THE PERILS OF CLUMPFIND: THE MASS SPECTRUM OF SUBSTRUCTURES IN MOLECULAR CLOUDS

Jaime E. Pineda[1], Erik W. Rosolowsky[2], and Alyssa A. Goodman[1]

[1] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA; jpineda@cfa.harvard.edu
[2] University of British Columbia Okanagan, 3333 University Way, Kelowna, BC V1V 1V7, Canada

ABSTRACT

We study the mass spectrum of substructures in the Perseus Molecular Cloud Complex traced by ^{13}CO(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 two- and three-dimensional Clumpfind versions are tested, using 850 μm dust emission and ^{13}CO spectral-line observations of Perseus, respectively. The effect of varying threshold is not important, but varying stepsize produces a different effect for two- and three-dimensional cases. In the two-dimensional 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 three-dimensional case, however, where the ^{13}CO emission traces the bulk of the molecular cloud (MC), the number of clumps and the derived mass spectrum are highly correlated with the stepsize used. The distinction between "two dimension" and "three dimension" 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 intercomparisons, 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 ^{13}CO mass spectrum depends on the observations resolution, due to the hierarchical structure of the MC.

Key words: ISM: clouds – ISM: individual (Perseus molecular complex) – ISM: molecules – stars: formation

1. INTRODUCTION

Molecular clouds (MCs) usually have been studied using ^{12}CO and ^{13}CO(1–0) transition line maps, because they trace low-density material. When these observations are used, the emission comes from the whole MC, and the emission is "crowded." They also provide information about the velocity structure of the cloud, and we refer to them as three-dimensional data. Thanks to the new generation of bolometers, large-scale dust emission maps of entire MCs are now possible (Motte et al. 1998; Hatchell et al. 2005; Johnstone et al. 2004; Kirk et al. 2006; Enoch et al. 2006). But, due to the observing technique, much of the emission on large scales is removed, yielding a map with "sparse" emission. These dust emission maps are mostly used to find the densest objects in an MC: dense cores. However, these data do not provide velocity information to assess if a core is bound or not, and we refer these data as two dimensional. Extinction maps provide another tool to study MCs and dense cores using two-dimensional data (Cambrésy 1999; Lombardi et al. 2006). These maps give an estimate of the total column density in a region hence capturing the large-scale structures in MCs, and therefore the map is crowded. However, thanks to some post-processing techniques the extended emission can be removed, finally yielding a map with just sparse emission (Alves et al. 2007). Finally, by observing molecular lines with higher critical density or with an interferometer, most of the large-scale structure is not traced, yielding a three-dimensional data set but with sparse emission.

The structure of MCs has been studied using a variety of decomposition algorithms on ^{12}CO and/or ^{13}CO(1–0) emission maps (e.g., Stutzki & Guesten 1990; Kramer et al. 1998; Williams et al. 1994, 1995). Algorithms that decompose the MC typically take all the emission (above some threshold) and split it into clumps, which can later be easily used to calculate a mass function. In such studies, it has been shown

that the mass function of clumps follows a power law with $dN/dM \propto M^{-1.6\pm0.2}$ (Blitz 1993).

One of the most widely used cloud decomposition algorithms is Clumpfind (Williams et al. 1994), since it is readily available and has only two user-controlled parameters. It was designed to study a whole MC using three-dimensional molecular line data in a systematic fashion (Williams et al. 1994, 1995), and the data historically had coarse angular resolution allowing the study of only the largest structures in the cloud. Clumpfind has also been modified to handle two-dimensional data with sparse emission, and successfully applied to studying the core mass function (CMF; e.g., Johnstone et al. 2004; Kirk et al. 2006; Alves et al. 2007; Reid & Wilson 2005, 2006a, 2006b), and to comparing that CMF to the initial mass function (IMF) of stars which appears as an almost invariant power law (Salpeter 1955; Muench et al. 2000; Kroupa 2001): $dN/dM \sim M^{-2.35}$ for $M > 0.6\ M_\odot$. In just a few cases, molecular line data from a higher density tracer have been used to study dense cores (Ikeda et al. 2007; Walsh et al. 2007), adding velocity information to the mostly sparse emission.

In this work, we study the robustness of the mass spectrum derived using Clumpfind in crowded (^{13}CO(1–0)) and sparse emission (SCUBA 850 μm). By using the ^{13}CO(1–0) and SCUBA 850 μm data collected by the COMPLETE team in the Perseus Molecular Cloud Complex (Ridge et al. 2006; Kirk et al. 2006),[3] we are able to study the algorithm in both its three- and two-dimensional 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

---

[3] All of the data from the COMPLETE (COordinated Molecular Probe Line Extinction Thermal Emission) Survey are available online at http://www.cfa.harvard.edu/COMPLETE.
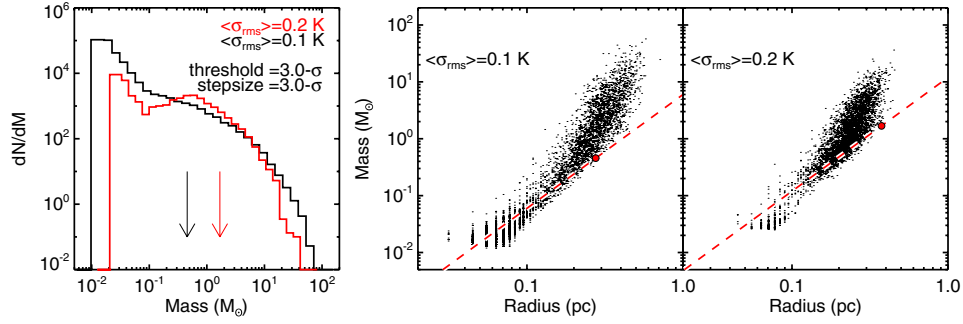
Figure 1. Comparison of clump mass functions for original and noise-added ^{13}CO data. The left panel shows the change in mass function of two Clumpfind runs using the same threshold and stepsize (in σ units), but with different noise levels. The arrow shows the corresponding completeness limit. The middle and right panels show the mass and radius for each identified clump. The red dashed line shows the sensitivity curve, and the red filled circle shows the sensitivity limit estimated for each Clumpfind run. The possible mass function turnover is close to where the identified clumps cross the sensitivity curve.

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 the T_A^* 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 rms noise in the map is 0.1 K in T_A^*, 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 μm 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 any structure larger than $\sim$2′ is removed during the data reduction process (in addition to the observational problems of detecting extended structure with bolometers), making the map mostly devoid of extended emission. Hence the SCUBA map is substantially different from the ^{13}CO data, because the later traces the more extended material. Small areas near the SCUBA map's edge are removed to avoid some image artifacts.

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 2σ, where σ is the noise in the data. Clumpfind assigns all the emission above the given threshold into clumps, and it can be applied to both two- and three-dimensional data sets.

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

is set to 3σ, 5σ, and 7σ; and the stepsize is varied between 2σ and 16.5σ, with a spacing of 0.5σ.

Some stepsize values seem unusually large, but given the improvement on the data available larger stepsizes 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):

$$A_V = 0.350 \, W(^{13}\text{CO}). \qquad (1)$$

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 $N(\text{H})$ and $E(B-V)$ is 5.8×10^{21} cm^{-2} mag^{-1} (Bohlin et al. 1978), and $R_V = 3.1$.

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

$$M_{850} = 0.48 \, S_{850} \left(\frac{\kappa_{850}}{0.02 \text{ cm}^2 \text{ gr}^{-1}} \right)^{-1} \, M_\odot, \qquad (2)$$

where S_{850} is the flux at 850 μm, κ_{850} is the opacity at 850 μm, and we assume a dust opacity of 0.02 cm^2 gr^{-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 dashed 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 a red filled circle. Here, a minimum size of three velocity channels (Δv) is assumed, and given that the brightness in each pixels must be larger than the threshold:

$$M_{\min}(R) = 0.091 \left(\frac{3\Delta v}{\text{km s}^{-1}} \right) \left(\frac{\pi R^2}{(46 \text{ arcsec})^2} \right) \left(\frac{\xi}{K} \right) \, M_\odot, \qquad (3)$$
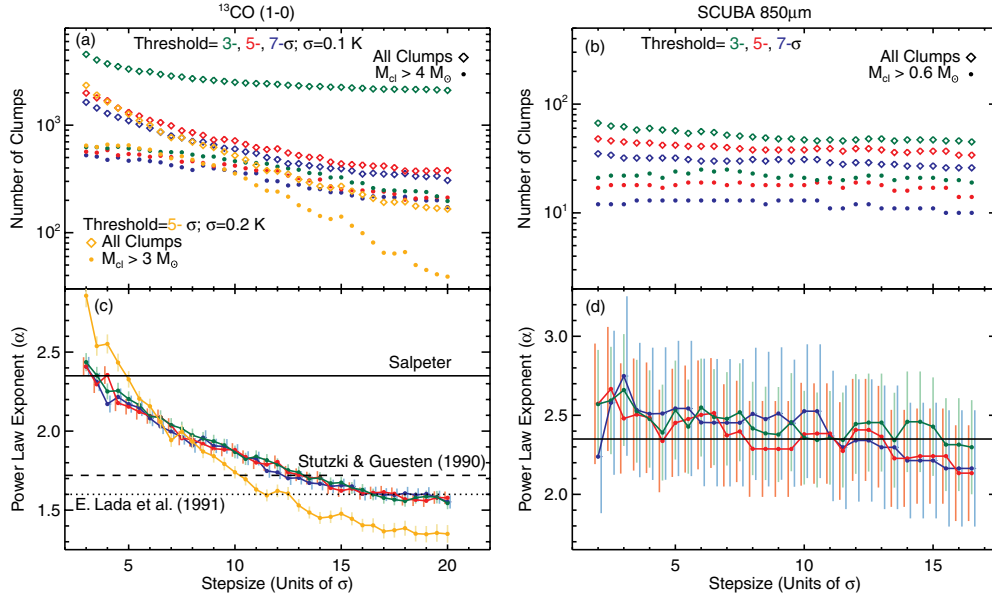
Figure 2. Summary of all Clumpfind runs as a function of stepsize. Color represent different thresholds: blue, red, and green for 3σ, 5σ, and 7σ, respectively; we also show in orange results with a threshold of 5σ for ^{13}CO data with added noise. Left and right columns show results for ^{13}CO and SCUBA data, respectively. Panels (a) and (b) show the number of clumps under a given category per model. Total number of clumps found, and total number of clumps with mass larger than the completeness limit are shown in open diamonds and filled circles, respectively. Panels (c) and (d) show the exponent of the fitted mass spectrum of clumps above the completeness limit, $dN/dM \propto M^{-\alpha}$, with error bars estimated from Equation (6). Horizontal black lines show some fiducial exponents for comparison. Average noise in ^{13}CO, ^{13}CO with added noise, and SCUBA data is 0.1 K, 0.2 K, and 0.06 Jy beam^{-1}, respectively. Completeness limit is estimated to be 4 $M_\odot$, 3 $M_\odot$, and 0.6 $M_\odot$ for ^{13}CO, ^{13}CO with added noise, and SCUBA data. Panel (c) also shows that for different noise level in the data, if a threshold of $\sim$2 K (20σ and 10σ for original and noise-added data, respectively) is used, then the fitted power-law exponents are closer to previous works.

where ξ 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 data set, a value larger than the completeness limit estimated for any of the individual Clumpfind runs: 4 $M_\odot$ 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 two or three dimensions. 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 3σ 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 5σ and 7σ 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, $dN_{cl}/dM \approx \Delta N_{cl}/\Delta M$, is usually approximated by a power law, $dN_{cl}/dM \propto M^{-\alpha}$. 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,

$$\frac{dN}{dM} = N_{cl}\frac{\alpha - 1}{M_{\min}}\left(\frac{M}{M_{\min}}\right)^{-\alpha}, \qquad (4)$$

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 α is the power-law exponent of the distribution. Using MLE, the exponent is estimated by

$$\alpha = 1 + N_{cl}\left[\sum_{i=1}^{N_{cl}}\ln\frac{M_i}{M_{\min}}\right]^{-1}, \qquad (5)$$

and the standard error on α is approximated by

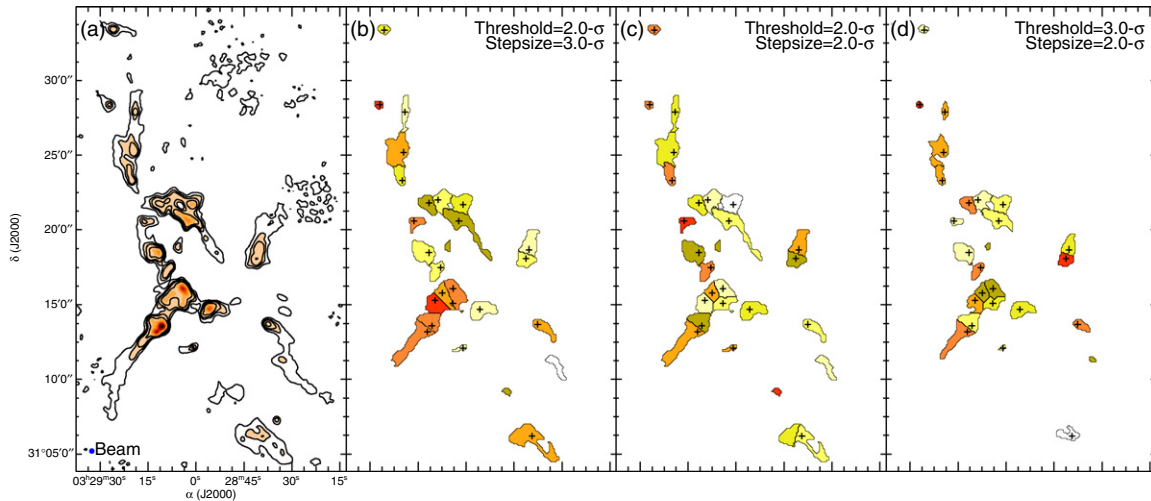$$\sigma_\alpha = \frac{\alpha - 1}{\sqrt{N_{cl}}}. \qquad (6)$$

Figure 3. Comparison of different Clumpfind runs on the NGC 1333 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 NGC 1333 region, with the overlaid contours at 1σ, 3σ, 5σ, and 10σ levels; in addition, the beam size ($19\rlap{.}''9$) is shown in the bottom left corner. Small changes in the parameters used generate small (but important) changes the catalog obtained.

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

The exponent, α, 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 3σ, 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, α, is correlated with the stepsize. This variation in α can be as high as 40%, and the correlation appears in both versions of Clumpfind: three and two dimensions. For two-dimensional 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 two-dimensional 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 catalog 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 be recovered in crowded cases.

Clumpfind is also run on the noise-added ^{13}CO data with 5σ 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 technique, 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 μm maps of Perseus of the COMPLETE Survey are used to study the two- and three-dimensional 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 ^{13}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 an MC or the difference between two regions. However, this also means that it is not possible to derive a *single* mass function describing the substructure in MCs when using a nonhierarchical decomposition. Most likely, a better way to study the structure in MCs 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.

REFERENCES

Alves, J., Lombardi, M., & Lada, C. J. 2007, A&A, 462, L17
Blitz, L. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson, AZ: Univ. Arizona Press), 125
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Cambrésy, L. 1999, A&A, 345, 965
Clauset, A., Rohilla Shalizi, C., & Newman, M. E. J. 2007, arXiv:0706.1062
Enoch, M. L., et al. 2006, ApJ, 638, 293
Goodman, A. A., Rosolowsky, E. W., Borkin, M. A., Foster, J. B., Halle, M., Kauffmann, J., & Pineda, J. E. 2009, Nature, 457, 63
Hatchell, J., Richer, J. S., Fuller, G. A., Qualtrough, C. J., Ladd, E. F., & Chandler, C. J. 2005, A&A, 440, 151
Ikeda, N., Sunada, K., & Kitamura, Y. 2007, ApJ, 665, 1194
Johnstone, D., Di Francesco, J., & Kirk, H. 2004, ApJ, 611, L45
Kainulainen, J., Lada, C. J., Rathborne, J. M., & Alves, J. F. 2009, A&A, 497, 399
Kirk, H., Johnstone, D., & Di Francesco, J. 2006, ApJ, 646, 1009
Kramer, C., Stutzki, J., Rohrig, R., & Corneliussen, U. 1998, A&A, 329, 249
Kroupa, P. 2001, MNRAS, 322, 231
Lada, E. A., Bally, J., & Stark, A. A. 1991, ApJ, 368, 432
Li, D., Velusamy, T., Goldsmith, P. F., & Langer, W. D. 2007, ApJ, 655, 351
Lombardi, M., Alves, J., & Lada, C. J. 2006, A&A, 454, 781
Motte, F., Andre, P., & Neri, R. 1998, A&A, 336, 150
Muench, A. A., Lada, E. A., & Lada, C. J. 2000, ApJ, 533, 358
Muñoz, D. J., Mardones, D., Garay, G., Rebolledo, D., Brooks, K., & Bontemps, S. 2007, ApJ, 668, 906
Ossenkopf, V., & Henning, T. 1994, A&A, 291, 943
Pineda, J. E., Caselli, P., & Goodman, A. A. 2008, ApJ, 679, 481
Rathborne, J. M., Johnson, A. M., Jackson, J. M., Shah, R. Y., & Simon, R. 2009, ApJS, 182, 131
Reid, M. A., & Wilson, C. D. 2005, ApJ, 625, 891
Reid, M. A., & Wilson, C. D. 2006a, ApJ, 644, 990
Reid, M. A., & Wilson, C. D. 2006b, ApJ, 650, 970
Ridge, N. A., et al. 2006, AJ, 131, 2921
Rosolowsky, E. 2005, PASP, 117, 1403
Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338
Salpeter, E. E. 1955, ApJ, 121, 161
Smith, R. J., Clark, P. C., & Bonnell, I. A. 2008, MNRAS, 391, 1091
Stutzki, J., & Guesten, R. 1990, ApJ, 356, 513
Walsh, A. J., Myers, P. C., Di Francesco, J., Mohanty, S., Bourke, T. L., Gutermuth, R., & Wilner, D. 2007, ApJ, 655, 958
Williams, J. P., Blitz, L., & Stark, A. A. 1995, ApJ, 451, 252
Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693