the outflow drivers, emerge in the clumps, shock features are rarely spotted in this region. Our study identified 14 molecular lobes, which show different spatial distributions from those of shocks. In the “SW2 main” region, Bally et al. (2014) paired the oppositely facing bow shocks as well as their HH counterparts to the north and south of the strongest clump in the whole region and designated it as MHO 3417, which spans ∼10$^3$ in total. Unlike the HH features that often extend for many parsecs, molecular outflows are generally observed close to the driving source (Bally 2016). We found that the blueshifted lobe of DMOC-0040, coupled with a weak redshifted one, is associated with the northern portion of MHO 3417. On the other hand, although the redshifted lobe of DMOC-0036 coincides with the southern jet, its gas component is blended with another outflow known as MHO 3418, driven by the southwest clump. Observations with higher resolution are needed to determine the spatial distributions and orientations of these lobes.

The “SW2 SE” and “SW2 S” regions are two sections of a discontinuous filament. A bipolar outflow, DMOC-0039, is revealed within the easternmost clump on the filament. The outflow extends along the east–west direction, and its blueshifted lobe is associated with HH 956 (Armond et al. 2011). However, such orientation is inconsistent with that of nearby MHOs, which suggest there are multiple outflows from different driving sources. Another active region on the filament is where it breaks and Class I YSOs emerge. The blueshifted lobe of DMOC-0032 is elongated and extends from the west end of “SW2 SE” to the east end of “SW2 S.” Such structure is likely to be unresolved outflows because it is associated with two groups of H$_2$ knots (MHO 3421/3422) next to two clusters of YSOs. DMOC-0028 is a bipolar outflow associated with a small clump to the north of the “SW2 S” region. The outflow presents clear line wing structures with rather strong emission and large average velocity, but the intensities of dense gas and dust in the clump are relatively weak.

Bally et al. (2014) found a long chain of H$_2$ shocks throughout the “SW1” region which could be driven by a dim Class 0 protostar in the most opaque portion of the clump. However, no outflow is identified by our procedure. The shocks present a large aspect ratio in distribution and clear bow shock features, which suggest that the inclination of the outflow could be large and most of the high-velocity components are parallel to the plane of the sky. On the other hand, after carefully checking the spectra, we find several velocity components converging near this region, which prevents the detection of low-velocity line wings.

The number density of shock features in the “core” region presents a clear enhancement compared with other subregions. Toujima et al. (2011) pointed out that this subregion has a higher star formation efficiency. Four molecular outflows lie along the filamentary clouds, which seems less scattered than the distributions of shock features. DMOC-0061 is a pair of lobes associated with a YSO cluster near the center of this region. Both lobes spatially coincide with the molecular component at ∼5 km s$^{-1}$ and present a similar average velocity. However, due to the complicated distribution of the YSOs, we cannot rule out the possibility that the lobes are driven by different sources. A bipolar outflow has been identified in “core E” by Green et al. (2011) with CO $J = 2−1$ observations. Dunham et al. (2012) using their SMA
observations confirmed such result, and further concluded that it is ejected by a Class 0 source embedded in the millimeter source MM3 rather than the FU Ori candidate HBC 722 nearby. The redshifted lobe, DMOC-0063, is \( \sim 1' \) to the northeast of "core E" in our observations, while the blueshifted counterpart is missing. The position of the red lobe matches the orientation of the outflow and the distributions of associated MHOs (Dunham et al. 2012; Bally et al. 2014).

4.3.2. Pelican’s Head

The Pelican’s Head region lies on the northeastern tip of the Pelican Nebula (IC 5070). As shown in Figure 9, molecular clouds in Pelican’s Head could be subdivided into two subregions: Pelican’s Hat to the east and Pelican’s Neck to the west. In Pelican’s Neck, there are four cometary clouds (designated as PN1–PN4) extending to the north behind the UV-excited edges mentioned by Bally et al. (2014). PN4 could be further subdivided into three condensations, designated as MM1–MM3.

The blueshifted lobe DMOC-0018 is associated with the strongest clumps of PN1, which is surrounded by MHOs and HH objects. Such outflow could be driven by a Class I YSO, J205041.02+441808.1, embedded in the clump identified by Rebull et al. (2011). Several shock features emerge around the two western clumps in PN2, but only the blueshifted lobe of DMOC-0021 is detected in the weaker one, which is associated

Figure 6. Distribution of average velocity (upper left), size (upper middle), outflow mass (upper right), momentum (lower left), energy (lower middle), and dynamical timescale (lower right). The distribution of the two subgroups, less and greater than 2 kpc, are shown with orange and green bars, respectively.

Figure 7. Mass and distance relation. The dashed blue line indicates the 3\( \sigma \) detection limit calculated based on our searching procedure. The dotted red line shows the detection limit based on the identified outflows.

Figure 8. Outflows in the Gulf of Mexico. Integrated intensity grayscale map of C\(^{18}\)O is overlaid with contours showing the 1.1 mm continuum emission from the Bolocam Galactic Plane Survey (BGPS). The spectra are integrated over \(-10\) to 10 km s\(^{-1}\). Blue and red circles indicate the positions of identified blue-/redshifted molecular outflow lobes. Purple crosses and X’s show the locations of MHOs and HH objects. The purple double-headed vector indicates the size and orientation of MHO 3417, the largest outflow in this region. The green stars, diamonds, and dots represent YSOs of Class I, flat, and II, respectively.
Figure 9. Outflows in the Pelican’s Head. The background image, contours, and symbols are the same as those in Figure 8. The red dashed lines outline the fluorescent edge of H₂ emission (Bally et al. 2014).

with a reflection nebula (Bally et al. 2014). Molecular emission in the stronger clump to the north blends with a weak component at 5.9 km s⁻¹, which prevents us from extracting the blueshifted line wing. Shock features emerge in each clump of PN4, while molecular outflows are only found in the northern two. The outflow associated with MM1, DMOC-0025, is a bipolar one, whose blueshifted lobe contains two peaks. Its orientation matches the position angle of the H₂ knots well (Bally et al. 2014). DMOC-0030, coinciding with MM2, shows a prominent redshifted lobe but no localized blueshifted line wing. Bally et al. (2014) reported an X-shaped outflow (MHO 3400) embedded in the dust clump, which is interpreted as precessing jets of the binary.

The molecular emission in the Pelican’s Hat has an elongated structure, most of which is infrared dark in the Spitzer 24 μm image. Several dusty clumps lie in the filament with similar spatial intervals of ~3'/5. Outflows in the Pelican’s Hat are relatively weak, with size and mass below average. DMOC-0044 is a bipolar outflow without a YSO association as the driving source. The X-ray observations conducted by Damiani et al. (2017) provide us with other possible driving candidates, which could be interpreted as weak-line T-Tauri stars missed by YSO searches. In the strongest dust core of this subregion, we detected another bipolar outflow, DMOC-0054, which shows a prominent blueshifted lobe. Its redshifted counterpart is weak and may be contaminated by a H II region, ~1' northeast of the outflow.

4.3.3. IC 5146

IC 5146 consists of clumps and filamentary structures across the Cocoon Nebula and the two Streamers as shown in Figure 10. The CO emission shows that the Cocoon Nebula is connected to the long filament in the Northern Streamer, while the Southern Streamer, which appears to be parallel with the Northern one, is attached to its opposite end through several subfilaments. Aside from their spatial connections, the velocities of the three subregions in IC 5146 are consistent, which provides strong evidence for their codistant property (Dobashi et al. 1993). The existence of variable stars, young pre-main-sequence (pre-MS) stars, Hα-emission stars, and 200 YSO candidates (Herbig & Dahm 2002; Harvey et al. 2008) is indicative of active star formation within the region.

The Cocoon Nebula separates the surrounding molecular cloud into three parts. DMOC-0126 is associated with a clump at the top of the northeast arc. Both of its lobes are marginally spotted and are assigned as grade B, whereas a C¹⁸O dense clump is detected just between them. The strong bipolar outflow, DMOC-0123, lying in the northwest part was first discovered by Levreault (1985). It spatially coincides with a Herbig Be star, BD+46°3471. However, the ZAMS distance to the pre-MS star is 355 pc, which is inconsistent with the cloud distance (~0.9 kpc) we derived and those in the literature (Walker 1959; Elias 1978; Harvey et al. 2008). Thus, the outflow candidate is either driven by the pre-MS star with a suspicious distance (Harvey et al. 2008; Johnstone et al. 2017), or ejected from other YSO candidates.

Unlike the Southern Streamer, which shows no sign of any line wing, the Northern Streamer possesses eight outflow candidates in its two parallel subfilaments and their junctions. DMOC-0116 is a known outflow reported by Myers et al. (1988). DMOC-0112, 0113, 0114, and 0117 were first identified by Dobashi et al. (1993) to be associated with four IRAS sources. Their results are reproduced by our observations in terms of position, morphology, and line profile of the lobes. As shown in Figure 11, Dobashi et al. (1993) identified a massive redshifted lobe and two marginally resolved blue-shifted counterparts designated as S-blue and N-blue near IRAS 21432+4719. Apart from recovering the lobes, a new blueshifted one is detected to the west, which is beyond the mapping coverage of Dobashi et al. (1993). S-blue seems to be symmetric with the redshifted lobe, but the new blueshifted one is stronger and closer to it. This is indicative of unresolved multiple outflows within the beam, and observations with
higher resolution are needed to resolve their origins. As for N-blue, it is highly possible that it originates from other dense clumps in another subfilament.

4.3.4. G084.9−0.4

The G084.9−0.4 region possesses 16 outflow candidates, which make it one of the most active region in the Perseus Arm. The consistency of the velocities of the rings (∼40 km s⁻¹) suggests that they are at the same distance. Though most molecular gas gathers in the southern ring, only five outflows are scattered in it. The most active cloud where outflows crowd is the broken arc to the north as shown in Figure 12. There are five candidates arranged in the middle portion of a filamentary cloud. Though it may be ambiguous to confirm outflows in a distant spiral arm, the bipolarity and distinct distributions between blueshifted and redshifted lobes illustrate that these line wings originate from outflows rather than other shock features, such as a H II region or supernova. The probed lobes, especially those in DMOC-0060, 0062, and 0066, show high mass (>10 M☉) and high velocity (5–10 km s⁻¹), and their sizes are comparable to the sizes of the associated clumps. There is another outflow, DMOC-0073, lying to the east of the arc, but the velocity of its associated cloud (∼−50 km s⁻¹) indicates that it might be at a farther distance.

4.3.5. Individuals

DMOC-0074 (V1057 Cyg)—this outflow is located to the east of the Gulf of Mexico, and is associated with a famous FU Ori star, V1057 Cyg. Only the blueshifted lobe has been detected by Levreault (1988). Its redshifted lobe uncovered in this work is weaker than its counterpart in intensity, but the lobes present a similar average velocity and good spatial symmetry.

DMOC-0069—This outflow is the most distant candidate though its distance estimation is doubtful. It originates from an isolated molecular cloud. Though both lobes are weak, they have similar properties and good symmetry.

DMOC-0077—Together with DMOC-0069, they are the only two candidates identified in the Outer Arm. The outflow is in a filamentary cloud protruding from a giant molecular cloud. Its lobes show similar intensity, but the blueshifted one has a higher average velocity (4.7 km s⁻¹) than the redshifted one (3.0 km s⁻¹).

5. Discussion

5.1. Comparing Properties with Other Outflow Surveys

Wu et al. (2004) conducted a large-scale statistical study of the outflows in the literatures available at that time. The properties measured for their samples cover a range orders of magnitude larger than ours and show large scatter. The sizes of the outflows are compared in Figure 13 as an example. The large scatter could result from their heterogeneous data, which include a large variety of outflow tracers, sensitivities, resolutions, and methods of size measurement. On the other hand, some scatter would be introduced into our samples if we include more line wings at low velocity that blend with ambient gas, as in these works. Although the sizes in both catalogs are similar within a distance of ∼2 kpc, the distant samples in our work present a slightly larger size than those in Wu et al. (2004). Moreover, a similar drop in angular size at larger distance, less severe for our samples, could be spotted, which
could be attributed to the limit size of unresolved outflow clusters. The mass, momentum, energy, and dynamical timescale share similar ranges and trends, increasing with distance for both catalogs. To reduce the discrepancy arising from different observational and detection methods, we pay further attention to some unbiased outflow surveys with the same tracer and almost the same resolution and sensitivity. Perseus is one of the regions where molecular outflows have been thoroughly searched for by Arce et al. (2010). Sixty new candidates are detected using their three-dimensional visualization method. The Taurus region is another well-studied example. Narayanan et al. (2012) and Li et al. (2015) both carried out outflow searches in the same data set from the FCRAO CO survey. Slight biases could be spotted in their methods. Narayanan et al. (2012) focused on high-velocity outflows (>3 km s\(^{-1}\)), and candidates in Li et al. (2015) are biased toward outflows driven by YSOs detectable in Spitzer infrared photometry, which means those associated with Class 0 and other embedded objects are missing. Another two unbiased studies, which are based on the MWISP large-scale CO survey as well, were carried out by Li et al. (2018, 2019). They found 198 and 459 outflow candidates toward the Gemini OB1 region and the W3/4/5 complex, respectively.

We extract and compare two fundamental properties, the lobe size and mass, which were commonly measured in the aforementioned literatures, but the methods used to measure the properties vary among different studies. The lobe sizes estimated at different contour levels are unified in size at the half-maximum contour by simply assuming a Gaussian column density profile, which means a size measured at 30% contour is multiplied by a factor of 0.76 for comparison. As for the mass, Arce et al. (2010) adopted a column density calculation method different from other studies to correct opacity. We managed to recalculate the masses of their samples following the method we used in Section 4.2. Other parameters such as excitation temperature, CO abundance, and mean molecular weight are also unified. Eventually, the correlation between the lobe sizes and masses is shown in Figure 14.

A tight power-law correlation appears in the plot for the unbiased studies, while such correlation is not significant in the samples from Wu et al. (2004), due to large scatter. It is worth noting that the properties of the points in Figure 14 are not corrected for inclination, opacity, excitation temperature, low-velocity outflow component, and high-velocity outflow emission below the sensitivity. We managed to correct the first four factors with Monte Carlo experiments. The correction factor and method are mentioned in Section 4.2. After adding factors

![Figure 14](image_url)
for 500 experiments, we use filled yellow contours to illustrate the number density of the expanded results in Figure 14. A larger intercept and extra scatter are introduced into the correlation, because masses in distant outflows are more likely to be underestimated due to beam dilution. More observations combined with theoretical studies are needed to measure all correction factors comprehensively. The correlation between size and mass is usually interpreted as self-similar fractal structures found in interstellar medium (ISM) on different spatial scales, spanning from giant molecular clouds (GMCs) to cores (Elmegreen & Falgarone 1996; Kauffmann et al. 2010; Lombardi et al. 2010; Roman-Duval et al. 2010). Such self-similarity also emerges in filamentary molecular clouds, as rich structures of subfilaments are commonly resolved in MWISP filament survey (Hacar et al. 2018). Mandelbrot & Whitrow (1983) pointed out that a fractal structure is naturally expected in turbulent ISM. As small-scale and high-velocity substructures of molecular clouds, low-J CO outflows could reflect the structures of molecular gas around the driving sources. Our correlation indicates that the self-similar structures of ISM may persist in outflows from the scale of outflow clusters to individual cavities, though jets may replace certain roles of turbulence in them.

When discussing fractals in ISM, the fractal dimension is usually used to describe how small-scale turbulent gas fills the upper-level space and forms self-similar hierarchical structures. A similarly defined fractal dimension (Roman-Duval et al. 2010), $D_M = 2.37 ± 0.03$, could be derived by applying an ordinary least-squares bisector linear fitting between the logarithms of sizes and masses shown in Figure 14. Though such a mass-based fractal dimension is a projection of the ISM in three-dimensional space, a simulation study carried by Sánchez et al. (2005) found that the value of the three-dimensional fractal dimension is quite close to $D_M$. A fractal dimension over 2 indicates either outflows or that their clusters are complicated spongy structures filling the three-dimensional spaces rather than simple sheets. Our fractal dimension is quite similar to the exponent derived from molecular cloud surveys (Roman-Duval et al. 2010; Colombo et al. 2019), as well as several GMCs (Elmegreen & Falgarone 1996), which is dominated by turbulence over parsec scales, but smaller than the case of M17, where a bright H II region provides extra ambient pressure to the molecular clouds (Stutzki & Guesten 1990; Elmegreen & Falgarone 1996). An observation using near-infrared extinction map presented by Lombardi et al. (2010) reported a smaller exponent of $≈2.0$ on parsec scales, dropping to $≈1.6$ on subparsec scales, which agrees with the value obtained by Kauffmann et al. (2010). The higher fractal dimension in the outflows might coincide with the simulation results reported by Whitney et al. (2013) that outflows raise the clumpiness in the clumps and may suggest that the outflows contain more turbulent motions than expected.

5.2. The Impact of Molecular Outflows

The large-scale search for outflows could help us assess the process of transferring material, momentum, and energy from outflows to their parent clouds. Table 3 shows the total mass, number, total mass, momentum, and energy of lobes. The “Cloud” section includes the area within the half-maximum contour line of $^{13}$CO emission, mass estimated using $^{13}$CO under the LTE assumption, and momentum and energy contained in turbulent motions.

| Name             | Distance (kpc) | $n$ | $M_{\text{flow}}$ ($M_\odot$) | $P_{\text{flow}}$ ($M_\odot$ km s$^{-1}$) | $E_{\text{flow}}$ ($10^{44}$ erg) | Area (pc$^2$) | $M_{\text{cloud}}$ ($10^2 M_\odot$) | $P_{\text{turb}}$ ($10^2 M_\odot$ km s$^{-1}$) | $E_{\text{turb}}$ ($10^{51}$ erg) |
|------------------|----------------|----|-------------------------------|------------------------------------------|----------------------------------|---------------|-----------------------------------|---------------------------------|----------------------------------|
| L931             | 0.71           | 6  | 5.2                           | 19.2                                     | 0.8                              | 4.3           | 1.0                               | 3.1                             | 2.1                              |
| Gulf of Mexico   | 0.60           | 28 | 34.4                          | 198.5                                    | 12.6                             | 5.8           | 5.7                               | 21.2                            | 10.8                             |
| Pelican’s Neck   | 0.60           | 13 | 50.9                          | 195.5                                    | 8.8                              | 3.5           | 3.2                               | 11.8                            | 7.3                              |
| Pelican’s Hat    | 0.60           | 10 | 16.9                          | 60.7                                     | 2.6                              | 4.6           | 1.6                               | 5.2                             | 2.7                              |
| L941             | 0.45           | 6  | 26.1                          | 89.7                                     | 3.4                              | 2.5           | 0.7                               | 4.7                             | 5.2                              |
| IC 5146 Streamer | 0.45           | 14 | 6.5                           | 29.4                                     | 1.5                              | 2.5           | 1.5                               | 4.0                             | 3.2                              |
| IC 5146 Cocoon   | 0.45           | 4  | 6.1                           | 13.9                                     | 0.4                              | 0.4           | 0.3                               | 0.7                             | 0.5                              |

Notes. Comparison of properties between outflows and molecular clouds. The first two columns list the name and distance of the regions. The “Outflow” section gives the number, total mass, momentum, and energy of lobes. The “Cloud” section includes the area within the half-maximum contour line of $^{13}$CO emission, mass estimated using $^{13}$CO under the LTE assumption, and momentum and energy contained in turbulent motions.

L931 is not fully covered in our observations, thus the parameters are lower limits.
However, the estimation is rough as we need to further consider the timescale of energy injection and dissipation, because turbulent energy dissipates rapidly in molecular clouds (Mac Low & Klessen 2004). Mac Low (1999) related the timescale of turbulence dissipation to the freefall time in their numerical computations and gave $t_{\text{diss}}/t_{\text{ff}} \approx 3.9 \kappa / M_s$, where $\kappa$ is the ratio between the turbulence-driving length and Jeans length, and $M_s$ is the Mach number of turbulence. The driving length of a continuous outflow is approximately the lobe length, as shown in numerical simulations (Nakamura & Li 2007; Cunningham et al. 2009). By using a gas temperature of $\sim 10$ K and mean density of $1.1 \times 10^4$ cm$^{-3}$, we could derive a sound speed of $0.3$ km s$^{-1}$, a Jeans length of $0.3$ pc, and a freefall time of $0.36$ Myr. We could then deduce a dissipation timescale of around $0.31$ Myr, which is larger than the mean dynamical timescale of the outflow by a factor of 2.5. Thus, after considering the timescale, the spatial scale, in which the dissipation rate of turbulent energy is equivalent to the energy injection rate of a lobe, turns into $\sim 0.19$ pc. This result confirms the prediction of Brunt et al. (2009) that outflow is important in providing turbulent energy on core or clump scales on short timescales.

5.3. Missing Outflow Candidates

Within our mapping region, two outflows reported in the literature are not detected. One instance, associated with the molecular core B361 (Beichman et al. 1986), was found by Wu et al. (1992) in their CO (2−1) survey. No sign of a line wing could be found in our $J = 1−0$ transition. The other one near IRAS 21428+4732 was found by Dobashi et al. (2001) in their CO (1−0) observations with the 45 m telescope at Nobeyama Radio Observatory. Due to beam dilution, only the blueshifted lobe is marginally detected in our observations. Thus, as a monopolar grade B sample, it is not included in our catalog. The failure in detecting these two sources reminds us of possible factors that may affect the result of outflow searches.

The tracer, observational sensitivity, and resolution are the dominant factors that limit the amount of protostellar outflows detected in the molecular cloud, but it is difficult to estimate how many outflows may be missing. We looked up the record of YSOs and detected molecular outflows in two nearby star-forming regions, the Perseus and Taurus regions, from the literature. The amount of detected molecular outflows is a factor of $\sim 7$ less than the number of corresponding YSOs (Evans et al. 2009; Arce et al. 2010; Curtis et al. 2010; Rebull et al. 2010; Li et al. 2015). For NGC 1333, a subregion in Perseus, Bally (2016) pointed out that the number of known flows is less than that of YSOs by a factor of at least 4. If we adopt the factor of 4 for the North America and Pelican Nebulae where YSOs are found in detail by Rebull et al. (2011), there should be $>60$ outflows in the Gulf of Mexico or the Pelican region. Such estimated amount is four times more than the outflows we actually detected. An obvious reason is that most outflows are not resolved, especially in the outflow clusters, as our spatial resolution ($\sim 0.15$ pc) is 2−4 times larger than those in Perseus and Taurus. Another important reason is the different compositions of YSOs, because the outflow detection rates vary in different classes of YSOs (Li et al. 2015).

Other than the observational factors, outflow search methods also determine the results and may accidentally miss some particular samples. Most outflow studies based on low-$J$ CO (Arce et al. 2010; Li et al. 2015; Li et al. 2018, and this work) use the combination of $^{12}$CO and $^{13}$CO to probe line wing and cloud structure, respectively, which means that similar features that emerged on $^{13}$CO or C$^{18}$O lines are usually missing. The detection rate might be reduced when the $^{12}$CO line width is heavily broadened or multiple velocity components are blended. On the other hand, unlike higher $J$ transitions of CO which are better at tracing the high-velocity component of outflow, the $J = 1−0$ transition requires the detection of $^{13}$CO to estimate the velocity dispersion in ambient gas. Consequently, outflow cannot be detected if there is no associated $^{13}$CO emission. Moreover, to improve the reliability of detection, we discard the monopolar grade B and all grade C samples, because they could be highly contaminated by other kinematic structures such as expanding H II regions, supernova remnants, fragmentation in clumps, velocity gradient on filaments, and even emission from ambient gas. And as a result, we cannot rule out the possibility that weak outflows are within them, and inevitably miss them.

6. Summary

We have conducted a survey of molecular outflows for a complex of dark clouds in the Cygnus region using $^{12}$CO, $^{13}$CO, and C$^{18}$O molecular line maps from the MWISP survey. Our 46.75 deg$^2$ maps cover a large number of cold dark clouds and provide abundant data to study outflow properties and feedback during the formation of stars. We developed a searching method that combines a machine-learning algorithm, SVM, with traditional visual inspection to efficiently recognize outflow line wing features in position-position-velocity space. Our main results are summarized as follows.

1. A total of 130 outflow candidates are identified within the mapping area, and 118 of them are new detections. Outflows usually emerge in dense molecular clouds and form clusters or linear structures along molecular filaments according to their environments. There are 97, 31, and 2 candidates located in the Local, Perseus, and Outer Arms, respectively.

2. Of the outflows in our catalog, 59.2%, 22.3%, and 18.5% are bipolar, mono-blue, and mono-red polar, respectively. The mean size of our outflows is 0.54 pc, and the mean mass is 0.72 $M_\odot$ in the Local Arm. Some outflows show shock indicators such as HH objects, MHOs, and water masers. Most outflows may be driven by YSOs, while there are still 36 candidates not associated with any detectable YSO.

3. Molecular clouds and clusters of outflows are revealed in several star-forming regions, which reflect the diverse amount of star-forming activities. Outflows shown by CO also present a different distribution from that traced by shock features.

4. The outflow properties we measured agree with those in the literature. A tight power-law correlation between the lobe sizes and masses appears when comparing samples of different works, which suggests self-similar structures emerge in outflows. A fractal dimension of $2.37 \pm 0.03$, which is higher than that in clumps, indicates that turbulence brought by outflows may raise the clumpiness in molecular clumps.

5. By comparing the energy injected by outflows and turbulent energy on different spatial scales, we note that
outflow is insufficient to energize cloud-scale turbulence, but may be important to support turbulence on core or clump scales on short timescales.

This work is based on MWISP data acquired with the Delingha 13.7 m telescope of the Purple Mountain Observatory. The authors appreciate all the staff members of the observatory for their help with the acquisition and reduction of the data. Our gratitude also goes to the MWISP team for their support in operating the project. This work is supported by National Key Research & Development Program of China grant No. 2017YFA0402700; Key Research Program of Frontier Sciences, CAS, grant No. QYZDJ-SSW-SLH047; the National Natural Science Foundation of China through grants NSF 11803091, 11873093, 11773077, 11873019, 1673066, and 11629302; and by the Key Laboratory for Radio Astronomy, CAS.

Software: GILDAS/CLASS (Pety 2005; Gildas Team 2013), SVMlight (Joachims 1999), IDL Astronomy Library (Landsman 1993).

Appendix

Accuracy and F1 Score

When applying a classification algorithm to a data set prelabeled with positive and negative tags, four situations appear in the predictions:

1. True positive (TP): positive sample predicted as positive
2. False negative (FN): negative sample predicted as negative
3. False positive (FP): negative sample predicted as positive
4. True negative (TN): negative sample predicted as negative

A straightforward metric to evaluate the performance of the classifier is the accuracy:

\[
\text{Accuracy} = \frac{(TP + TN)}{\text{SampleSize}}.
\]

However, accuracy becomes insensitive to imbalanced classes in which positive samples are far less than negative ones, especially when a study is focusing on the positive ones. Thus, two more metrics are introduced:

\[
\text{Precision} = \frac{TP}{(TP + FP)}; \\
\text{Recall} = \frac{TP}{(TP + FN)}.
\]

In our outflow search study, “precision” indicates the proportion of how many actual LHW pixels (labeled with positive) are among all selected candidates (predicted as positive), while “recall” indicates the proportion of how many actual LHW pixels are retrieved from the data set. The relationship between precision and recall for our models is shown in Figure A1. The area under the curve represents the overall capability of a model. Apparently, a precise model tends to give low recall rate, and precision cannot be guaranteed by a model that recalls all samples. To balance the trade-off between precision and recall, the F1 score could be defined as

\[
F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},
\]

which makes it a more useful metric than accuracy to assess classification models for imbalanced problems.

Figure A1. Trade-off between precision and recall. Each curve represents a model trained from a different sample size.

References

Arce, H. G., Borkin, M. A., Goodman, A. A., Pineda, J. E., & Halle, M. W. 2010, ApJ, 715, 1170
Arce, H. G., Shepherd, D., Gaeth, F., et al. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. Arizona Press), 445
Armond, T., Reipurth, B., Bally, J., & Aspin, C. 2011, A&A, 528, A125
Bally, J. 2016, ARA&A, 54, 491
Bally, J., Ginsburg, A., Probst, R., et al. 2014, AJ, 148, 120
Bally, J., & Lada, C. J. 1983, ApJ, 265, 824
Bally, J., & Reipurth, B. 2003, AJ, 126, 893
Bally, J., & Scoville, N. Z. 1980, ApJ, 239, 121
Beaumont, C. N., Williams, J. P., & Goodman, A. A. 2011, ApJ, 741, 14
Beichman, C. A., Myers, P. C., Emerson, J. P., et al. 1986, ApJ, 307, 337
Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 359
Brunt, C. M. 2010, A&A, 513, A67
Brunt, C. M., Heyer, M. H., & Mac Low, M. M. 2009, A&A, 504, 883
Colombo, D., Rosolowsky, E., Duarte-Cabral, A., et al. 2019, MNRAS, 483, 4291
Cunningham, A. J., Frank, A., Carroll, J., Blackman, E. G., & Quillen, A. C. 2009, ApJ, 692, 816
Curtis, E. I., Richer, J. S., Swift, J. J., & Williams, J. P. 2010, MNRAS, 408, 1516
Damiani, F., Pilippitteri, I., & Prisinzano, L. 2017, A&A, 602, A115
Dobashi, K. 2011, PASJ, 63, S1
Dobashi, K., Bernard, J.-P., Yonekura, Y., & Fukui, Y. 1994, ApJS, 95, 419
Dobashi, K., Onishi, T., Iwata, T., et al. 1993, AJ, 105, 1487
Dobashi, K., Uchida, H., Kandori, R., et al. 2005, PASI, 57, 1
Dobashi, K., Yonekura, Y., & Hayashi, Y. 2001, PASI, 53, 811
Downes, T. P., & Cabrit, S. 2007, A&A, 471, 873
Dunham, M. M., Arce, H. G., Bourke, T. L., et al. 2012, ApJ, 755, 157
Dunham, M. M., Arce, H. G., Mardones, D., et al. 2014, ApJ, 783, 29
Elias, J. H. 1978, ApJ, 223, 859
Elmegreen, B. G., & Falgarone, E. 1996, ApJ, 471, 816
Elmegreen, B. G., & Scalo, J. 2004, ARA&A, 42, 211
Evans, N. J. I., Dunham, M. M., Jørgensen, J. K., et al. 2009, ApJS, 184, 321
Federrath, C., Schrön, M., Banerjee, R., & Klessen, R. S. 2014, ApJ, 790, 128
Foster, T., & Brunt, C. M. 2015, AJ, 150, 147
Frank, A., Ray, T. P., Cabrit, S., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 451

ORCID iDs

Shaobo Zhang @ https://orcid.org/0000-0003-2549-7247
Ji Yang @ https://orcid.org/0000-0001-7768-7320
Yang Su @ https://orcid.org/0000-0002-0197-470X
Yan Sun @ https://orcid.org/0000-0002-3904-1622
Xin Zhou @ https://orcid.org/0000-0003-2418-3520
