THE PERILS OF CLUMPFIND: THE MASS SPECTRUM OF SUB-STRUCTURES IN MOLECULAR CLOUDS

JAMIE E. PINEDA1, ERIK W. ROSOLOWSKY2, AND ALYSSA A. GOODMAN1

1 Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
2 University of British Columbia Okanagan, 3333 University Way, Kelowna, BC V1V 1V7, Canada

Abstract

We study the mass spectrum of sub-structures in the Perseus Molecular Cloud Complex traced by 13CO (1–0), finding that $dN/dM \propto M^{-2.4}$ for the standard Clumpfind parameters. This result does not agree with the classical $dN/dM \propto M^{-1.6}$. To understand this discrepancy we study the robustness of the mass spectrum derived using the Clumpfind algorithm. Both 2D and 3D Clumpfind versions are tested, using 850 $\mu$m dust emission and 12CO spectral-line observations of Perseus, respectively. The effect of varying threshold is not important, but varying stepsize produces a different effect for 2D and 3D cases. In the 2D case, where emission is relatively isolated (associated with only the densest peaks in the cloud), the mass spectrum variability is negligible compared to the mass function fit uncertainties. In the 3D case, however, where the 13CO emission traces the bulk of the molecular cloud, the number of clumps and the derived mass spectrum are highly correlated with the stepsize used. The distinction between “2D” and “3D” here is more importantly also a distinction between “sparse” and “crowded” emission. In any “crowded” case, Clumpfind should not be used blindly to derive mass functions. Clumpfind’s output in the “crowded” case can still offer a statistical description of emission useful in inter-comparisons, but the clump-list should not be treated as a robust region decomposition suitable to generate a physically-meaningful mass function. We conclude that the 13CO mass spectrum depends on the observations resolution, due to the hierarchical mass spectrum derived using Clumpfind in crowded (13CO (1–0)) and sparse emission (SCUBA 850$\mu$m). By using the 13CO (1–0) and SCUBA 850$\mu$m data collected by the COMPLETE team in the Perseus Molecular Cloud Complex (Ridge et al. 2006; Kirk et al. 2006) we are able to study the algorithm in both its 3D and 2D versions on real data sets (with overlapping sky coverage) with high resolution and sensitivity.
2. DATA

We use the $^{13}$CO (1-0) molecular line map obtained by the COMPLETE Survey (Ridge et al. 2006) using the SEQUOIA 32-element focal plane array at the FCRAO telescope. Observations were carried out using the on-the-fly technique. The data cube covers an area of $\sim$6.25$''$$\times$3$''$ with a 46$''$ beam on a 23$''$ grid, and it is presented in $T_{RMS}^*$ scale.

The map is beam sampled and Hanning smoothed in velocity. The final pixel size is 46$''$ and the velocity resolution is $\Delta v = 0.066$ km s$^{-1}$. The median Root Mean Square (RMS) noise in the map is 0.1 K in $T_{RMS}^*$, and all positions with RMS > 0.3 K are removed from the map. In addition, a noise-added $^{13}$CO cube (with RMS=0.2 K) is also used.

We also use the 850 $\mu$m map obtained with SCUBA on the JCMT (Kirk et al. 2006). The pixel scale is 6$''$, while the effective beam is 19.9$''$. The mean RMS in the map is $\sim$0.06 Jy beam$^{-1}$. The map coverage is smaller than the $^{13}$CO data, but it covers the densest regions in the cloud, where the dense cores identified by SCUBA lie (Kirk et al. 2006; Hatchell et al. 2005). It is important to note that in the SCUBA map map any structure larger than $\sim$Hatchell et al. 2005). It is important to note that in the SCUBA data, but it covers the densest regions in the cloud, where one; while the stepsize is varied between 3$-$ and 20$-$ $\sigma$

\[ \Delta v = \frac{\Delta v}{RMS} \]

\[ M_{RMS} = 0.48 S_{850} \left( \frac{S_{850}}{0.02 \text{ cm}^{-2} \text{ ms}^{-1}} \right) M_{\odot}, \]

where $S_{850}$ is the flux at 850 $\mu$m, $S_{850}$ is the opacity at 850 $\mu$m, and we assume a dust opacity of 0.02 cm$^{-2}$ ms$^{-1}$ (Ossenkopf & Henning 1994), dust temperature of $T_d = 15$ K and a distance to Perseus of 250 pc. These adopted values are the same used by Kirk et al. (2006).

3. CLOUD STRUCTURE

3.1. Clump Identification

Clumpfind needs only two parameters (threshold and stepsize) to decompose the emission onto a set of clumps. The threshold parameter sets the minimum emission required to be included in the decomposition, while the stepsize defines how finely separated the iso-surfaces (or iso-contours) are drawn by the algorithm in order to check for structures. In other words, threshold sets the number of pixels included in the decomposition, and stepsize sets the contrast needed between two features to be identified as different objects. Williams et al. (1994) suggest using a threshold and stepsize value of 2$-$, where $\sigma$ is the noise in the data. Clumpfind assigns all the emission above the given threshold into clumps, and it can be applied to both 2D and 3D data sets.

For this analysis, different values for both parameters are used to test the robustness of the results derived using Clumpfind. In the 3D case ($^{13}$CO), the threshold is set to 3$-$, 5$-$ and 7$-$ $\sigma$ for the original data and 5$-$ $\sigma$ for the noise-added one; while the stepsize is varied between 3$-$ and 20$-$ $\sigma$ with an spacing of 0.5$-$ $\sigma$. In the 2D case (SCUBA), the threshold is set to 3$-$, 5$-$, and 7$-$ $\sigma$; and the stepsize is varied between 2$-$ and 16.5$-$, with a spacing of 0.5$-$ $\sigma$.

Some stepsize values seem unusually large, but given the improvement on the data available larger stepizes are required to identify the largest structures in MC (see Rathborne et al. 2009)

3.2. Mass Estimate

We adopt the conversion between $^{13}$CO integrated intensity, $W$($^{13}$CO), and extinction, $A_V$, derived by Pineda et al. (2008):

\[ A_V = 0.350 W^{13CO} . \]  

This conversion is derived for Perseus using the COMPLETE extinction map and FCRAO data (assuming a main-beam efficiency of 0.49). To convert from visual extinction to column density, we assume that the ratio between N(H) and $E(B-V)$ is $5.8 \times 10^{11}$ cm$^{-2}$ mag$^{-1}$ (Bohlin et al. 1978), and $R_V = 3.1$.

For the dust continuum emission, we assume it is optically thin,

\[ M_{RMS} = 0.48 S_{850} \left( \frac{S_{850}}{0.02 \text{ cm}^{-2} \text{ ms}^{-1}} \right) M_{\odot}, \]

where $S_{850}$ is the flux at 850 $\mu$m, $S_{850}$ is the opacity at 850 $\mu$m, and we assume a dust opacity of 0.02 cm$^{-2}$ ms$^{-1}$ (Ossenkopf & Henning 1994), dust temperature of $T_d = 15$ K and a distance to Perseus of 250 pc. These adopted values are the same used by Kirk et al. (2006).

3.3. Completeness limit

For the $^{13}$CO data, the completeness limit is estimated by comparing the derived mass and radius of each clump and a sensitivity curve. The sensitivity curve is estimated as the smallest realistic clump that can be found by Clumpfind given a radius (red dash line in Figure 1). The completeness limit is then estimated as the largest mass below this sensitivity curve (for each Clumpfind run), i.e. the mass where data and red line merge in Figure 1 and shown with red filled circle. Here, a minimum size of 3 velocity channels ($\Delta v$) is assumed, and given that the brightness in each pixels must be larger than the threshold:

\[ M_{min}(R) = 0.091 \left( \frac{3 \Delta v}{\text{km s}^{-1}} \right) \left( \frac{\pi R^2}{\text{46}$''\text{r}^2} \right) \left( \frac{\xi}{K} \right) M_{\odot}, \]

where $\xi$ is the threshold. In Figure 1 two Clumpfind runs are shown: original and noise-added $^{13}$CO. In both cases the possible change in slope of the mass function happens close to the completeness limit shown by the arrow.

For simplicity, we use a single completeness limit for each dataset, a value larger than the completeness limit estimated for any of the individual Clumpfind runs: 4 and 3 $M_{\odot}$ for the original and noise-added $^{13}$CO, respectively. The completeness limit for the objects identified in the SCUBA map is estimated by Kirk et al. (2006) as 0.6 $M_{\odot}$, not as the mass where the mass function changes, but as the object that would be missed given the typical size of the cores found.

4. RESULTS

The simplest comparison between different Clumpfind runs is how many clumps are defined. In panels a and b of Figure 2 we show that the total number of clumps identified in each run (filled circles) decreases when increasing the threshold or stepsize, in either 2D or 3D. This decrease is not a surprise, because with a higher threshold there are fewer pixels available, and therefore a smaller volume to define clumps; in the case of the stepsize, a larger stepsize can miss some real structure, but also small stepsize can identify spurious clumps from structure due to noise (i.e. split a single clump into two or more because the noise creates fake structure above the stepsize level). However, Clumpfind runs with thresholds of 3$-$, 5$-$, and 7$-$ $\sigma$ can identify twice as many clumps as runs with higher thresholds and the same stepsize (see panels a and b in Figure 2), while runs with 5$-$ and 7$-$ $\sigma$ thresholds follow a similar curve (with the runs of lower threshold still finding more clumps as expected) for a given data set. It is important to note that objects identified in $^{13}$CO are not necessarily bound.

Despite the difference in the total number of clumps, the number of clumps above the completeness limit (shown as
filled circles in panels a and b of Figure 2) is comparable between different Clumpfind runs. Not only between different thresholds, but also when changing stepsize. However, the number of clumps above the completeness limit is usually less than half the total number of clumps, and therefore, most of the identified clumps are not even considered in mass function analysis.

The differential mass function, \( dN/dM \approx \Delta N_{cl}/\Delta M \), is usually approximated by a power-law, \( dN/dM \propto M^{-\alpha} \). However, if the data are binned, then variations in the fitted power-law exponent are generated by changing the bin width and shifting the bins (Rosolowsky 2005); and when analyzing the cumulative function special care must be taken to avoid the undesired effects of truncation (Muñoz et al. 2007; Li et al. 2007). To avoid both problems we perform a fit of the differential mass function, but without binning the data, by using the Maximum Likelihood Estimate (MLE, see Clauset et al. 2007).

We fit the following function,

\[
\frac{dN}{dM} = N_{cl} \frac{\alpha - 1}{M_{min}} \left( \frac{M}{M_{min}} \right)^{-\alpha},
\]

where \( M_{min} \) is the minimum mass of the sample to be used in the fitting, \( N_{cl} \) is the number of clumps more massive than \( M_{min} \), and \( \alpha \) is the power-law exponent of the distribution.
Using MLE, the exponent is estimated by

\[ \alpha = 1 + \frac{1}{N_{cl}} \left[ \sum_{i=1}^{N_{cl}} \ln \frac{M_i}{M_{min}} \right]^{-1}, \]

and the standard error on \( \alpha \) is approximated by

\[ \sigma_{\alpha} = \frac{\alpha - 1}{\sqrt{N_{cl}}}. \]

This can be estimated as a lower limit in the uncertainty, because it does not take into account uncertainties in the mass measurements. For this work we use \( M_{min} \) equals to the completeness limit.

The exponent, \( \alpha \), estimated for every Clumpfind run is shown as a function of stepsize in panels c and d in Figure 2 for \(^{13}\)CO and SCUBA, respectively. The derived values for the clump mass spectrum from Lada et al. (1991); Stutzki & Guesten (1990) (using different methods and in different regions) and Salpeter’s exponent (for the IMF) are also shown for comparison. For a standard stepsize of 3–\( \sigma \), the \(^{13}\)CO clump mass spectrum is similar to the IMF, and steeper than the values derived by previous works.

An interesting result is that the clump mass spectrum agrees (within the uncertainties of the fit) for different threshold values used if the same stepsize is used. However, most important is the fact that the estimated power-law exponent, \( \alpha \), is correlated with the stepsize. This variation in \( \alpha \) can be as high as 40%, and the correlation appears in both versions of Clumpfind: 3D and 2D. For 2D Clumpfind, we find that this correlation is negligible compared with the uncertainties associated with the fitted power-law exponent.

5. DISCUSSION

From Figure 2 we can clearly see that the power-law exponent fitted to the decomposition done by Clumpfind of the \(^{13}\)CO data is strongly correlated with the stepsize, and therefore not unique. In fact, our results show that Clumpfind is not very useful to identify small structures within a map, unless they are isolated. The reason for this correlation between fitted power-law and stepsize is that for a small stepsize less contrast is required to identify the structure, generating more but smaller objects and therefore having a steeper mass distribution; this effect is more important in crowded regions (see Figure 3). Despite the fact that the previous conclusion seems obvious, the amount of variation in the fitted exponent has typically been deemed negligible. An independent analysis carried out by Smith et al. (2008) on numerical simulations also found different results from Clumpfind when investigating the effect of the data resolution on the Clumpfind analysis. Smith et al. (2008) “observe” a numerical simulation using different spatial resolutions for the final “data”, and then run Clumpfind on them. They show that for the same region, Clumpfind identifies a different number of clumps and their derived properties are variable when using different spatial resolution.

In the 2D case, we notice that the exponent does not vary significantly with stepsize. In addition, the exponent fluctuation is almost negligible when compared with the associated uncertainties. However, Figure 3 shows an example of how different the Clumpfind decomposition is for three different input parameters. By comparing different panels in Figure 3 we see that some structures appear or are split under different parameters. These subtle differences suggest that to create a reliable catalogue manual check is needed to ensure meaningful structures. Moreover, Kainulainen et al. (2009) recently showed, using the Pipe MC extinction data, that the CMF cannot not be recovered in crowded cases.

Clumpfind is also run on the noise-added \(^{13}\)CO data with 5–\( \sigma \) threshold. The fitted power-law exponents for noise-added \(^{13}\)CO structures are similar to those derived for the original \(^{13}\)CO data only for large stepsizes (\( \sim 2 \) K). Also, only for large stepsizes the fitted power-law exponent is close to results from previous studies of the structure in MCs. But, this should not be a surprise, since Clumpfind will assign the \(^{13}\)CO extended emission into several clumps, and by adding noise the boundaries of these clumps are changed. This generates more less-massive clumps and also changes the slope of the power-law. However, there must be a point where Clumpfind identifies the largest structures in the cloud and the exponent should not change much for larger stepsize. We estimate that this effect must be less dramatic when the emission is sparse (e.g. SCUBA map, interferometer data or higher density tracer), because there is less room to change the boundaries and masses of the objects. Also, a different structure identification techniques, dendrogram (Rosolowsky et al. 2008; Goodman et al. 2009), that allows for hierarchical structure is already available and could be used to derive mass function of bound structures or any specific structure under consideration.

6. SUMMARY

The \(^{13}\)CO and 850 \( \mu \)m maps of Perseus of the COMPLETE Survey are used to study the 2D and 3D versions of the Clumpfind algorithm, respectively.

The total number of identified structures is highly correlated with the parameters used (threshold and/or stepsize). Decompositions run with a smaller threshold and stepsize produce more objects.

We use a new method to estimate the completeness limit for a sample of clumps. The mass spectrum of the identified structures, \( dN/dM \), is fitted with a power-law above the completeness limit. For the standard Clumpfind parameters, the mass function exponent is closer to Salpeter than to the classical result from Blitz (1993). Despite the small variation in the number of objects above the completeness limit, the fitted power-law exponent for \(^{13}\)CO structure is a strong function of the stepsize, while it is independent of the threshold used. The power-law exponent of SCUBA objects is also correlated with stepsize, but this effect is negligible compared to the associated uncertainties from the fitting. The \(^{12}\)CO power-law exponent variation shows that the cloud structure changes as we approach smaller scales, and that Clumpfind is still a useful tool to study the structure of a Molecular Cloud or the difference between two regions. However, this also means that it is not possible to derive a single mass function describing the sub-structure in molecular clouds when using a non-hierarchical decomposition. Most likely, a better way to study the structure in molecular clouds is by using some identification scheme that takes into account the hierarchical nature of these regions (e.g. dendrograms). In such case, mass distribution functions of the bound material could be used as an observable.

We thank Jonathan B. Foster, and Jens Kauffmann for helpful discussions. We thank the referee for useful comments that helped to improve the paper. JEP is supported by the National Science Foundation through grant #AF002 from the
Association of Universities for Research in Astronomy, Inc., under NSF cooperative agreement AST-9613615 and by Fundación Andes under project No. C-13442. This material is based upon work supported by the National Science Foundation under Grant No. AST-0407172 to AAG. EWR’s work is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship (AST-0502605) and a Discovery Grant from NSERC of Canada.

REFERENCES


FIG. 3.—Comparison of different Clumpfind runs on the NGC1333 region. Panels b, c, and d show clumps found in three different Clumpfind runs, and crosses mark the position of cores found by Kirk et al. (2006). Panel a shows the dust emission map in the NGC1333 region, with the overlaid contours at 1–, 3–, 5–, and 10–σ level; in addition the beam size (19.9′′) is shown in bottom left corner. Small changes in the parameters used generate small (but important) changes the catalogue obtained.