



# Numerical Simulation and Completeness Survey of Bubbles in the Taurus and Perseus Molecular Clouds

Mengting Liu<sup>1,2</sup> , Di Li<sup>1,2,3</sup> , Marko Krčo<sup>1</sup>, Luis C. Ho<sup>4,5</sup> , Duo Xu<sup>6</sup>, and Huixian Li<sup>1</sup>

<sup>1</sup> CAS Key Laboratory of FAST, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

[liumengting@nao.cas.cn](mailto:liumengting@nao.cas.cn)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China; [dili@nao.cas.cn](mailto:dili@nao.cas.cn)

<sup>3</sup> NAOC-UKZN Computational Astrophysics Centre, University of KwaZulu-Natal, Durban 4000, South Africa

<sup>4</sup> Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China

<sup>5</sup> Department of Astronomy, School of Physics, Peking University, Beijing 100871, People's Republic of China

<sup>6</sup> Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

Received 2019 June 1; revised 2019 September 24; accepted 2019 September 25; published 2019 November 7

## Abstract

Previous studies have analyzed the energy injection into the interstellar matter due to molecular bubbles. They found that the total kinetic energies of bubbles are comparable to, or even larger than, those of outflows but still less than the gravitational potential and turbulence energies of the hosting clouds. We examined the possibility that previous studies underestimated the energy injection due to being unable to detect dim or incomplete bubbles. We simulated typical molecular bubbles and inserted them into the <sup>13</sup>CO Five College Radio Astronomical Observatory maps of the Taurus and Perseus Molecular Clouds. We determined bubble identification completeness by applying the same procedures to both simulated and real data sets. We proposed a detectability function for both the Taurus and Perseus molecular clouds based on a multivariate approach. In Taurus, bubbles with kinetic energy less than  $\sim 1 \times 10^{44}$  erg are likely to be missed. We found that the total missing kinetic energy in Taurus is less than a couple of  $10^{44}$  erg, which only accounts for around 0.2% of the total kinetic energy of identified bubbles. In Perseus, bubbles with kinetic energy less than  $\sim 2 \times 10^{44}$  erg are likely to be missed. We found that the total missing kinetic energy in Perseus is less than  $10^{45}$  erg, which only accounts for around 1% of the total kinetic energy of identified bubbles. We thus conclude that previous manual bubble identification routines used in Taurus and Perseus can be considered to be energetically complete. Therefore, we confirm that energy injection from dynamic structures, namely outflows and bubbles, produced by star formation feedback are sufficient to sustain turbulence at a spatial scale from  $\sim 0.1$  to  $\sim 2.8$  pc.

*Key words:* ISM: bubbles – ISM: jets and outflows – ISM: kinematics and dynamics

## 1. Introduction

Stellar feedback plays a crucial role in the dynamics and energy balance of the interstellar medium (ISM; Zinnecker & Yorke 2007). Feedback associated with protostars injects momentum and energy to the parent molecular cloud, altering the velocity field and density distribution of the cloud (Arce & Goodman 2001, 2002; Arce et al. 2010), contributing to the mass loss of the surrounding dense gas (Fuller & Ladd 2002; Arce & Sargent 2006; Arce et al. 2010; Offner & Arce 2015), affecting star formation efficiency (Frank et al. 2014; Federrath 2015; Kroupa et al. 2018), inducing changes in the chemical composition of the impacted media (Bally et al. 2007), sustaining or generating turbulence (e.g., Fukui et al. 1986; Arce & Goodman 2001; Arce et al. 2011; Nakamura et al. 2011, 2011; Plunkett et al. 2013; Federrath 2015; Feddersen et al. 2018—henceforth Ar11, Fe18, see also Li et al. 2015—henceforth L15), resisting gravitational collapse and even disrupting the surrounding gas to limit the lifetime of their parent molecular cloud (Solomon et al. 1981; Arce & Goodman 2001; Hartmann et al. 2001; Arce & Goodman 2002; Duarte-Cabral et al. 2012; Plunkett et al. 2013).

The primary manifestations of stellar feedback are molecular outflows and bubbles. On the Galactic scale, superbubbles are shown to be globally important, with structures spanning hundreds of parsecs, such as chimneys and large cavities (Heiles 1979; Norman & Ikeuchi 1989). Abundant studies of outflows have been carried out (Kwan & Scoville 1976; Snell et al. 1980;

Lada & Harvey 1981; Arce et al. 2010; Nakamura et al. 2011, 2011; Narayanan et al. 2012; Mottram et al. 2017). In local to main star-forming molecular cloud complexes, relatively less attention has been paid to “local” molecular bubbles that are hollowed out by low- to intermediate-mass stars. Despite being expected to be less influential than superbubbles, they are substantial in number and could inject much more energy back into their natal clouds than that from outflows (Ar11, L15, Fe18).

Molecular bubbles are partially or fully enclosed three-dimensional structures whose projections on the sky resemble partial or full rings (Churchwell et al. 2006). Young stellar objects have sufficient stellar winds to entrain and accelerate ambient gas to sculpt spherical or ring-like cavities in their surrounding molecular clouds. Compared to collimated outflows, bubbles affect a larger volume of ambient molecular gas. Compared to supernova remnants, bubbles occur around more stars and persist for a longer period of time (Matzner 2002; Arce et al. 2011). In the Taurus molecular cloud complex (TMC), the kinetic energy and energy injection rate of identified bubbles are larger than those of outflows, implying a substantial input of mechanical energy into the TMC (L15). Due to their complex morphology, bubbles are more difficult to identify (Beaumont et al. 2014) in comparison to the relatively clear characteristics of outflows. In this work, we examine the completeness of bubble identification from previous studies.

The Five College Radio Astronomical Observatory (FCRAO) CO survey (Goldsmith et al. 2008) remains one of the largest

molecular spectral-line maps of a continuous cloud complex. Following an empirical and iterative procedure, L15 identified 37 bubbles in the TMC. Meanwhile, Ar11 detected 12 bubbles in the Perseus molecular cloud complex (PMC). The identified Taurus bubbles inject  $\sim 9.2 \times 10^{46}$  erg into the surrounding ISM. The energy injection rate was measured to be  $\sim 6.4 \times 10^{33}$  erg  $s^{-1}$ , surprisingly comparable to the turbulence dissipation rate. In the PMC, the kinetic energy of identified bubbles is around  $\sim 7.6 \times 10^{46}$  erg with an energy input rate of around  $\sim 8.9 \times 10^{33}$  erg  $s^{-1}$ , which is similar to the turbulence dissipation rate. Small or slowly expanding bubbles may be missed. However, large or irregular bubbles may also be missed due to confusion. We approach this problem with semiempirical numerical simulations of artificial molecular bubbles based on the bubble model from Cazzolato & Pineault (2005). We generated artificial spherical and partially spherical  $^{13}\text{CO}$  bubbles and embedded them into the real data cubes of Taurus and Perseus. We then carried out parameter studies including average antenna temperature, number of pixels, and expansion velocity to examine the completeness of the empirical identification procedures.

In Section 2, we provide a basic description of the Taurus and Perseus data sets and the bubble identification routines of previous surveys. The bubble model is presented in Section 3, including a detailed description of our method for simulating bubbles and artificial injection into the data sets. In Section 4, we describe the bubble detection procedures and quantify detectability. The properties of missing bubbles and the implication for stellar feedback in Taurus and Perseus are discussed in Section 5. In Section 6, we discuss and summarize the conclusions which may be drawn from this research.

## 2. Data and Empirical Procedures of Bubble Identification

$^{13}\text{CO}(1-0)$  is a tracer frequently used to trace molecular hydrogen and study the dynamic morphology of the ISM due to its much smaller opacities than those of  $^{12}\text{CO}(1-0)$ . The Taurus  $^{13}\text{CO}(1-0)$  and the Perseus  $^{13}\text{CO}(1-0)$  data sets we used are from the FCRAO CO surveys conducted between 2003 and 2005 (Ridge et al. 2006; Goldsmith et al. 2008; Narayanan et al. 2008; COMPLETE team 2011).

The common empirical procedures used to identify bubble structures in molecular clouds are as follows:

1. To search the regions with circular (or arc) structures brighter than the surrounding molecular gas (Ar11, L15, Fe18).
2. To visualize the high-velocity features of these regions by plotting a position–velocity ( $P$ – $V$ ) diagram. If there is an expanding bubble with a redshifted part or a blueshifted part, we can identify the  $\cup$  or  $\cap$ -shaped feature on the  $P$ – $V$  diagram of each region (Ar11, L15, Fe18).
3. To search for candidate sources such as pre-main-sequence stars or protostars associated with the cloud (Ar11, L15, Fe18).
4. To compare the bubble structures in  $^{13}\text{CO}(1-0)$  with an infrared map to determine whether they have similar morphologies (Ar11, L15, Fe18).

The more conditions identified bubbles satisfy, the more likely they are to be real dynamic feedback driven by protostars rather than superpositions of unrelated patterns in the cloud. The radius and thickness of each bubble can be estimated by fitting a Gaussian to the azimuthally averaged profile of the CO integrated intensity map (Ar11, L15).

## 3. Comparison of Simulated Bubble and Identified Bubble

We simulated spherical and partially spherical  $^{13}\text{CO}(1-0)$  bubbles based on the model proposed by Cazzolato & Pineault (2005) for the TMC and the PMC. This model relies on three fundamental assumptions:

1. The density of the surrounding medium and within the bubble is different but homogeneous.
2. The bubble expansion is isotropic.
3. All pixels on the bubble are at the same distance from Earth.

There are eight significant parameters in this model, listed in Table 1. Since identified bubbles in Taurus and Perseus often have partial circular or arc-like morphology, we employed completeness of bubble  $\beta$  as another parameter. The detailed bubble simulation calculation is shown in the Appendix A.

These artificial bubbles were embedded individually into the Taurus and Perseus  $^{13}\text{CO}(1-0)$  data cubes at random positions and channels. We compared inserted artificial bubbles with real identified bubbles by morphology and high-velocity features in Figures 1–2. Many of the identified bubbles in Taurus are ambiguous and difficult to identify, as in the right column of Figure 1. The identified bubbles in Perseus tend to have more prominent and identifiable apparent characteristics.

We simulated artificial bubbles with a similar radius, thickness,  $\beta$ , and expansion velocity for each cloud, as shown in the left columns of Figures 1–2. Circular structures brighter than the surrounding molecular gas may be identified in the intensity and channel maps. Meanwhile, the  $P$ – $V$  diagram illustrates the  $\cup$ - or  $\cap$ -shaped features. We found that in general artificial bubbles are somewhat easier to detect than real ones because they are not really embedded within the surrounding environment but are added on to the original images, and their morphology is necessarily independent of the surrounding environment. This means that our detectability function actually presents an upper limit for the bubble detectability function with the same parameters. We found it challenging to simulate realistic artificial bubbles in part to the ambiguous morphology of identified bubbles.

## 4. Bubble Detection Experiment

Identifying the circular structure in the channel maps is the first step in bubble identification. Therefore, we searched for the artificial bubbles by going through the channel maps. When we detected a circular structure, we compared the position of detected structure with the position of the inserted artificial bubble. If they are consistent with each other, we consider the artificial bubble detectable. When we cannot detect any circular structure through channel maps, we consider the artificial bubble undetectable.

We generated 500 artificial bubbles with different parameters for each cloud. Each bubble was embedded into Taurus and Perseus separately in random positions and channels one at a time. We undertook a blind search for each inserted simulated bubble to determine the detectability function. The artificial bubble identification routine largely follows the previous bubble manual identification procedures in Section 2.

### 4.1. Description of Variable Parameters

Number of pixels  $N$ , average antenna temperature  $T_a$ , and expansion velocity  $V_e$  are three observable parameters which

**Table 1**  
Artificial Bubble Parameters<sup>a</sup>

Parameters	Definition	Taurus	Perseus
$R$	Bubble radius	[0.28 pc–1.90 pc]	[0.14 pc–2.79 pc]
$\Delta R$	Bubble thickness	[0.03 pc, 0.44 pc]	[0.1 pc, 0.68 pc]
$V_{\text{exp}}$	Bubble expansion velocity	[1.0 km s <sup>-1</sup> –3.3 km s <sup>-1</sup> ]	[1.2 km s <sup>-1</sup> –6.0 km s <sup>-1</sup> ]
$\sigma$	Velocity dispersion	$\sim 1.4$ km s <sup>-1</sup>	$\sim 2.0$ km s <sup>-1</sup>
$\beta$	The completeness of bubble	[90°, 360°]	[90°, 360°]
$n'_0$	Bubble H <sub>2</sub> number density with assumed <sup>13</sup> CO abundance of $1.43 \times 10^{-6}$ . This is estimated based on intensity of the high-velocity line wings and used to calculate the amount of gas being moved by bubbles (see more explanation in the Discussion section).	35 cm <sup>-3</sup>	280 cm <sup>-3</sup>
$\Delta V$	Velocity width of a spectrometer channel	0.266 km s <sup>-1</sup>	0.066 km s <sup>-1</sup>
$a_{\text{pc}}$	The bubble physical depth along the LOS in parsecs.	Calculations are presented in Appendix A of Cazzolato & Pineault (2005)	

**Note.**

<sup>a</sup> Parameter name, definition, values for Taurus, and values for Perseus. First five parameters are measured values from L15 and Ar11, while the last four parameters are adopted values from calculation and observation property.

affect the bubble detection. The number of pixels  $N$  depends on the bubble radius, thickness  $\Delta R$ , and completeness  $\beta$ . The values of the parameters of each of the 500 experimental bubbles are generated randomly from the range of the maximum and minimum value of the observed bubble with uniform distribution to parameterize the bubble detectability in the whole parameter space equally. The experimental bubbles were inserted into the Taurus and Perseus <sup>13</sup>CO data cubes in random positions and channels one at a time. If we cannot detect the inserted artificial bubble, the detection result of this bubble is 0, otherwise, it is 1.

#### 4.2. Analysis Method

The previous bubble identification procedures are subjective and depending on the apparent detection threshold resulting from the bubble identification procedures in L15, Ar11, and Fe18. We quantified the bubble detection results of previous bubble identification procedures with the detectability function. Since the bubble detection results correspond to a Bernoulli distribution, which is one of the exponential family distributions, and the experimental bubble parameters can be considered as independent to each other, we constructed the bubble detectability function using generalized linear models (GLMs; Nelder & Wedderburn 1972). GLMs are widely used regression models for dependent variables which follow an exponential family distribution, such as Gaussian distribution, Poisson distribution, and chi-squared distribution. Derivation of the detectability function is discussed in Appendix B. The detectability function is in the form of

$$h(\mu_{T_a}, \mu_N, \mu_{V_e}) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1^* \mu_{T_a} + \alpha_2^* \mu_N + \alpha_3^* \mu_{V_e})}}, \quad (1)$$

where  $\alpha_0$  to  $\alpha_3$  are constants that need to be fitted,  $\mu_{T_a}$ ,  $\mu_N$ , and  $\mu_{V_e}$  are the bubble average antenna temperature, number of pixels, and expansion velocity divided by their maximum values (see Appendix B). The maximum value of  $T_a$ ,  $N$ , and  $V_e$  are obtained from real observed bubbles.

#### 4.3. Parametric Detectability

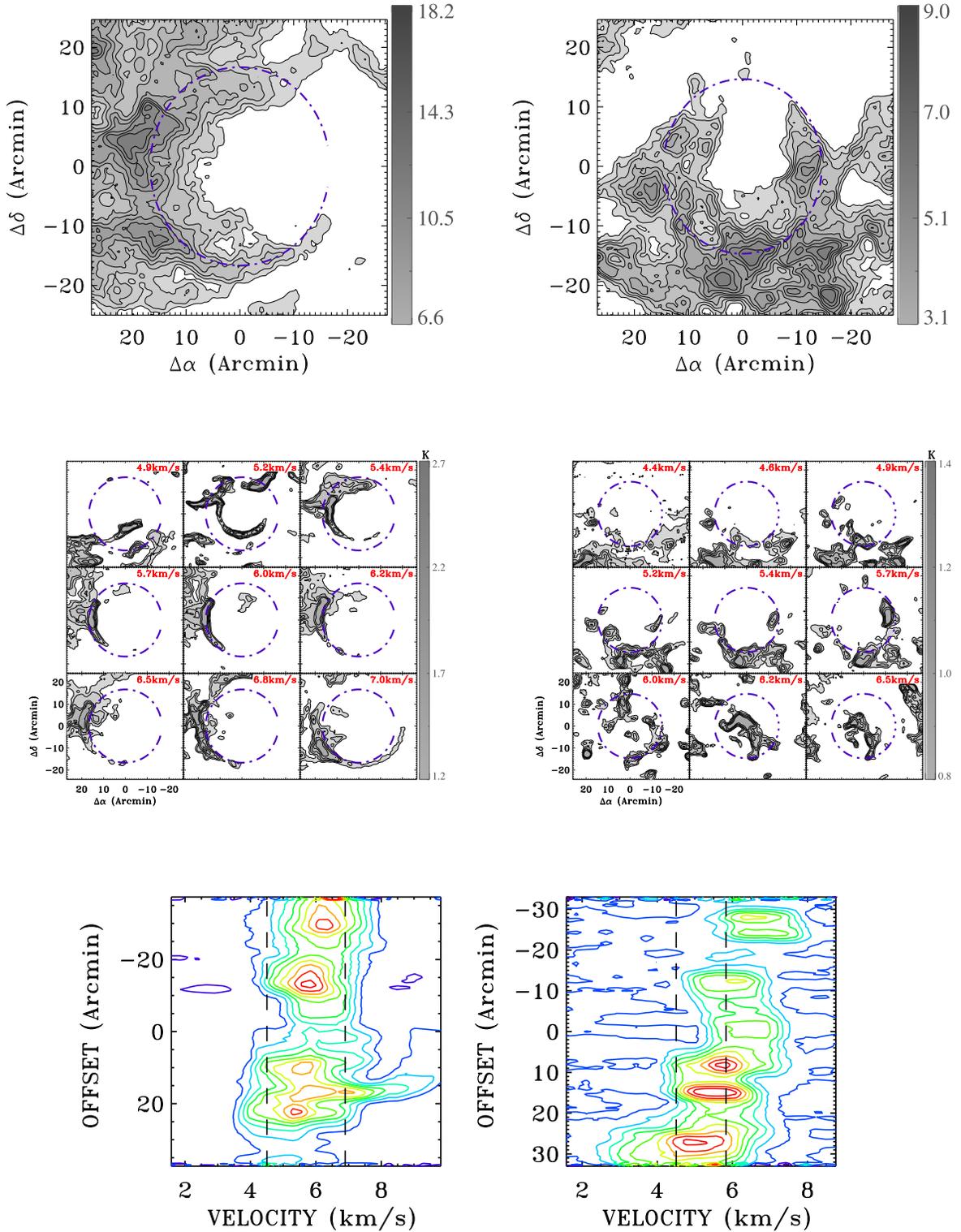
We performed 500 trials in both the TMC and the PMC to obtain the data set of  $N$ ,  $T_a$ , and  $V_e$ , and the detection result (0 or 1).  $N$ ,  $T_a$ , and  $V_e$  are then divided by their maximum values. The maximum values are obtained from real observed bubbles. We performed tenfold cross-validation on the experimental data sets to fit the constants, while testing how well the detectability function performs on the bubble identification and detectability prediction in each cloud. In 10-fold cross-validation, the data sets are randomly partitioned into 10 equal size subsets. One single subset is used as the validation data for testing the model, while the remaining nine subsets are used as training data to fit the model. We used the maximum likelihood estimator (MLE; see Appendix B) to get an estimate for each parameter from training sets and use the remaining subset for validation. We repeated this procedure 10 times. Our estimate for the parameters is the average over the 10 training runs, while the error is the average over the 10 validation runs. Tenfold cross-validation is frequently used when evaluating performance of models on multiclass data (James et al. 2013).

##### 4.3.1. Results in Taurus

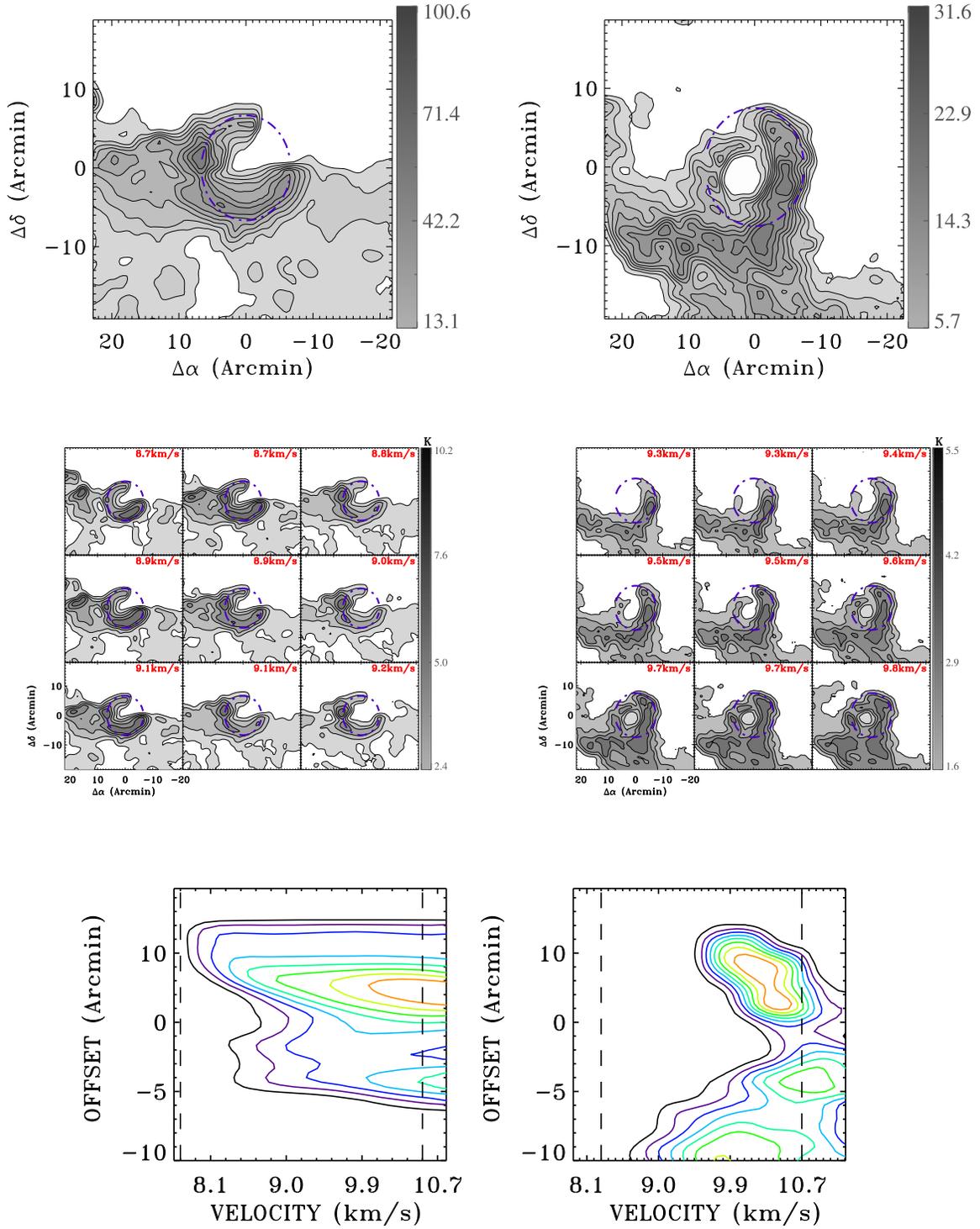
We performed tenfold cross-validation on the Taurus experimental data set with random segment selection on all the data for each fold to fit the detectability function. We adopt the average value of 10 times the fitting results for  $\alpha_0$  to  $\alpha_3$ , where  $\alpha_0 = -6.0$ ,  $\alpha_1 = 112.0$ ,  $\alpha_2 = 42.9$ , and  $\alpha_3 = 3.3$ . The uncertainties are derived from the standard deviation of 10 times the fitting results for  $\alpha_0$  to  $\alpha_3$ , which are 0.3, 8.3, 3.9, and 0.5, respectively. The detectability function in Taurus is described as

$$h^T(\mu_{T_a}, \mu_N, \mu_{V_e}) = \frac{1}{(1 + e^{6.0 - 112.0 * \mu_{T_a} - 42.9 * \mu_N - 3.3 * \mu_{V_e}})}, \quad (2)$$

where  $\mu_{T_a}$ ,  $\mu_N$ , and  $\mu_{V_e}$  are scaled experimental bubble parameters—average antenna temperature, number of pixels, and expansion velocity—in Taurus.



**Figure 1.** Comparison of artificial (left column) and real (right column) bubbles in integrated intensity (top), channel maps (middle), and  $P$ - $V$  diagrams (bottom) in the TMC. Upper left panel:  $^{13}\text{CO}$  integrated intensity map of artificial bubbles inserted in position of R.A. (J2000) =  $4^{\text{h}}38^{\text{m}}3^{\text{s}}$ , decl. (J2000) =  $25^{\circ}43'33''$ , integrated over  $4.2$ – $6.1$   $\text{km s}^{-1}$  with  $R = 0.7$  pc,  $\beta = 260^{\circ}$ ,  $\Delta R = 0.2$  pc,  $V_{\text{exp}} = 2.0$   $\text{km s}^{-1}$ ,  $n_0 = 0.5 \times 10^{-4}$   $\text{cm}^{-3}$ , and  $\sigma = 1.4$   $\text{km s}^{-1}$ . Upper right panel:  $^{13}\text{CO}$  integrated intensity map of identified bubble TMB<sub>35</sub> in position of R.A. (J2000) =  $4^{\text{h}}46^{\text{m}}12^{\text{s}}$ , decl. (J2000) =  $25^{\circ}07'33''$ , integrated over  $4.4$ – $6.5$   $\text{km s}^{-1}$  with  $R = 0.62$  pc and  $V_{\text{exp}} = 1.5$   $\text{km s}^{-1}$ . The blue dashed circles show the approximate extent of the CO emission of the bubble. Middle left panel: channel maps of the artificial bubble of  $^{13}\text{CO}$  emission. Middle right panel: channel maps of identified bubble TMB<sub>35</sub> of  $^{13}\text{CO}$  emission. The blue dashed circles are the same as that in the upper panel. Lower left panel:  $P$ - $V$  diagram of artificial bubble of  $^{13}\text{CO}$ , through the slice shown by the angle of  $110^{\circ}$ . Lower right panel:  $P$ - $V$  diagram of identified bubble TMB<sub>35</sub> of  $^{13}\text{CO}$ , through the slice shown by the angle of  $110^{\circ}$ . The two black dashed vertical lines show the extent of the  $^{13}\text{CO}$  emission associated with the bubble determined by visual inspection.



**Figure 2.** Comparison of artificial (left column) and real (right column) bubbles in integrated intensity (top), channel maps (middle), and  $PV$  diagrams (bottom) in the PMC. Upper left panel:  $^{13}\text{CO}$  integrated intensity map of artificial bubbles inserted in the position of R.A. (J2000) =  $3^{\text{h}}43^{\text{m}}5^{\text{s}}$ , decl. (J2000) =  $32^{\circ}00'51''$ , integrated over 7.2–10.0  $\text{km s}^{-1}$  with  $R = 0.44$  pc,  $\beta = 288^{\circ}$ ,  $\Delta R = 0.27$  pc,  $V_{\text{exp}} = 2.5$   $\text{km s}^{-1}$ ,  $n_0 = 4 \times 10^{-4}$   $\text{cm}^{-3}$ , and  $\sigma = 2.0$   $\text{km s}^{-1}$ . Upper right panel:  $^{13}\text{CO}$  integrated intensity map of identified bubble CPS\_12 in the position of R.A. (J2000) =  $3^{\text{h}}29^{\text{m}}26^{\text{s}}$ , decl. (J2000) =  $31^{\circ}25'10''$ , integrated over 8.8–10.4  $\text{km s}^{-1}$  with  $R = 0.44$  pc, and  $V_{\text{exp}} = 2.5$   $\text{km s}^{-1}$ . The blue dashed circles show the approximate extent of the CO emission of the bubble. Middle left panel: channel maps of the artificial bubble of  $^{13}\text{CO}$  emission. Middle right panel: channel maps of identified bubble CPS\_12 of  $^{13}\text{CO}$  emission. The blue dashed circles are the same as that in the upper panel. Lower left panel:  $P$ - $V$  diagram of artificial bubble of  $^{13}\text{CO}$ , through the slice shown by the angle of  $110^{\circ}$ . Lower right panel:  $P$ - $V$  diagram of identified bubble CPS\_12 of  $^{13}\text{CO}$ , through the slice shown by the angle of  $65^{\circ}$ . The two black dashed vertical lines show the extent of the  $^{13}\text{CO}$  emission associated with the bubble determined by visual inspection.

Meanwhile, we estimated the training error from the training set and generalization error from the testing set to analyze how well  $h^T$  performs on bubble identification and prediction in

Taurus. The average training error is about 0.14 which is the probability that  $h^T$  misclassifies samples in training sets. The average generalization error for  $h^T$  is about 0.12 which is

the probability if we draw a new set of bubble parameters and bubble detection results,  $h^T$  misclassifies it. In Taurus bubble experiments, there are 330 identified experimental bubbles, which are 66% of total experiments. As long as the detectability function can fit and predict the bubble detection result with correctness larger than 66%, we can consider that the detectability function can be used for bubble identification. According to the average training error and generalization error, the correctness of  $h^T$  to fit and predict the bubble detection result is 86% and 88%, respectively, which indicates that  $h^T$  can well fit and predict the bubble detection result in Taurus.

#### 4.3.2. Results in Perseus

We performed tenfold cross-validation on the Perseus experimental data set with random segment selection on all the data for each fold to fit the detectability function. We adopt the average value of 10 times fitting results for  $\alpha_0$  to  $\alpha_3$ , where  $\alpha_0 = -10.2$ ,  $\alpha_1 = 121.3$ ,  $\alpha_2 = 6.3$ , and  $\alpha_3 = 2.9$ . The uncertainty is derived from the standard deviation of 10 times the fitting results for  $\alpha_0$  to  $\alpha_3$ , which are 1.5, 23.6, 1.1, and 0.2, respectively. The detectability function in Perseus is described as

$$h^P(\mu_{T_a}, \mu_N, \mu_{V_e}) = \frac{1}{(1 + e^{10.2 - 121.3 * \mu_{T_a} - 6.3 * \mu_N - 2.9 * \mu_{V_e}})}, \quad (3)$$

where  $\mu_{T_a}$ ,  $\mu_N$ , and  $\mu_{V_e}$  are scaled experimental bubble parameters—average antenna temperature, number of pixels, and expansion velocity—in Perseus.

Meanwhile, we estimated the training error from the training set and generalization error from the testing set to analyze how well  $h^P$  performs on bubble identification and prediction in Perseus. The average training error is about 0.06 which is the probability that  $h^P$  misclassifies samples in training sets. The average generalization error for  $h^P$  is about 0.08 which is the probability if we draw a new set of bubble parameters and bubble detection result,  $h^P$  would misclassify it. In Perseus bubble experiments, there are 321 identified experimental bubbles, which are 64.2% of total experiments. As long as the detectability function can fit and predict the bubble detection result with correctness larger than 64.2%, we can consider that the detectability function can be used for bubble identification. According to the average training error and generalization error, the correctness of  $h^P$  to fit and predict bubble detection result is 94% and 92%, respectively, which indicate that  $h^P$  can well fit and predict the bubble detection result in Perseus.

#### 4.4. Comparison

The detectability functions for bubble average antenna temperature, number of pixels, and expansion velocity for each cloud are illustrated in Figure 3. In Taurus, the change from yellow to deep blue is more gradual than in Perseus, which is caused by the larger training error in Taurus. We found that bubble detectability in Taurus and Perseus both strongly depend on average antenna temperature and number of pixels. The brightness of the bubble understandably dominates the detectability of the bubbles. In the case of a large, relatively regularly shaped bubble, the detectability roughly scaled with the number of pixels. Weak dependence is seen to occur for expansion velocity. However, we still find that bubbles with larger expansion velocity are easier to detect. In general, bubbles in Perseus are easier to detect than Taurus with the

same  $N$ ,  $T_a$ , and  $V_e$ . According to the behaviors of the detectability function, bubbles with low average antenna temperature, small spatial size, and slow expansion velocity are more likely to be missed during manual identification. We quantify the kinetic energy of the missing bubble in the following section.

### 5. Kinetic Energy of Missing Bubbles

Bubble kinetic energy is associated with the input of mechanical energy into the parent molecular cloud, which is the crucial parameter to evaluate the impact of bubbles on the surrounding ISM. We estimated the completeness of previous bubble surveys by comparing the kinetic energy of missing bubbles with identified bubbles in Taurus and Perseus. The identified bubbles here and afterward represent the real identified bubbles in previous bubble surveys.

In order to estimate the kinetic energy distribution of all bubbles, including identified and missing bubbles in each molecular cloud, we assumed that the kinetic energy distribution of identified bubbles results from the kinetic energy distribution of all bubbles modified by the bubbles' detectability. Therefore, the number of the missing bubble can be estimated by the difference between the kinetic energy distribution of identified bubbles and the total number of bubbles predicted by our model.

The relation between the kinetic energy distribution of identified bubbles and all bubbles in the molecular cloud can be expressed as

$$\int N_D \cdot p(E_k) dE_k = \int N_A \cdot p^{\text{tot}}(E_k) \cdot h(E_k) dE_k, \quad (4)$$

where  $N_D$  is the number of real identified bubbles,  $N_A$  is the number of all bubbles in a given cloud,  $p(E_k)$  is the probability density function (PDF) of identified bubbles with respect to kinetic energy,  $p^{\text{tot}}(E_k)$  is the PDF of all bubbles with respect to kinetic energy within the cloud, and  $h(E_k)$  denotes the detectability function with respect to kinetic energy.

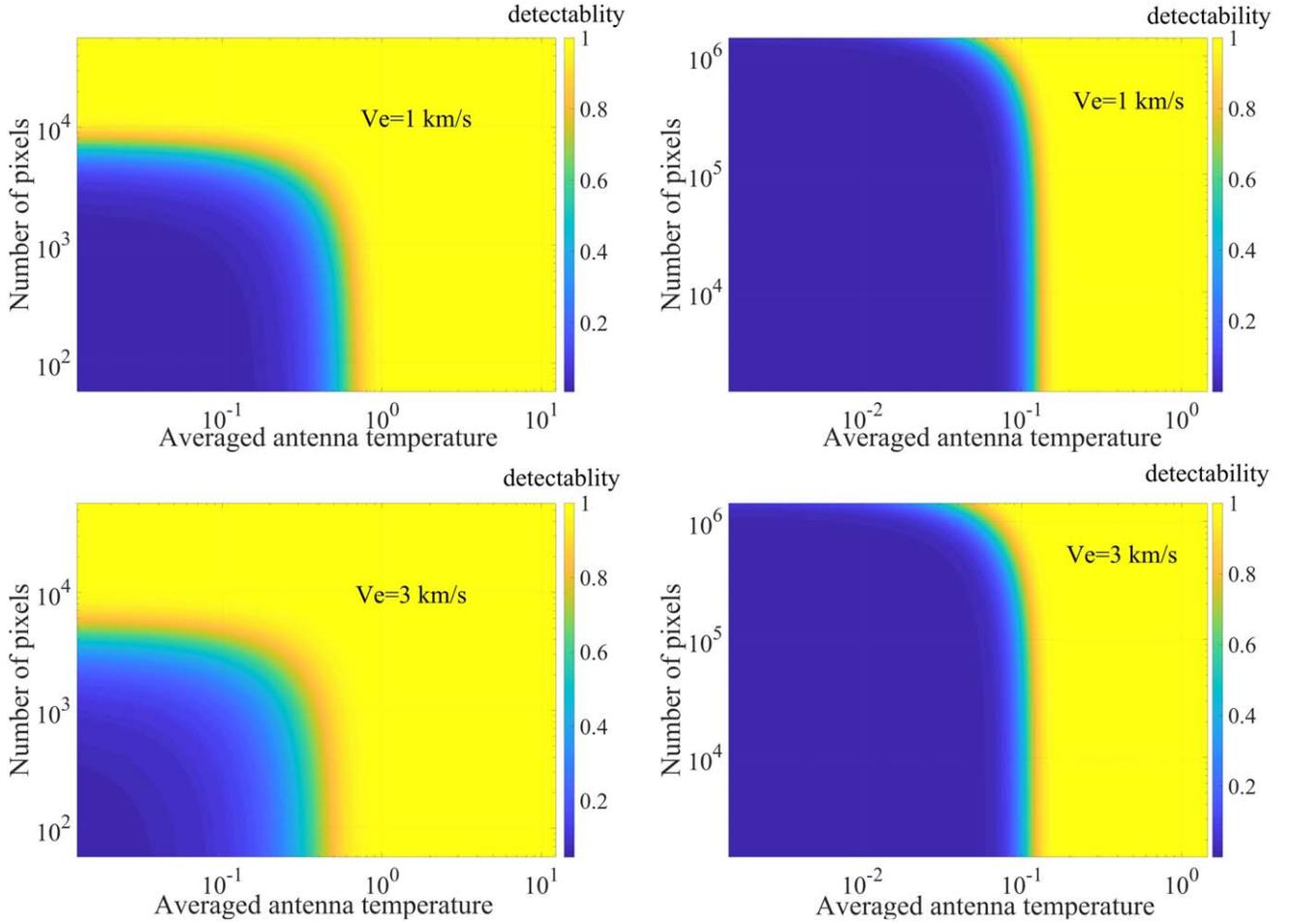
The number of missing bubbles  $N^{\text{miss}}$  for kinetic energy is described as

$$\begin{aligned} N^{\text{miss}} &= \int_{E_k^{\text{min}}}^{E_k^{\text{max}}} [N_A \cdot p^{\text{tot}}(E_k) - N_D \cdot p(E_k)] dE_k, \\ &= N_D \cdot \int_{E_k^{\text{min}}}^{E_k^{\text{max}}} [p(E_k)/h(E_k) - p(E_k)] dE_k. \end{aligned} \quad (5)$$

The kinetic energy of missing bubbles  $E_k^{\text{miss}}$  is written by

$$\begin{aligned} E_k^{\text{miss}} &= \int_{E_k^{\text{min}}}^{E_k^{\text{max}}} E_k [N_A \cdot p^{\text{tot}}(E_k) - N_D \cdot p(E_k)] dE_k, \\ &= N_D \cdot \int_{E_k^{\text{min}}}^{E_k^{\text{max}}} E_k [p(E_k)/h(E_k) - p(E_k)] dE_k. \end{aligned} \quad (6)$$

We constructed the kinetic energy of each experimental bubble based on their number of pixels, average antenna temperature, and expansion velocity. The detectability of each experimental bubble was evaluated by the detectability function of Equations (2) and (3). By fitting  $E_k$  with detectability using MLE, we estimate parameters of the kinetic energy detectability function. The parameters of PDF of the kinetic energy of identified bubbles is estimated from the MLE.



**Figure 3.** Comparison of bubble detectability in Taurus and Perseus in different average antenna temperature, number of pixels, and expansion velocity. Left column of the diagrams displays the bubble detectability in Taurus when  $V_e = 1 \text{ km s}^{-1}$  and  $3 \text{ km s}^{-1}$ , respectively. Right column of the diagrams displays the bubble detectability in Perseus when  $V_e = 1$  and  $3 \text{ km s}^{-1}$ , respectively.

A detailed description of the fitting method is illustrated in Appendix C.

### 5.1. The Kinetic Energy of Missing Bubbles in Taurus

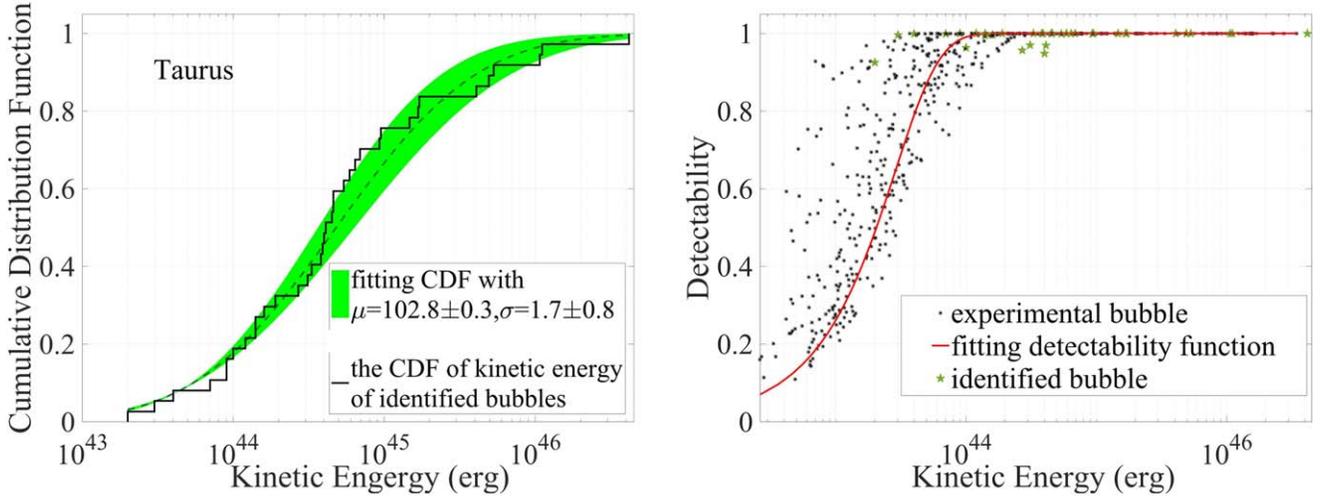
There are 37 identified bubbles  $N_D$  in Taurus. Their total kinetic energy is about  $9.2 \times 10^{46} \text{ erg}$ . The maximum and minimum kinetic energies of identified bubbles are  $E_k^{\max T} = 4.18 \times 10^{46} \text{ erg}$  and  $E_k^{\min T} = 0.2 \times 10^{44} \text{ erg}$ , respectively. The detectability of each experimental bubble in Taurus is evaluated from Equation (2). We plotted  $E_k$  versus detectability for both experimental bubbles and identified bubbles in the right panel of Figure 4. Our detectability function apparently overestimates the detectability of the bubble with low kinetic energy. However, such overestimation does not affect energy calculation related to feedback, as these bubbles add up to negligible energies compared to bright ones. One possibility is that there is a large number of small, low-energy, undetectable bubbles that might contribute significant energy to the clouds. In order for this to be true, there would need to be at least 4600 such bubbles in the TMC and 379 in the PMC. We find this to be unlikely. By fitting  $E_k$  with detectability using MLE, we obtain parameters of the kinetic energy detectability function by

$$h(E_k) = \tanh(925.5 \times E_k^u), \quad (7)$$

where  $E_k^u = \frac{E_k}{E_k^{\max T}}$  refers to uniformed kinetic energy. We chose the tanh function because it changes from 0 to 1 gradually when kinetic energy is larger than 0, which is consistent with the detectability of all experimental bubbles from small kinetic energy to large kinetic energy. The distribution of detectability of all experimental bubbles is not a Bernoulli distribution. Therefore, we cannot use the logistic function to fit the kinetic energy detectability directly.

The bubble detectability increases as the kinetic energy increases, which makes the bubble with low kinetic energy hard to identify. The number of identified bubbles with low kinetic energy should be small. Although the bubble with high kinetic energy is easier to identify, it is unlikely that low-mass star formation in Taurus and Perseus would generate a lot of high kinetic energy bubbles. Therefore, the number of identified bubbles with high kinetic energy should be small as well. Lognormal distribution satisfies the above conditions and can fit the real identified bubbles well. The probability distribution of kinetic energy for identified bubbles  $p(E_k)$  can be well fitted in truncated lognormal distribution using MLE (see Appendix C) by

$$p(E_k) = \frac{-\sqrt{2} e^{-\frac{1}{2} \frac{1}{\sigma^2} (\ln \frac{E_k}{m})^2}}{\sqrt{\pi} \sigma (\text{erf}(a_3) - \text{erf}(a_8)) E_k}, \quad (8)$$



**Figure 4.** Bubble detectability and cumulative distribution function for kinetic energy in Taurus. The left panel shows the CDF of identified bubbles for kinetic energy as black stairs. The CDF of the fitting distribution is illustrated in green, while the green area is the fitting uncertainty. The right panel shows the experimental bubble detectability as black stars and identified bubble detectability as green stars with the fitting detectability function shown in red.

where  $\sigma$  is the fitting shape parameter which is  $1.7 \pm 0.80$  in logarithm, erf is the error function, we defined  $\mu = \ln(m)$  to be the logarithmic fitting scale parameter which is  $102.8 \pm 0.27$ ,  $a_3$  and  $a_8$  are compact notations which are given by

$$a_3 = \frac{1}{2} \frac{\sqrt{2} (\ln(E_k^{\min T}) - \ln(m))}{\sigma},$$

$$a_8 = \frac{1}{2} \frac{\sqrt{2} (\ln(E_k^{\max T}) - \ln(m))}{\sigma}.$$

The fitting uncertainty was derived from 1000 sets of Monte Carlo experiments, each of size 37 (the number of identified bubbles in Taurus) with the kinetic energies ranging from  $E_k^{\min T}$  to  $E_k^{\max T}$ . For each Monte Carlo experiment, the kinetic energy is randomly generated from the truncated lognormal distribution with  $\mu = 102.8$ ,  $\sigma = 1.7$ . We applied the same fitting algorithm to those 37 random samples to estimate  $\mu$  and  $\sigma$  and repeated the Monte Carlo experiment 1000 times to get sets of the estimated value of  $\mu$  and  $\sigma$ . We took the standard deviation of the estimated 1000 sets of  $\mu$  and  $\sigma$  to be the fitting uncertainty. Since there are only 37 identified bubbles in Taurus, we did not plot the PDF of  $E_k$  of identified bubbles, instead we plotted the cumulative distribution function (CDF) of  $E_k$ . The CDF of  $E_k$  for identified bubbles and CDF of fitting distribution for Taurus are illustrated in the left panel of Figure 4.

By combining Equations (7) and (8) with Equation (5), the number of missing bubbles in Taurus can be derived, which ranges from 2 to 3. This is the probable number of bubbles not identified in surveys of L15 in Taurus. The total kinetic energy of the missing bubbles is estimated by combining Equations (7) and (8) with Equation (6). However, the kinetic energy of most missing bubbles aggregates at less than  $1 \times 10^{44}$  erg as illustrated in the green line of the right panel in Figure 6, which leads to the total kinetic energy of missing bubbles is ranging from  $7.2 \times 10^{43}$  to  $8.6 \times 10^{43}$  erg. This corresponds to about 0.01% of the kinetic energy of identified bubbles. Therefore, the Taurus bubble survey can be considered to be energetically complete.

## 5.2. Kinetic Energy of Missing Bubbles in Perseus

There are 12 identified bubbles  $N_D$  in Perseus. Their total kinetic energy is about  $7.58 \times 10^{46}$  erg. The kinetic energy of identified bubbles ranges from  $2 \times 10^{44}$  to  $1.88 \times 10^{46}$  erg. We adopt the minimum and maximum kinetic energy of identified bubbles in Taurus (which has a wider range than in Perseus) to be the range used in estimating the missing kinetic energy in Perseus.

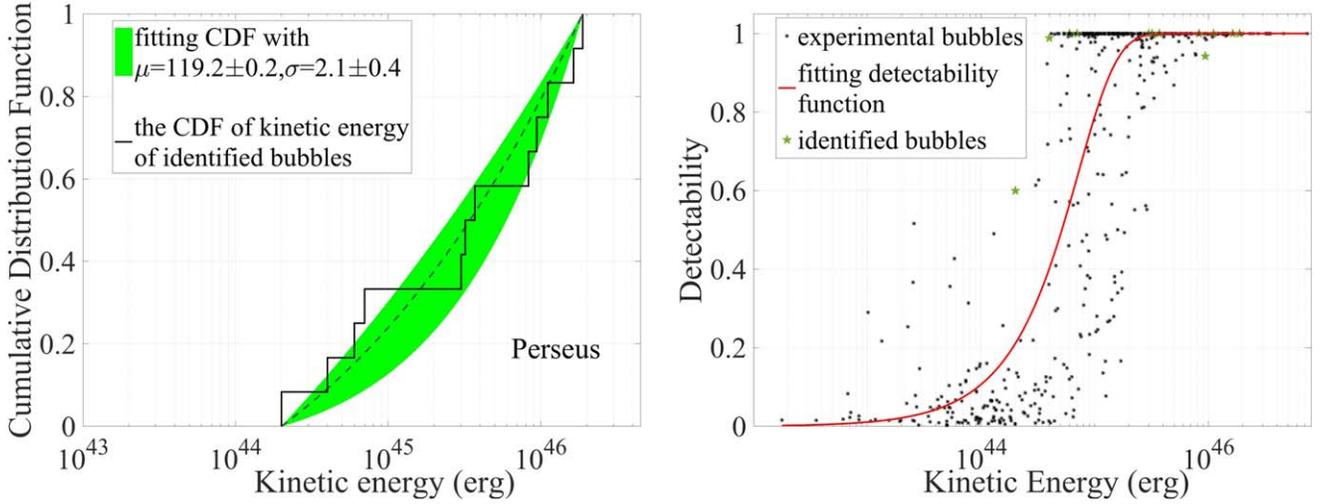
The detectability of each experimental bubble in Perseus is evaluated from Equation (3). We plotted  $E_k$  versus detectability for both experimental bubbles and identified bubbles in the right panel of Figure 5. Similar to Taurus, we obtain parameters of the kinetic energy detectability function by fitting  $E_k$  with detectability using MLE, which can be written as

$$h(E_k) = \tanh(79.34 \times E_k^\mu), \quad (9)$$

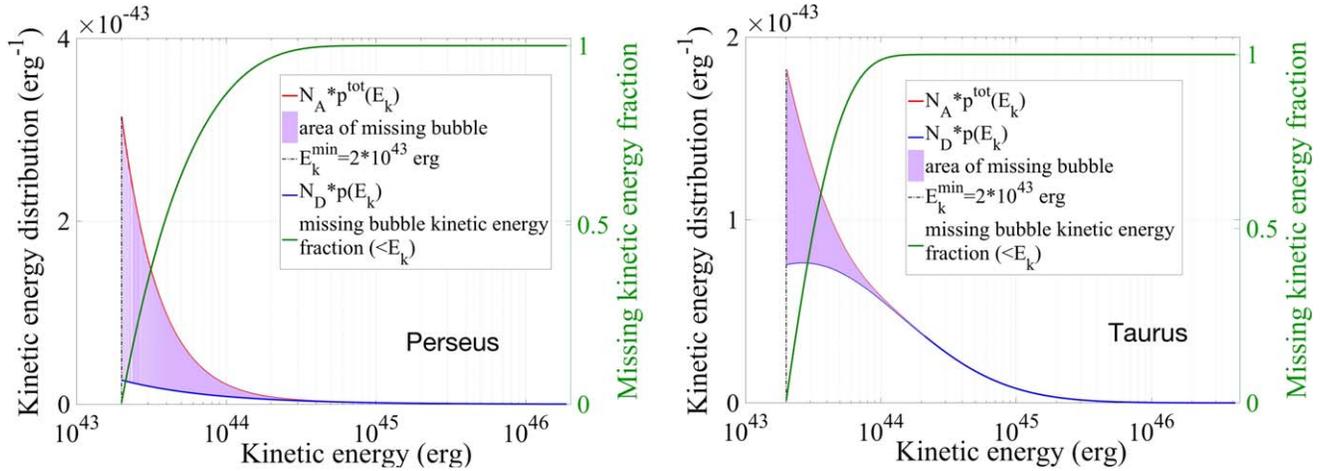
where  $E_k^\mu = \frac{E_k}{E_k^{\max T}}$  refers to uniformed kinetic energy.

Similarly, the probability distribution of kinetic energy for identified bubbles in Perseus can also be well fitted in truncated lognormal distribution using MLE shown in Equation (8). The fitting shape parameter is  $2.10 \pm 0.43$  in logarithm. The logarithmic fitting scale parameter is  $119.92 \pm 0.22$ . The fitting uncertainty was derived from 1000 sets of Monte Carlo experiments, each of size 12 (the number of identified bubbles in Perseus) with the kinetic energies ranging from  $E_k^{\min P}$  to  $E_k^{\max P}$ . For each Monte Carlo experiment, the kinetic energy is randomly generated from the truncated lognormal distribution with  $\mu = 119.92$  and  $\sigma = 2.10$ . We applied the same fitting algorithm to those 12 random samples to estimate  $\mu$  and  $\sigma$  then repeated the Monte Carlo experiment 1000 times to get sets of the estimated value of  $\mu$  and  $\sigma$ . The fitting uncertainty is the standard deviation of the estimated 1000 sets of  $\mu$  and  $\sigma$ . There are only 12 identified bubbles in Perseus. Similar to Taurus, we plotted up the CDF of  $E_k$  for identified bubbles and CDF of fitting distribution for Perseus, which are illustrated in the left panel of Figure 5.

By combining Equations (8) and (9) with Equation (5), the number of missing bubbles in Perseus can be derived which is



**Figure 5.** Bubble detectability and cumulative distribution function for kinetic energy in Perseus. The left panel shows the CDF of identified bubbles for kinetic energy as black stairs. The CDF of the fitting distribution is illustrated in green, while the green area is the fitting uncertainty. The right panel shows the experimental bubble detectability as black stars and identified bubble detectability as green stars with the fitting detectability function in red.



**Figure 6.** Comparison of the distribution and missing bubble kinetic energy fraction. The left panel shows the distribution of bubbles for kinetic energy in Perseus; the red curve presents the distribution of all bubbles in Perseus for kinetic energy and the blue curve presents the distribution of identified bubbles in Perseus for kinetic energy. The purple area illustrates the missing bubble. The green curve shows the missing bubble kinetic energy fraction, which represents the percentage of missing kinetic energy lower than  $E_k$  to the total missing kinetic energy. The right panel shows the distribution of bubbles for kinetic energy in Taurus. Same as the left panel.

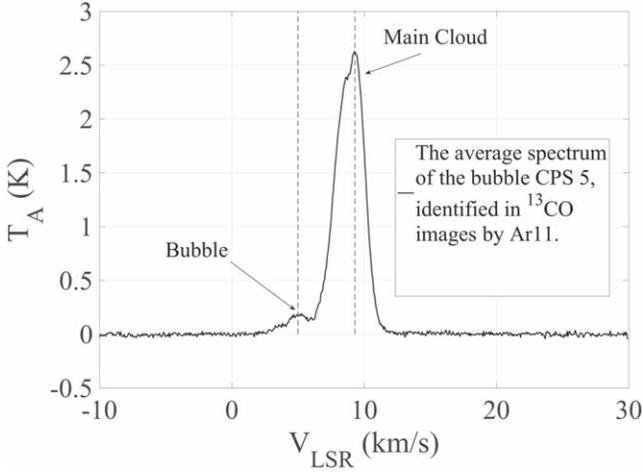
ranging from 3 to 14. This is the probable number of bubbles not identified in the Perseus survey by Ar11. The total kinetic energy of the missing bubbles is estimated by combining Equations (9) and (8) with Equation (6). However, the kinetic energy of most missing bubbles aggregates at less than  $2 \times 10^{44}$  erg as illustrated in the green line of the left panel of Figure 6, which leads to the total kinetic energy of missing bubbles is ranging from  $1.4 \times 10^{44}$  to  $8.0 \times 10^{44}$  erg. This corresponds to about 1% of the kinetic energy of identified bubbles. Therefore, the Perseus bubble survey can be considered to be energetically complete.

## 6. Discussion and Conclusions

Stellar feedback has a significant impact on the surrounding gas. The total kinetic energy of bubbles detected in Taurus and Perseus are  $\sim 3.9 \times 10^{46}$  erg and  $\sim 7.6 \times 10^{46}$  erg, respectively. The gravitational potential energy and turbulence energy

( $E_G$ ,  $E_{\text{tur}}$ ) of Taurus and Perseus are about  $1.5 \times 10^{48}$  erg,  $3.2 \times 10^{47}$  erg, and  $6 \times 10^{47}$ ,  $1.6 \times 10^{47}$  erg, respectively. The energy contained in bubbles are orders of magnitude smaller than those of either gravity or turbulence, which are expected to be equal in a virialized supersonic cloud. The maximum energy injection (into the natal cloud) rate of molecular bubbles can be estimated from their momentum. They are  $\sim 6.4 \times 10^{33}$  erg  $\text{s}^{-1}$  and  $\sim 2 \times 10^{33}$  erg  $\text{s}^{-1}$  for bubbles in Taurus and Perseus, respectively. Following the methods laid out in Ar11 and L15, we estimated the turbulence dissipation rate of Taurus and Perseus to be  $\sim 3.1 \times 10^{33}$  erg  $\text{s}^{-1}$  and  $\sim 1 \times 10^{33}$  erg  $\text{s}^{-1}$ , respectively. The observed injection rates are thus similar to or even slightly larger than the dissipation rates.

Typically, when discussing molecular clouds we refer to total proton volume densities of  $10^3 \text{ cm}^{-3}$  for regions with sufficient volume density to produce  $^{13}\text{CO}$  emission. This is typical for regions with  $A_V \geq 1$ . However, the volumes of gas that are affected by bubbles are large and not necessarily



**Figure 7.** Average spectrum of the bubble CSP 5, identified in  $^{13}\text{CO}$  by Ar11. The bubble was identified between 2.0 and 6.0  $\text{km s}^{-1}$ . The dashed lines present the bubble velocity offset from the emission of the main cloud.

located in the dense regions. In fact, they almost necessarily include large volumes of very diffuse gas. The gas that is dynamically affected (“moved”) by bubble expansion and can be identified as such has to exhibit a velocity offset (Figure 7). Such “high” velocity gas has much lower density. Due to this, we chose to use our estimates of the average  $\text{H}_2$  volume densities within Taurus and Perseus ( $35 \text{ cm}^{-3}$  and  $280 \text{ cm}^{-3}$  respectively) when estimating the energy injected into the cloud by each bubble. As a comparison, if the main cloud density was assumed to be  $3.5 \times 10^3 \text{ cm}^{-3}$ , the energy injected into the cloud by star formation feedback in Taurus and Perseus would be  $\sim 3.9 \times 10^{48}$  erg and  $\sim 9.1 \times 10^{47}$  erg, respectively, which could disperse the whole cloud. Similarly, if we use such a high volume density for the undisturbed gas in our simulations, then the resulting bubbles are much brighter than any that are actually observed.

We simulated bubbles for the Taurus and Perseus molecular cloud. We randomly inserted artificial bubbles into the Taurus and Perseus  $^{13}\text{CO}$  data cubes from the 13.7 m FCRAO telescope. With changing average antenna temperature, the number of pixels, and expansion velocity of artificial bubbles, we parameterized bubble detections according to the bubble detectability functions to evaluate how the detectability is affected by the parameters for two molecular clouds. According to the properties of identified bubbles in Taurus and Perseus and the detectability functions, we estimated the energetic completeness of previously stellar feedback studies. Our conclusions regarding the completeness of bubble identification in Taurus and Perseus and their properties are as follows:

1. The detectability of bubbles can be described as a logistic function, which is derived from GLMs. The distribution of bubble detection results is a Bernoulli distribution. Therefore we adopted GLMs to fit the bubble detectability functions. In the TMC, the detectability function can be described as  $h^T(\mu_{T_a}, \mu_N, \mu_{V_e}) = \frac{1}{(1 + e^{6.0 - 112.0 * \mu_{T_a} - 42.9 * \mu_N - 3.3 * \mu_{V_e}})}$ , while in the PMC as  $h^P(\mu_{T_a}, \mu_N, \mu_{V_e}) = \frac{1}{(1 + e^{10.2 - 121.3 * \mu_{T_a} - 6.3 * \mu_N - 2.9 * \mu_{V_e}})}$ . We then

fitted the bubble kinetic energy distributions in Taurus and Perseus with truncated lognormal distribution.

2. The number of missing bubbles in Taurus is less than 8% of the number of identified bubbles. The number of missing bubbles in Perseus is about 25%–125% of the number of identified bubbles.
3. We used bubble kinetic energy distributions and the bubble detectability functions to estimate the total kinetic energy of missing bubbles, which suggests that although the numbers of missing bubbles are large, their kinetic energies are relatively small (usually less than  $2 \times 10^{44}$  erg). The total kinetic energies of missing bubbles in Taurus and Perseus during manual identification range from  $7.2 \times 10^{43}$  erg to  $8.6 \times 10^{43}$  erg and  $1.4 \times 10^{44}$  erg to  $8.0 \times 10^{44}$  erg, respectively. Such potential incompleteness only accounts for  $\sim 0.2\%$  and  $\sim 1\%$  of the total kinetic energy of identified bubbles in Taurus and Perseus, respectively.
4. The empirical surveys in L15 and Ar11 for identifying bubble structures in Taurus and Perseus can be considered as energetically complete.
5. The total energy of bubbles in a cloud is orders of magnitude smaller than those of either turbulence or gravity. The bubbles cannot generate the observed turbulence or disperse the cloud. The observed energy injection rate from bubbles, now considered complete, is similar to the turbulence dissipation rate. We conclude that, even in low-mass star-forming regions, the feedback from star formation is sufficient to sustain turbulence at ranges from  $\sim 0.1$  pc to  $\sim 2.8$  pc scales.

This work is supported by the National Natural Science Foundation of China grant No. 11988101, No. 11725313, No. 11721303, the International Partnership Program of Chinese Academy of Sciences grant No. 114A11KYSB20160008, the National Science Foundation of China No. 11721303, and the National Key R&D Program of China No. 2016YFA0400702.

## Appendix A Bubble Simulation Description

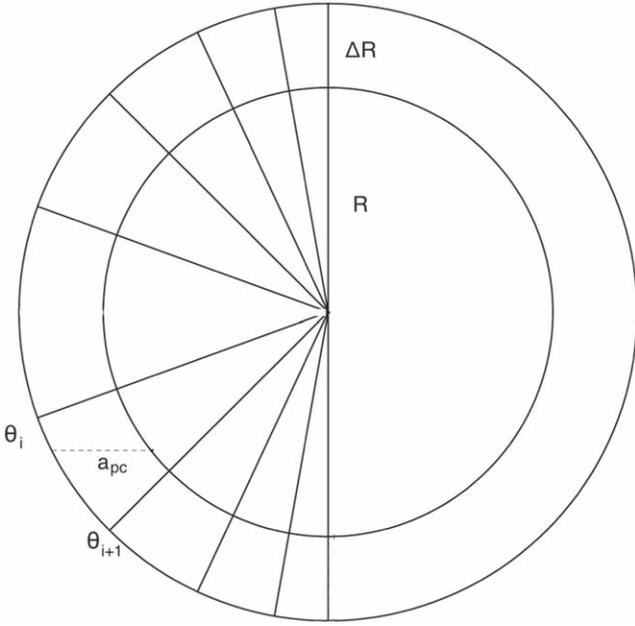
The  $^{13}\text{CO}$  column densities may be described by

$$N_{^{13}\text{CO}} = 1.18 \times 10^{15} T_b \Delta V, \quad (10)$$

where  $T_b$  is the brightness temperature of each pixel,  $N_{^{13}\text{CO}}$  is the column density, and  $\Delta V$  represents the velocity width of a spectrometer channel which is listed in Table 1. This relation is derived under several assumptions. We assumed that the excitation temperature  $T_{\text{ex}}$  of  $^{13}\text{CO}$  is 25 K, the background temperature is 2.7 K, the  $^{13}\text{CO}$  emission from the bubble is generally optically thin ( $\tau(^{13}\text{CO}) \ll 1$ ) and in local thermal equilibrium.  $N_{^{13}\text{CO}}$  can also be expressed as

$$N_{^{13}\text{CO}} = 3.0857 \times 10^{18} n_s a_{\text{pc}} \text{ cm}^{-2}, \quad (11)$$

where  $a_{\text{pc}}$  describes the bubble physical depth along the line of sight (LOS) in parsecs. By combining Equation (10) with Equation (11), we obtain the  $T_b$  for each pixel in the artificial



**Figure 8.**  $^{13}\text{CO}$  bubble sampled by constant velocity channels with an expansion velocity of  $V_{\text{exp}}$ . The radius and thickness of the bubble is  $R$  and  $\Delta R$ , respectively. An observer located far to the left would see a series of caps and rings of different depths. The dashed line shows the bubble physical depth at a certain line of sight.  $\theta_i$  and  $\theta_{i+1}$  are the starting angle and ending angle of the second velocity channel.

bubbles with known  $a_{\text{pc}}$ ,  $\Delta V$ , and  $n_s$ , which are given by

$$T_b = 2.615 * 10^3 \frac{n_s a_{\text{pc}}}{\Delta V}. \quad (12)$$

We assume that the  $^{13}\text{CO}$  abundance relative to  $\text{H}_2$  is  $1.43 \times 10^{-6}$ . The densities of  $^{13}\text{CO}$  affected by bubble expansion  $n_0$  are  $5 \times 10^{-5} \text{ cm}^{-3}$  and  $4 \times 10^{-4} \text{ cm}^{-3}$  for Taurus and Perseus, respectively. If all of the  $^{13}\text{CO}$  within the radius  $R + \Delta R$  is distributed within the bubble, we can get  $n_s/n_0 = 1 - \xi^{-3}$ , where  $\xi = (R + \Delta R)/R$ .

Figure 8 illustrates how  $a_{\text{pc}}$  is calculated. It shows a  $^{13}\text{CO}$  bubble sampled by five velocity channels in the left hemisphere as seen by an observer located to the far left. The starting and ending angles of the second velocity channel are  $\theta_i$  and  $\theta_{i+1}$ , respectively. The angular dimension of each velocity channel is given by  $\cos(\theta_{i+1}) = \cos(\theta_i) + \Delta V/V_{\text{exp}}$ . Thereby, the angle of any LOS within the velocity channel can be determined. The physical depth  $a_{\text{pc}}$  of each LOS can be calculated from the angle of this LOS, bubble radius, and thickness. The detailed calculations are presented in Appendix A of Cazzolato & Pineault (2005). By estimating the physical depth  $a_{\text{pc}}$  of each LOS for a given expansion velocity, and knowing  $n_s$  of the bubble, we can obtain antenna temperature for each LOS using Equation (12).

## Appendix B

### Bubble Detectability Function Derivation and Fitting

According to the definition, the experimental bubble detection result,  $y$ , is either 0 or 1 which corresponds to a Bernoulli distribution. The probability mass function  $p$  of the

detection results  $y$  is given by

$$p(y = 1; \phi) = \phi,$$

$$p(y = 0; \phi) = 1 - \phi,$$

where  $p(y = 1; \phi)$  is the probability of detecting a bubble and  $p(y = 0; \phi)$  is the probability of not detecting a bubble.  $p$  can be also expressed as (Nelder & Wedderburn 1972)

$$\begin{aligned} p(y; \phi) &= \phi^y (1 - \phi)^{1-y} \\ &= \exp(\log(\phi^y (1 - \phi)^{1-y})) \\ &= \exp(y \log(\phi) + (1 - y) \log(1 - \phi)) \\ &= \exp\left(\log\left(\frac{\phi}{1 - \phi}\right) y + \log(1 - \phi)\right). \end{aligned}$$

The exponential family is written as

$$f(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta)),$$

where  $\eta$  is the natural parameter of the distribution;  $T(y)$  is the sufficient statistic; and  $a(\eta)$  is the log partition function. When  $\eta = \log\left(\frac{\phi}{1 - \phi}\right)$ ,  $a(\eta) = -\log(1 - \phi)$ ,  $T(y) = y$ , and  $b(y) = 1$ , we can derive the  $p(y; \phi)$  from  $f(y; \eta)$ . This indicates that the probability mass function of the experimental bubble detection result is one of exponential family. The experimental bubble parameters:  $N$ ,  $T_a$  and  $V_e$  can be considered as independent to each other, so that we can combine them linearly after scaling. The scaling rule is as follows:

$$\mu_{\text{scale}} = \frac{\mu}{\mu_{\text{max}}} \quad (13)$$

where  $\mu$  represents the bubble parameter,  $\mu_{\text{min}}$  and  $\mu_{\text{max}}$  refer to the minimum and the maximum value of parameter.

Thus,  $\eta$  can be constructed as  $\eta = \alpha^T X$ . Here,  $X$  denotes a three-dimensional vector containing  $N$ ,  $T_a$ , and  $V_e$ .  $\alpha$  denotes a four-dimensional coefficient vector containing an intercept  $\alpha_0$ . Our goal is to predict the expected value of the probability of detecting the bubble  $y$  with given  $X$ , which means we would like the detectability function  $h(X)$  output to satisfy  $h(X) = E[y|X]$ . In our formulation of the Bernoulli distribution as an exponential family distribution, we had  $\eta = \log\left(\frac{\phi}{1 - \phi}\right)$  which can be written as  $\phi = \frac{1}{1 + e^{-\eta}}$  and  $E[y|X; \alpha] = \phi$ . So that the bubble detectability function  $h(X)$  can be expressed as

$$h(X) = E[y|X; \alpha] \quad (14)$$

$$= \phi \quad (15)$$

$$= \frac{1}{1 + e^{-\alpha^T X}}. \quad (16)$$

$h(X)$  is the logistic function. With given experimental bubble parameter and bubble detection result data sets, we can estimate the  $\alpha^T$  for  $h(X)$  based on the maximizing the likelihood function  $l(\theta) = \sum_{i=1}^m y^{(i)} \log(h(X^{(i)})) + (1 - y^{(i)}) \log(1 - h(X^{(i)}))$ , where  $m$  is the row number of the data set.

## Appendix C

### The Kinetic Energy Probability Density Function Fitting

We adopt the MLE to estimate the parameters of PDF of kinetic energy for both Taurus and Perseus using the truncated lognormal distribution. The derivation for the MLE of the truncated lognormal distribution largely follows Zaninetti (2017). Consider a sample  $\mathcal{X} = x_1, x_2, \dots, x_n$ . The maximum and minimum value of sample can be expressed as  $x_l$  and  $x_u$ , respectively, which are given by

$$x_l = \max(x_1, x_2, \dots, x_n), \quad x_u = \min(x_1, x_2, \dots, x_n). \quad (17)$$

The PDF can be expressed as

$$\text{PDF}(x; m, \sigma, x_l, x_u) = \frac{-\sqrt{2} e^{-\frac{1}{2} \frac{(\ln(\frac{x}{m}))^2}{\sigma^2}}}{\sqrt{\pi} \sigma (\text{erf}(a_3) - \text{erf}(a_8)) x}, \quad (18)$$

where  $m$  is the scale parameter,  $\sigma$  is the shape parameter,  $a_3$  and  $a_8$  are compact notations which are given by

$$a_3 = \frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma},$$

$$a_8 = \frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}.$$

The CDF can be expressed as

$$\text{CDF}(x; m, \sigma, x_l, x_u) = \frac{-\text{erf}\left(\frac{1}{2} \frac{\sqrt{2}}{\sigma} \ln\left(\frac{x}{m}\right)\right) + \text{erf}(a_3)}{\text{erf}(a_3) - \text{erf}(a_8)}. \quad (19)$$

The MLE is obtained by maximizing

$$\Lambda = \sum_i^n \ln(\text{PDF}(x_i; m, \sigma, x_l, x_u)). \quad (20)$$

The two derivatives  $\frac{\partial \Lambda}{\partial m} = 0$  and  $\frac{\partial \Lambda}{\partial \sigma} = 0$  generate two nonlinear equations in  $m$  and  $\sigma$  which can be solved numerically,

$$\begin{aligned} \frac{\partial \Lambda}{\partial m} = & \left( \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma}\right) \right. \\ & \left. - \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}\right) \right) \\ & \times \left( n\sqrt{2} \sigma e^{-\frac{1}{2} \frac{(\ln(x_l) - \ln(m))^2}{\sigma^2}} - n\sqrt{2} \sigma e^{-\frac{1}{2} \frac{(\ln(x_u) - \ln(m))^2}{\sigma^2}} \right. \\ & \left. - \sqrt{\pi} \left( \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma}\right) \right) \right. \\ & \left. - \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}\right) \right) \left( n \ln(m) - \sum_{i=1}^n \ln(x_i) \right) = 0, \end{aligned} \quad (21)$$

and

$$\frac{\partial \Lambda}{\partial \sigma} = \frac{N}{D} = 0, \quad (22)$$

where

$$\begin{aligned} N = & \ln(x_u) \sqrt{2} e^{-\frac{1}{2} \frac{(\ln(x_u) - \ln(m))^2}{\sigma^2}} n\sigma - \ln(x_l) \sqrt{2} e^{-\frac{1}{2} \frac{(\ln(x_l) - \ln(m))^2}{\sigma^2}} n\sigma \\ & + \sqrt{2} e^{-1/2} \frac{(\ln(x_l) - \ln(m))^2}{\sigma^2} \ln(m) n\sigma - \sqrt{2} e^{-1/2} \frac{(\ln(x_u) - \ln(m))^2}{\sigma^2} \ln(m) n\sigma \\ & + n(\ln(m))^2 \sqrt{\pi} \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}\right) \\ & - n\sigma^2 \sqrt{\pi} \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}\right) \\ & - n(\ln(m))^2 \sqrt{\pi} \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma}\right) \\ & + n\sigma^2 \sqrt{\pi} \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma}\right) \\ & + \sum_{i=1}^n \ln(x_i) (\ln(x_i) - 2 \ln(m)) \sqrt{\pi} \\ & \times \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}\right) \\ & - \sum_{i=1}^n \ln(x_i) (\ln(x_i) - 2 \ln(m)) \sqrt{\pi} \\ & \times \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma}\right), \end{aligned} \quad (23)$$

$$\begin{aligned} D = & \sqrt{\pi} \left( -\text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_l) - \ln(m))}{\sigma}\right) \right. \\ & \left. + \text{erf}\left(\frac{1}{2} \frac{\sqrt{2} (\ln(x_u) - \ln(m))}{\sigma}\right) \right) \sigma^3. \end{aligned} \quad (24)$$

### ORCID iDs

Mengting Liu  <https://orcid.org/0000-0001-7790-5498>

Di Li  <https://orcid.org/0000-0003-3010-7661>

Luis C. Ho  <https://orcid.org/0000-0001-6947-5846>

### References

- Arce, H. G., Borkin, M. A., Goodman, A. A., Pineda, J. E., & Beaumont, C. N. 2011, *ApJ*, **742**, 105
- Arce, H. G., Borkin, M. A., Goodman, A. A., Pineda, J. E., & Halle, M. W. 2010, *ApJ*, **715**, 1170
- Arce, H. G., & Goodman, A. A. 2001, *ApJ*, **554**, 132
- Arce, H. G., & Goodman, A. A. 2002, *ApJ*, **575**, 911
- Arce, H. G., & Sargent, A. I. 2006, *ApJ*, **646**, 1070
- Bally, J., Reipurth, B., & Davis, C. J. 2007, in *Protostars and Planets VI*, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. Arizona Press), 215
- Beaumont, C. N., Goodman, A. A., Kendrew, S., et al. 2014, *ApJS*, **214**, 3
- Cazzolato, F., & Pineault, S. 2005, *AJ*, **129**, 2731
- Churchwell, E., Povich, M. S., Allen, D., et al. 2006, *ApJ*, **649**, 759
- COMPLETE team 2011, *FCRAO Perseus 13CO cubes and map*, Harvard Dataverse, V2, <https://hdl.handle.net/10904/10075>
- Duarte-Cabral, A., Chrysostomou, A., Peretto, N., et al. 2012, *A&A*, **543**, 140
- Fedderson, J. R., Arce, H. G., Shuo, K., et al. 2018, *ApJ*, **862**, 121
- Federrath, C. 2015, *MNRAS*, **450**, 4035
- 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
- Fukui, Y., Sugitani, K., Takaba, H., et al. 1986, *ApJ*, **311**, 85
- Fuller, G. A., & Ladd, E. F. 2002, *ApJ*, **573**, 699

- Goldsmith, P. F., Heyer, M., Narayanan, G., et al. 2008, *ApJ*, 680, 428
- Hartmann, L., Ballesteros-Paredes, J., & Bergin, E. A. 2001, *ApJ*, 562, 852
- Heiles, C. 1979, *ApJ*, 229, 533
- James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013, *An Introduction to Statistical Learning*, Vol. 112 (Berlin: Springer)
- Kroupa, P., Jeřábková, T., Dinnbier, F., Beccari, G., & Yan, Z. 2018, *A&A*, 612, 74
- Kwan, J., & Scoville, N. 1976, *ApJL*, 210, 139
- Lada, C. J., & Harvey, P. M. 1981, *ApJ*, 245, 58
- Li, H. X., Li, D., Xu, D., et al. 2015, *ApJS*, 219, 20
- Matzner, C. D. 2002, *ApJ*, 566, 302
- Mottram, J. C., van Dishoeck, E. F., Kristensen, L. E., et al. 2017, *A&A*, 600, A99
- Nakamura, F., Kamada, Y., Kamazaki, T., et al. 2011, *ApJ*, 726, 46
- Nakamura, F., Sugitani, K., Shimajiri, Y., et al. 2011, *ApJ*, 737, 56
- Narayanan, G., Heyer, M. H., Brunt, C., et al. 2008, *ApJS*, 177, 341
- Narayanan, G., Snell, R., & Bemis, A. 2012, *MNRAS*, 425, 2641
- Nelder, J. A., & Wedderburn, R. W. M. 1972, *J. R. Stat. Soc., Ser. A*, 135, 370
- Norman, C., & Ikeuchi, S. 1989, *ApJ*, 345, 372
- Offner, S. S. R., & Arce, H. G. 2015, *ApJ*, 811, 146
- Plunkett, A. L., Arce, H. G., Corder, S. A., et al. 2013, *ApJ*, 774, 22
- Ridge, N. A., Di Francesco, J., Kirk, H., et al. 2006, *AJ*, 131, 2921
- Snell, R. L., Loren, R. B., & Plambeck, R. L. 1980, *ApJ*, 239, 17
- Solomon, P. M., Huguenin, G. R., & Scoville, N. Z. 1981, *ApJ*, 245, 19
- Zaninetti, L. 2017, *AdAp*, 2, 197
- Zinnecker, H., & Yorke, H. W. 2007, *ARA&A*, 45, 481