**Astronomy
&
Astrophysics**

The fidelity of the core mass functions derived from dust column density data

J. Kainulainen[1,2], C. J. Lada[3], J. M. Rathborne[3], and J. F. Alves[4]

[1] TKK/Metsähovi Radio Observatory, Metsähovintie 114, 02540 Kylmälä, Finland
  e-mail: jouni.kainulainen@helsinki.fi
[2] Observatory, PO Box 14, 00014 University of Helsinki, Finland
[3] Harvard-Smithsonian Center for Astrophysics, Mail Stop 72, 60 Garden Street, Cambridge, MA 02138, USA
[4] Calar Alto Observatory, Centro Astronomico Hispano, Alemán, C/q Jesús Durbán Remón 2-2, 04004 Almeria, Spain

ABSTRACT

Aims. We examine the recoverability and completeness limits of the dense core mass functions (CMFs) derived for a molecular cloud using extinction data and a core identification scheme based on two-dimensional thresholding. We study how the selection of core extraction parameters affects the accuracy and completeness limit of the derived CMF and the core masses, and also how accurately the CMF can be derived in varying core crowding conditions.
Methods. We performed simulations where a population of artificial cores was embedded in the variable background extinction field of the Pipe nebula. We extracted the cores from the simulated extinction maps, constructed the CMFs, and compared them to the input CMFs. The simulations were repeated using a variety of extraction parameters and several core populations with differing input mass functions and differing degrees of crowding.
Results. The fidelity of the observed CMF depends on the parameters selected for the core extraction algorithm for our background. More importantly, it depends on how crowded the core population is. We find that the observed CMF recovers the true CMF reliably when the mean separation of cores is larger than the mean diameter of the cores ($f > 1$). If this condition holds, the derived CMF for the Pipe nebula background is accurate and complete above $M \gtrsim 0.8 \ldots 1.5\ M_\odot$, depending on the parameters used for the core extraction. In the simulations, the best fidelity was achieved with the detection threshold of 1 or 2 times the rms-noise of the extinction data, and with the contour level spacings of 3 times the rms-noise. Choosing a higher threshold and wider level spacings increases the limiting mass. The simulations also show that, when $f \gtrsim 1.5$, the masses of individual cores are recovered with a typical uncertainty of $25 \ldots 30\%$. When $f \approx 1$, the uncertainty is ~60%. In very crowded cases where $f < 1$ the core identification algorithm is unable to recover the masses of the cores adequately, and the derived CMF is unlikely to represent the underlying CMF. For the cores of the Pipe nebula $f \approx 2.0$ thereby justifying the use of the method in that region.

Key words. dust, extinction – ISM: clouds – ISM: structure – stars: formation – stars: luminosity function, mass function

1. Introduction

The dense cores of interstellar molecular clouds are the precursors of stars, providing the suitable physical conditions and the mass reservoir for the star-forming process to ensue. Recently, several studies of the dust emission from such cores have presented intriguing observational evidence suggesting that the mass function of the dense cores (CMF) could be directly linked to the initial mass function of stars (IMF). In particular, the CMFs derived for nearby molecular clouds have been found to resemble the high-mass, power-law slope of the stellar IMF presented by Salpeter (1955) (Motte et al. 1998; Testi & Sargent 1998; Johnstone 2000, 2001; Motte et al. 2001; Johnstone & Bally 2006; Johnstone et al. 2006; Stanke et al. 2006; Reid & Wilson 2006a,b; Young 2006; Nutter & Ward-Thompson 2007; Enoch et al. 2008; Simpson et al. 2008). Even though the uncertainties in the slopes of the derived CMFs are often large, the slopes seem to agree with that of the stellar IMF for stellar masses greater than roughly 0.5 $M_\odot$. Because of the scale-free nature of power laws, it is difficult to ascertain the true nature

of the physical connection between the CMF and stellar IMF from the similarity of these slopes alone. However, in some of these studies, the derived CMF is also interpreted as flattening or turning over at lower masses, making the similarity between the CMF and IMF even more striking and providing a physical scaling or characteristic mass for the CMF, which can be then compared to the characteristic mass similarly measured for the IMF. However, there are considerable variations in the possible peak position of the CMF among the studies, with the estimates ranging from ~0.1 $M_\odot$ to ~8 $M_\odot$, while in comparison, the characteristic mass scale for the IMF is relatively well-determined to be ~0.5 $M_\odot$ (Kroupa 2002).

It is quite possible that a large part of the observed variation between these CMFs results from the uncertainties inherent in their construction. Deriving the CMF accurately for a molecular cloud is not trivial because of three main difficulties. First, gathering the sensitive, uniformly calibrated data required to compile a large enough sample of cores is a challenge for any current observational technique. The variations in the shapes of the derived CMFs that are reported in the studies mentioned above are

probably in part due to sampling errors resulting from the small populations considered (typically 20 . . . 100 cores). Second, deriving a robust CMF requires an accurate and unbiased determination of individual core masses over a wide mass range. This, in turn, requires an accurate observational technique for measuring core mass. However, it is often difficult to assess the uncertainties inherent in a given methodology for measuring mass and consequently the effects of these uncertainties on the derived CMF. Third, identifying cores within a cloud requires a physically meaningful and clear operational definition of a dense core. Unfortunately, there is no consensus among observers on what exactly constitutes such a definition and different definitions are usually employed in the various studies. It is not simple to assess the effect of these differing definitions on the derived CMFs.

In this paper, our goal is to determine the reliability of the CMFs and the core masses derived from 2-dimensional column density maps. For this purpose we have conducted simulations where an artificial population of dense cores is embedded into an observed (real) field of variable dust column density. The cores are then extracted using a specific core extraction (definition) scheme and a CMF constructed from them. This is then compared to the original input CMF.

In order to minimize the difficulties discussed above, we have chosen to model dust extinction observations in these numerical experiments because, in contrast to the observations of dust emission, interpreting extinction data in terms of dust column density (and core mass) does not depend on the temperature of dust grains and only very weakly on the optical properties of the grains (i.e. the extinction law), which are well known and calibrated in the near-infrared (e.g. Mathis 1990). In particular, we have chosen to simulate the dust column density conditions in the Pipe Nebula, a nearby cloud with a well-populated and smooth distribution of background stars. These properties have enabled wide-field, near-infrared extinction mapping of the cloud (Lombardi et al. 2006, LAL06 hereafter) which provides both a good sensitivity to the low contrast features and a high spatial resolution which are required to accurately measure the masses of cores over a wide mass range.

Recently, Alves et al. (2007, ALL07 hereafter) presented a study of the mass spectrum of the dense cores in the Pipe nebula. In their work, ALL07 used the extinction map of LAL06 to trace the dust column density. With a wavelet decomposition technique and an automated thresholding routine, clfind2d, they identified 159 dense cores, measured their masses and constructed the CMF for the Pipe nebula. Consistent with previous dust emission studies, ALL07 found that the CMF is surprisingly similar in shape to the stellar IMF and includes a break or flattening at low mass. The measured break in the Pipe nebula CMF suggests a characteristic core mass of $M = 2 \ldots 3\ M_\odot$, higher than the corresponding value of the stellar IMF. This observation supported the view that the IMF might originate directly from the dense core mass function after modification by a uniform star formation efficiency.

To simulate various populations of dense cores in a realistic extinction field, we create simulated extinction images of the Pipe Nebula by embedding populations of artificial dense cores into the real, variable large-scale extinction component of the Pipe Nebula. We analyze these simulated images in the same way as the observers (ALL07) and, in doing so, evaluate the fidelity of the extracted CMFs. The details of the simulations are given in Sect. 2. The results are presented and discussed in Sects. 3 and 4. In Sect. 5 we give our conclusions.

2. Simulations

The simulated CMFs were derived by extracting cores from an artificial extinction image. The core extraction was repeated using several sets of input parameters for the identification algorithm. The main characteristics of the core population, namely the form of the input core mass function and the degree of crowding, were also altered to study their effect on the derived CMF. We made numerous realizations of each simulation to lower the counting errors of the histogram bins of the derived CMF, and thus to better isolate the effect of different variables. In the following the construction of the simulated extinction maps and the core extraction procedure are explained in more detail.

2.1. Simulated extinction maps

One of the main difficulties in constructing the CMF for a molecular cloud from extinction data is disentangling the dense cores from the variable "background" extinction caused by the more diffuse dust component of the cloud. In order to simulate the effects of the variable and noisy background in a realistic manner, we used real extinction data to create the background component for the simulations. We exploited the extinction map of the Pipe nebula, derived from near-infrared color-excess data by LAL06, to create the map of the diffuse dust component. This $8° \times 6°$ wide extinction map is presented in Fig. 7 of LAL06, and it features a dynamical range of $A_V = 0 \ldots 25$ mag with the 3-σ error of 0.5 mag at the FWHM resolution of 1′. At the distance of 130 pc (LAL06) the dimensions of the map translate to $\sim 18 \times 14$ pc, and the resolution to 0.038 pc.

We extracted the dense cores initially present in the extinction map by using the wavelet decomposition technique described in ALL07. We used the technique to produce a map of structures (the "cores-only" map) at 4′ scale and subtracted it from the original extinction map. The remaining, core subtracted map was used as the large-scale background extinction component in our simulations. Thus, the background maps have the same angular extent and the same resolution as the extinction map presented by LAL06, but prominent structures at angular scales smaller than 4′ ($\sim$0.15 pc) have been largely filtered out from it. The range of extinction values in this background map is $A_V = 0 \ldots 10$ mag, and the frequency distribution of the values is similar to that of the original extinction map when $A_V < 8$ mag (see Fig. 20 of LAL06). At higher extinctions, the frequency distribution of the background map goes quickly to zero. The rms-variation in the background map is a function of position, but a typical value for it is $\sigma^{\rm rms} \approx 0.4$ mag.

To mimic the population of dense cores with non-constant column density profiles in the cloud, we embedded an assemblage of elliptical Gaussians into the map of the diffuse background extinction. The individual Gaussians, i.e. the cores, were created in the following way. First, the mass of a core was drawn randomly from the input mass function. These input mass functions were defined using power-law distributions[1], i.e. $dN/d \log M = M^\Gamma$. In particular, we used $\Gamma = -1$, $\Gamma = 0$, and a broken power-law where $\Gamma = -0.3$ ($M < 3\ M_\odot$), and $\Gamma = -1.35$ ($M > 3\ M_\odot$). The range of the input mass function was always set to $\log M = -0.55 \ldots 1.2$. Second, a radius was calculated for the core from the mass-to-radius relation $\log 2R = 0.35 \times \log M + 0.65$ that was derived for the cores of the Pipe nebula by Lada et al. (2008). In the equation, R is given in arcminutes

---

[1] We note that the mass function is calculated per unit log mass. In this notation the Salpeter (1955) mass function has the exponent $\Gamma = -1.35$.
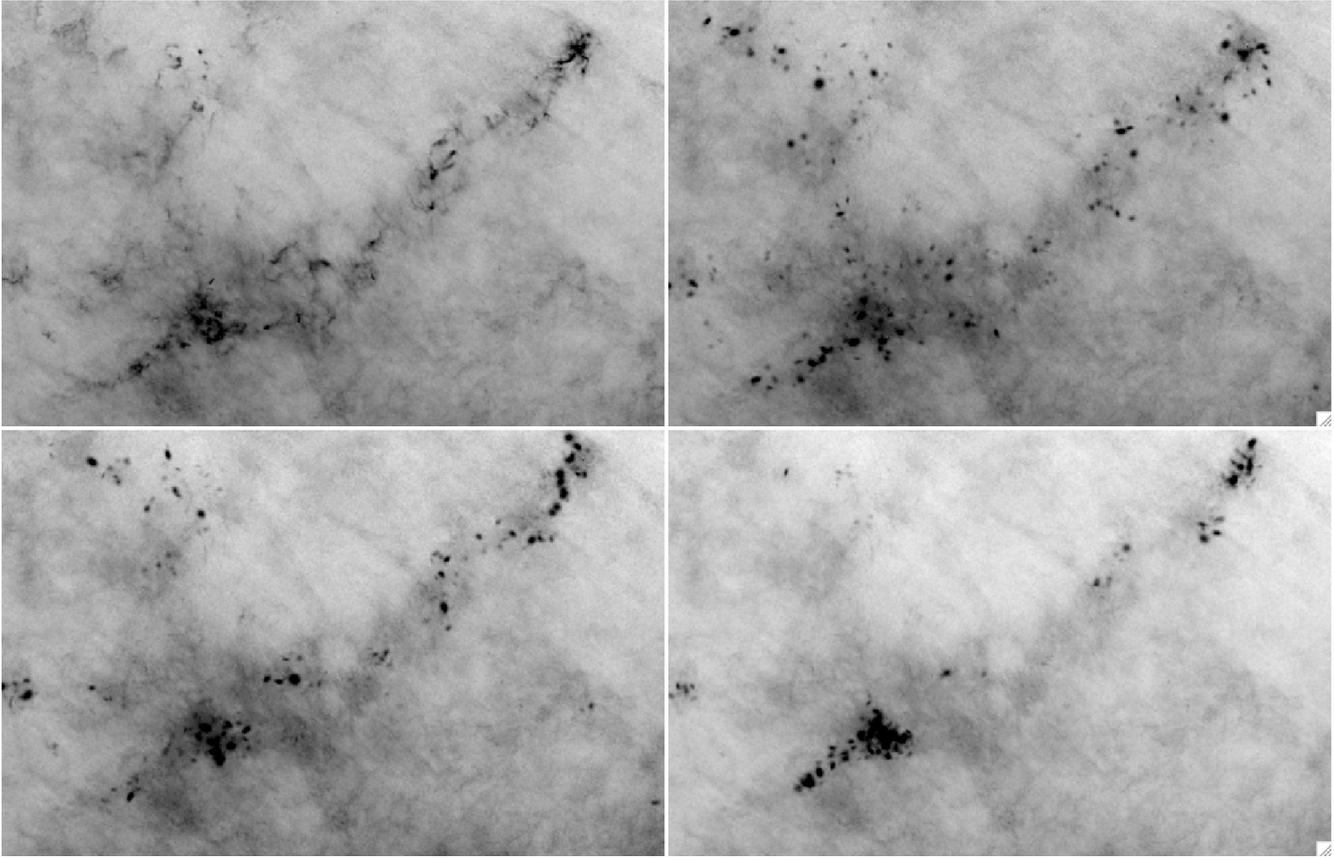
Fig. 1. Illustration of the simulated extinction data. *Top left*: the extinction map of the Pipe nebula, as presented in LAL06. The mean separation to mean diameter ratio of the cores detected by ALL07 is $f = 2.0$. *Top right*: an example of the extinction map used in the simulations. The map is composed of a diffuse background component similar to the large-scale structure of the Pipe nebula, and of a population of artificial cores (elliptical Gaussians). The figure corresponds to the simulation where $\Gamma = -1$ and $f = 1.5$. *Bottom left*: the same, but for simulation where $f = 1.0$. *Bottom right*: the same, but for simulation where $f = 0.5$.

and M in solar masses. The total mass and radius were then used to calculate the dispersion (σ) of the Gaussian profile. The mass of a core was distributed to a region within 3σ distance from its center point using the profile as a weighting function. The resulting sizes of the cores range between $\sim1.7'-6.5'$ ($0.06-0.25$ pc), and the upper end of their size range overlaps with the lower end of the range of the structures in the background map. Finally, both a uniformly distributed ellipticity ($0.5-1$) and position angle ($0-2\pi$) were applied to each of the cores. The resulting column density distribution for each core was then embedded into the background extinction map with a simple summation. In total, about 250 cores were embedded into the background map in each realization of the simulations.

The spatial distribution of the cores is one of the main parameters whose effect we want to study in this work. This is because the 2-dimensional core extraction schemes, such as clfind2d used by ALL07 and this study (see Sect. 2.2), can not easily disentangle overlapping cores, and it is likely that the derivation of CMF is severely hampered in very crowded regions. Obviously, the crowding of the cores depends not only on their separation, but also on their size. To take both of these parameters into account, we used the ratio f = mean core separation/mean core diameter as a metric to describe the crowding of cores in the simulations. By separation we refer to the distance measured from a core to its nearest neighbor.

The positions of the cores in the simulations were chosen randomly from the simulated image. The randomization was weighted with the spatial density function of the real cores identified from the Pipe nebula by ALL07. The spatial distribution of the cores in the Pipe nebula is characterized by the mean nearest neighbor distance of $9.5'$, which together with the mean core size of $4.7'$, gives the ratio $f = 2$. If the medians are used instead of means, the ratio $f = 1.7$ follows. The spatial distribution of the simulated cores can be tuned to have different ratios of f by choosing the size of the kernel function that is used to calculate the spatial density function of the cores. In this way, we generated distributions where $f = 2, 1.5, 1$, and 0.5. These cover the range where we expect both minor ($f = 2$) and substantial ($f = 0.5$) degrees of crowding.

An illustration of the simulated extinction maps is shown in Fig. 1. The figure shows the actual extinction map of the Pipe nebula, and the extinction maps for three simulations ($f = 1.5$, 1.0, and 0.5). Each simulated extinction map is a sum of the variable background component and the population of embedded cores. On a large scale, these maps are comparable to the original extinction map of LAL06.

2.2. The core extraction

We use a two-step procedure to construct the CMFs from the simulated extinction maps. The procedure is the same as used by ALL07, and it is outlined in their paper (see also Sect. 2.1 of Lada et al. 2008). First, the extinction map was decomposed with a wavelet transformation at $4'$ scale. The wavelet transform effectively removes the variations at scales larger than the selected

Table 1. The contour levels used in the core extraction.

#	A_V^{th} $[\sigma^{rms}]^a$	δA_V $[\sigma^{rms}]^b$
1	7	7
2	7	3
3	3	7
4	3	5
5	3	3
6	3	1[c]
7	2	3
8	2	1[c]
9	1	3
10	1	1[c]

[a] The detection threshold of the core extraction algorithm; [b] the contour level spacing; [c] contour increment $8\sigma^{rms}$ when $A_V > 5$ mag.

spatial scale. This results in a "cores-only" map that includes only the small-scale structure of the data. Second, the cores were identified from the cores-only map using the routine clfind2d (a 2D version of the routine clumpfind presented by Williams et al. 1994). The routine identifies the cores based on 2-dimensional contouring. The input parameters for the routine are the contour levels used in the identification, the lowest of which sets the threshold level for detection. Our main goal in this work is to examine how the selection of parameters for clfind2d affects the reliability and the completeness limit of the observed CMF, and thus we performed the extraction using ten different sets of parameters. In these sets we combined both low and high thresholds, and narrow and wide level spacings. The contour levels are listed in Table 1 and are given as multiplicatives of σ^{rms}, which is the rms-variation in extinction values within the background map ($\sigma^{rms} = 0.4$ mag). We note that ALL07 used contour levels [1.2, 4, 6] mag that are close to the lowest levels of the parameter set #3. They also required a "visual verification" of the cores to be included in the CMF.

We ran the clfind2d routine on the simulated extinction maps using all contour levels listed in Table 1. The routine outputs the characteristics of each identified core, including peak A_V and its position, total A_V, and the pixels belonging to the core. In this paper, when we refer to "an observed core" we refer to all the pixels assigned to a core by clfind2d. The masses of the cores were calculated from the total A_V within each core using the equation $M = 0.0074 \times A_V^{tot}$ $M_\odot$ mag^{-1}. This equation assumes the standard gas-to-dust ratio of $N(H)/A_V = 2 \times 10^{21}$ cm^{-2} mag^{-1} (Bohlin et al. 1978), the mean molecular mass of 2.35, and the distance of 130 pc to the Pipe nebula (LAL06). The mass functions were then constructed for each set of contour levels separately. We used the equation $N_{bins} = \Delta \log M/(2 \times n^{1/3})$ for the histogram binning, where $\Delta \log M$ is the mass range and n is the total number of cores.

3. Results

3.1. The detectability and masses of cores

Our first goal is to study how well the cores can be identified from the variable background extinction using the method described above. To do this, the crowding of cores needs to be eliminated from the simulations. For completely isolated cores, possible inaccuracies in the derived CMF result only from the inability of the method to disentangle the cores from the extended, diffuse background. To quantify this effect, we performed a series of simulations where the cores were positioned into a fixed

grid pattern so that they were isolated and did not overlap. In every other respect the simulations were performed as explained in Sect. 2.1. We also studied how the choice of contour levels for the core extraction algorithm affects the appearance of the CMFs in these simulations.

3.1.1. Fidelity and completeness limits

It should be noted that it is not trivial to present a metric that would quantify the completeness of the observed CMFs in a generally meaningful way. Ideally, the completeness at each input mass could be determined from the simulations by matching every input core with an output core (or with a non-detection). Then, the convolving function which transforms the input masses to output masses could be determined. However, this can only be done if every input core can be uniquely related to an output core (or with a non-detection). In crowded regions where several cores are blended together, and are recognized as one entity by the core extraction algorithm, unique matching may not be possible.

In this work we have made an effort to define practical mass limits, above which the observed CMFs can be regarded as good reproductions of the true CMFs. In practice, we do this simply by comparing the mass bins and the overall shape of the input and output CMFs. We will define mass limits below which less than 90% of the cores are detected, and we will refer to these limits as *fidelity limits*. We note that in a particular mass bin, having a lower number of cores than the input is not necessarily due to cores that are not detected, but may also be due to the cores whose masses are determined inaccurately. To provide insight on this, we also inspected how the observed masses of individual cores correspond to their input masses (Sect. 3.1.2).

Figure 2 shows the CMFs resulting from the simulations where the cores do not overlap. The three columns of the figure correspond to the simulations where the exponents of the input CMFs were $\Gamma = -1$, $\Gamma = 0$, and $\Gamma = [-0.3, -1.35]$ with the breakpoint at 3 $M_\odot$. The three rows show the CMFs derived using contour levels #3, #5, and #9 (see Table 1). We use these three contour levels in our illustrations throughout the paper, because they form a representative selection where we can possibly see the effects that the contour level parameters, A_V^{th} and δA_V, have on the derived CMF. The histograms of the input CMFs are also plotted in the panels. The fidelity limits, as defined earlier, are marked in the figure with dashed vertical lines. Generally, the simulations recover the shape of the input mass function well above the fidelity limits, while below them the observed CMF becomes flatter than the input CMF. The fidelity limits are not affected by crowding in this particular simulation, and thus they indicate the mass range over which the method is able to reproduce the underlying CMF for isolated cores accurately.

The fidelity limits of the simulations vary from $M \approx 0.6$ $M_\odot$ for contour levels #7−#10 to $M \approx 1.2$ $M_\odot$ for contour levels #1−#6. Thus, it appears that the CMF is recovered over a wider mass range when choosing lower thresholds and narrower level spacings for the extraction algorithm. However, there are only very subtle differences in the observed CMFs in the mass range where the CMF is recovered well with all parameter sets, i.e. at $M > 1.2$ $M_\odot$. This suggests that for isolated cores with masses $M > 1.2$ $M_\odot$, the derived CMF is insensitive to the exact choice of extraction parameters. Above the fidelity limits, the exponents of the CMFs are also accurately recovered, which was verified by fitting a power-law to the observed mass bins. In the simulations where the input CMF featured a break, the position of the observed break was also consistent with the input.
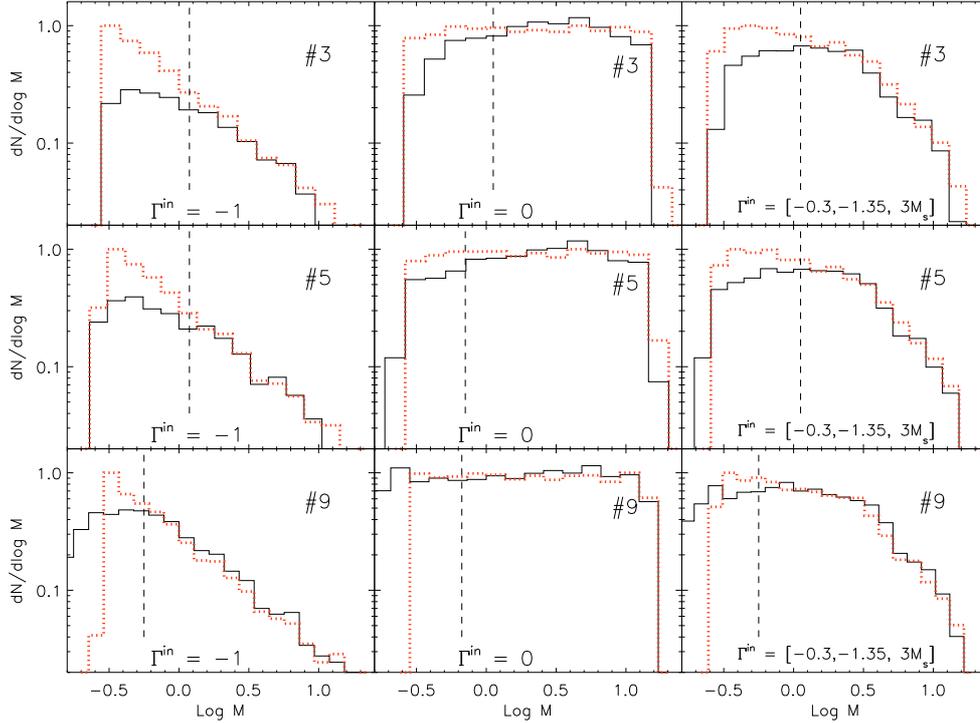
Fig. 2. The CMFs derived for three simulations where the input CMF was different. The input CMFs were $\Gamma = -1$ (leftmost column), $\Gamma = 0$ (center column), and $\Gamma = [-0.3, -1.35]$ with a break point at $3\ M_\odot$ (rightmost column). In these simulations the cores were positioned in an equally spaced grid pattern, so that they are isolated and do not overlap. The observed CMF is plotted with a solid line, and the input CMF with a dotted red line. The dashed vertical lines show the mass below which the derived CMF less than 90% complete. The CMFs derived with three different core extraction parameter sets are shown: #3 (top row), #5 (middle row), and #9 (bottom row). The histograms are normalized to the peak value of the input CMF. The range of input masses was $\log M = -0.55\ldots 1.2$. The input CMFs are always the same in each panel of the columns. However, the histograms are binned equally with the observed CMFs, and may therefore slightly differ from each other.

In the above, we defined the fidelity limits as masses below which the mass bins of the derived CMF are less than 90% complete. It should be noted that below these limits the detection efficiency of the cores does not necessarily drop immediately. Instead, the shape of the CMF may be altered due to the cores whose masses are determined incorrectly because of blending with the background or because of the inaccurate performance of the core extraction algorithm. Therefore, we also define actual *completeness limits* which reflect the mass below which the detection efficiency of the method decreases significantly.

In the simulations where the cores do not overlap, it is possible to match each input core uniquely with an observed core or with a non-detection. This was achieved by comparing the center position of each input core with the observed cores of that simulation. If there was a pixel of any observed core at the center position, the input core was counted as detected. Performing this check for all input cores within a simulation enables the examination of completeness, i.e. the ratio of the number of detected cores to the total number of cores, as a function of the input core mass. The 90% completeness mass limits resulting from this calculation were between $M = 0.45-0.7\ M_\odot$, and particularly $M = 0.7, 0.65,$ and $0.45\ M_\odot$ for the contour levels #3, #5, and #9, respectively. Below these masses, the detection efficiency decreases rapidly.

It is possible to assess the decrease in the detection efficiency caused by the wavelet filter by performing the same completeness calculation with a very low detection threshold of the clfind2d algorithm (e.g. 0.1 mag). In this case, practically all structures coming through the wavelet filter are identified as cores. Thus, if there is no observed core at the center position

of an input core, it can only be because the core has not passed the wavelet filter and, therefore, does not appear in the cores-only map at all. The mass below which the detection efficiency resulting from this calculation goes below 90% is $M \approx 0.4\ M_\odot$. Thus, the detection efficiency is dominated by the contouring algorithm rather than the wavelet filter.

3.1.2. The accuracy of detected masses

The simple comparison of the derived CMFs with the input CMFs does not provide information on whether, or how, the masses of individual cores are altered in the extraction process. It was noted in the previous section that the direct comparison does not necessarily provide meaningful fidelity limits for crowded simulations. To have another kind of metric for the feasibility of the extraction method, we calculated the ratio of observed and input masses, $M_{\mathrm{observed}}/M_{\mathrm{input}}$, for each input core.

The frequency distributions of the calculated $M_{\mathrm{observed}}/M_{\mathrm{input}}$ ratios for the simulations where $\Gamma = -1$ are shown in Fig. 3 with thick black lines. The three panels of the figure correspond to the contour levels #3, #5, and #9 (see Table 1). The shapes of the distributions are relatively similar for different contour levels and close to Gaussians in all cases. The distributions have practically the same mean and median values of 0.85. The standard deviations of $0.25\ldots 0.3$ are slightly higher for parameter sets #3 and #5 than for #9. We note that the histograms presented in Fig. 3 are composed of all cores detected in those simulations. Thus, they are dominated by the low mass cores that constitute the majority of the detected cores. The accuracy of the observed mass as a function of the input mass is
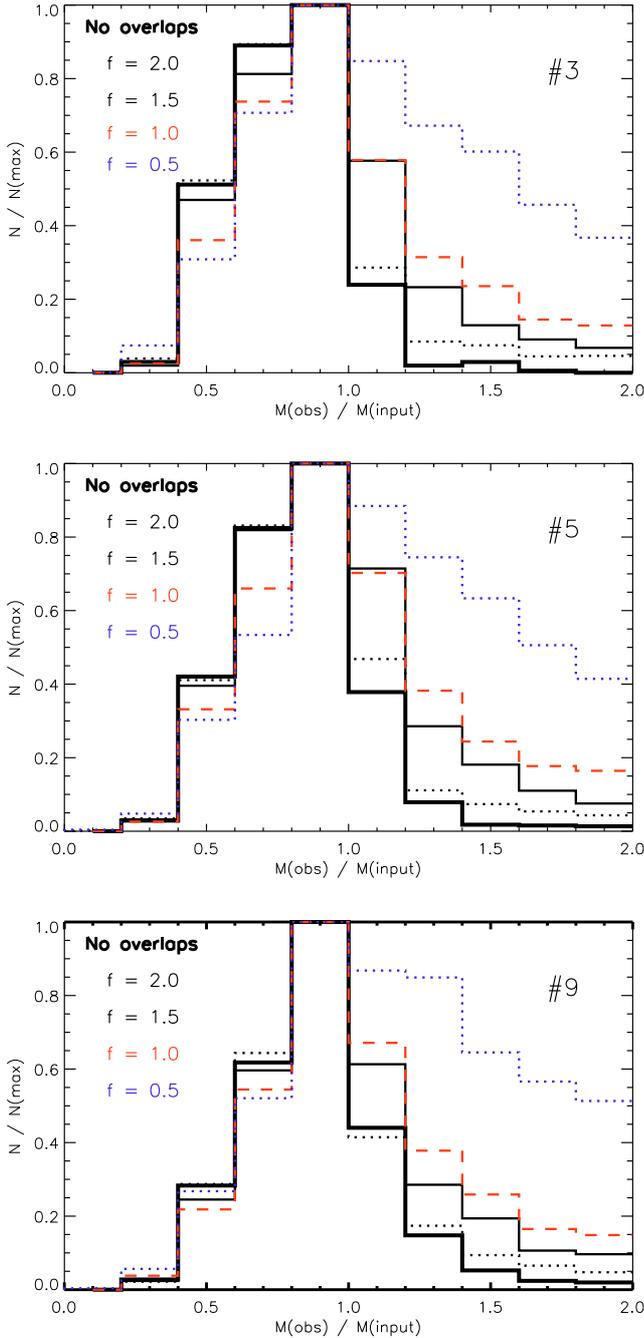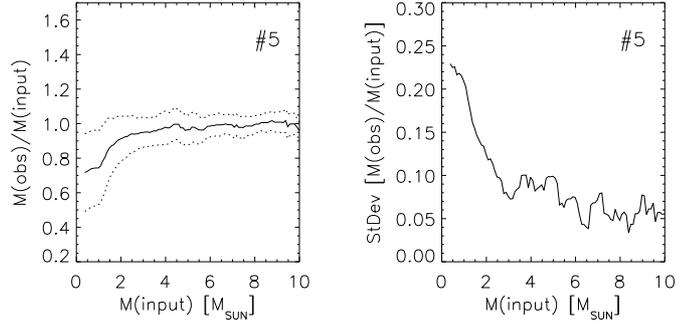
Fig. 4. Accuracy of the observed core masses as a function of the input mass of the cores in the simulation where $\Gamma = -1$ (contour level set #5). *Left*: the mean of the $M_{observed}/M_{input}$ ratio. The standard deviation of the mean is shown with dotted lines. *Right*: the standard deviation of the $M_{observed}/M_{input}$ ratio.

3.2. The effect of crowding

Because cores will not be isolated features in any real data, we need to determine the effect that core crowding has on the derived CMF. Cores along the same line of sight that partially overlap when projected in 2 dimensions may not necessarily be recognized as separate entities by the core extraction algorithm that is based only on 2-dimensional contouring. This may lead to an apparent absence of low-mass cores, and correspondingly to an overabundance of high-mass cores. In addition, the determination of the core masses suffers significantly from crowding, because in contouring algorithms each pixel is assigned uniquely to only one core, if any. As a consequence, if the crowding is severe enough, the form of the true CMF can be totally lost. Our aim is to determine the fidelity limits at different degrees of crowding, and how crowded regions can be reliably studied with this method.

To achieve this, we performed simulations where the ratio of the mean separation to the mean diameter was $f = 2.0, 1.5, 1.0$, and 0.5. Figure 5 shows the CMFs for these simulations in the case where the exponent of the input mass function is $\Gamma = -1$. The figure shows the CMFs derived with the core extraction parameter sets #3, #5, and #9 (see Table 1).

There is a clear change in the shape of the observed CMFs at different degrees of crowding. In the most sparse simulations where $f = 2$, the CMFs are well recovered above $M \gtrsim 1.5\ M_\odot$ with parameter sets #1−#4, above $M \gtrsim 1.2\ M_\odot$ with parameter sets #5−#6, and above $M \gtrsim 0.8\ M_\odot$ with parameter sets #7−#10. The simulations where $f = 1.5$ yield very similar results. The fidelity limits with these degrees of crowding are actually close to the simulations where the cores did not overlap (Fig. 2). Thus, the crowding seems to have negligible effect on the fidelity limit when $f \gtrsim 1.5$. (However, see also Sect. 3.1.2.)

When the mean separation of the cores is equal to the mean diameter of the cores ($f = 1$), the observed CMFs are more clearly altered. This is a rather logical level for the core crowding to become a severe issue for Gaussian shaped cores, because at this level the fraction of the mass of a core that is blended with neighboring cores becomes substantial compared to the original mass of the core. The effect of crowding is seen predominantly as an overabundance of cores at intermediate mass bins (log $M \approx 0.5$), which results from the growing number of cores that are blended together. The fidelity limit indicated by the direct comparison of input and output mass bins are close to those where $f > 1.0$, but we emphasize that this limit is less meaningful when the crowding is this severe. The shape of the
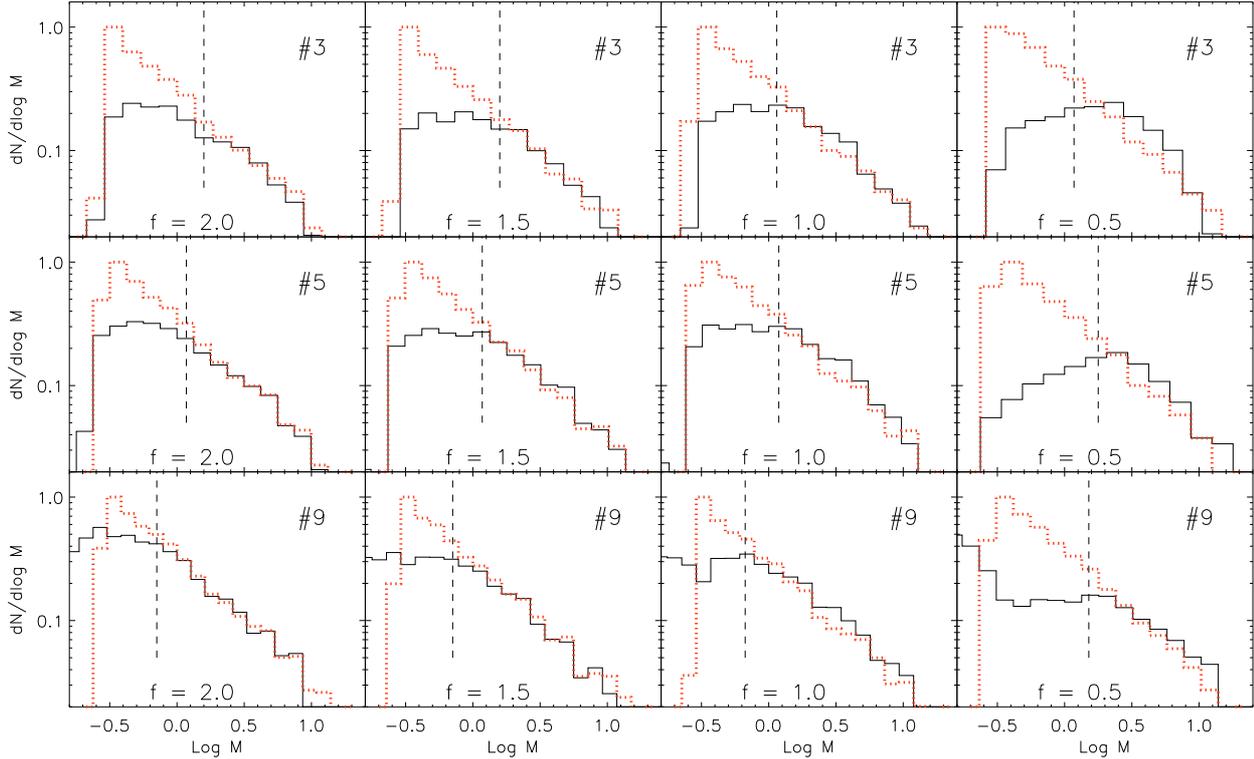

Fig. 3. Frequency distributions of the $M^{observed}/M^{input}$ ratio for the simulations where $\Gamma = -1$. The three panels show the histograms for the core extraction parameter sets #3 (*top*), #5 (*middle*), and #9 (*bottom*). Each panel shows five histograms, corresponding to different degrees of crowding: the simulations with no overlaps (thick and solid black line), and the simulations where the mean separation to diameter ratio is 2.0 (dotted black line), 1.5 (thin and solid black line), 1.0 (dashed red line), and 0.5 (dotted blue line). The histograms are normalized to their maximum values.

illustrated in Fig. 4 for one simulation (contour levels #5). The figure shows both the $M_{observed}/M_{input}$ ratio and its standard deviation as a function of input mass. The standard deviation of the $M_{observed}/M_{input}$ ratio is 5−10% for $M_{input} \gtrsim 2\ M_\odot$. Below that mass, the standard deviation increases rapidly, being 20−25% for $M_{input} < 1\ M_\odot$.

Fig. 5. The CMFs derived for the simulations with different degrees of crowding. Each row shows the CMFs for four degrees of crowding, namely $f = 2.0$, 1.5, 1.0, and 0.5. The three rows show the same CMFs derived using the core extraction parameter sets #3 (*top*), #5 (*middle*), and #9 (*bottom*). The panels show the observed CMF with a black solid line and the input CMF with a red dotted line. The dashed vertical lines show the mass below which the derived CMF less than 90% complete. The histograms are normalized to the peak value of the input CMF. The slope of the input mass function in these simulations was $\Gamma = -1$, and the range of input masses was $\log M = -0.55 \ldots 1.2$. The input CMFs are always the same in each panel of the columns. However, the histograms are binned equally with the observed CMFs, and may therefore differ slightly from each other.

derived CMF above this limit is not accurately recovered when $f = 1.0$.

In the most crowded simulation with $f = 0.5$ the observed CMFs recover the underlying mass function very inaccurately. The observed CMFs turn over at $M \approx 2.5\ M_\odot$ regardless of the core extraction parameters, and there is a serious overabundance of cores above that mass. Clearly, if the crowding is this severe, then the observed CMF does not represent the underlying CMF.

The degree of core crowding is strongly correlated with the accuracy to which the core masses are recovered (Fig. 3). In the simulations where $f = 2.0$ (dotted black line) the standard deviation of the M_{obs}/M_{input} distribution is almost similar to the case where the cores do not overlap (thick black line). However, when the core crowding is more severe ($f = 1.5$ and $f = 1.0$) the distribution begins to deviate increasingly from the Gaussian shape. The standard deviation clearly increases and the tail towards higher ratios becomes more prominent. Because of this, the mean and median values are higher. In the simulations with $f = 1.0$, the median of the ratio is 1.0 and the standard deviation is 0.6. In the most crowded simulations with $f = 0.5$ the distribution is clearly wider than the distribution for more sparsely separated cores. Thus, significant overlapping of the cores makes the reliable determination of core masses practically impossible.

4. Discussion

Near-infrared extinction data provide probably the most accurate and efficient way to trace the column density distribution of the most nearby molecular clouds over a wide spatial and

dynamical scale. Thus, A_V data can be exploited to study the CMF of dense cores. However, the derivation of the CMF based on 2-dimensional extinction data alone can be hampered by difficulties in identifying the cores and measuring their masses and sizes. The accuracy of which the identification can be done depends undoubtedly on the adopted identification scheme, and on the details of the core population itself.

In this work we studied how accurately the CMF can be derived using extinction data and the core identification scheme used recently by ALL07 to derive the CMF of the Pipe nebula. Our results, presented in Sect. 3, suggests that the CMF can indeed be determined reliably with this method, although some attention should be paid to selecting reasonable parameters for the core extraction algorithm. In our simulations, the identification performed with parameter sets #5, #7, and #9 (see Table 1) resulted in the most accurate reproduction of the input CMFs with the fidelity limits of $M = 0.8-1.5\ M_\odot$. Considering the CMF of the Pipe nebula derived by ALL07, this result confirms that the break of the power-law occuring at $M \approx 2-3\ M_\odot$ is not produced by incompleteness or by bias in the core identification method. However, the simulations suggest significantly higher fidelity limits than estimated by ALL07, and hence introduce some restrictions to the use of method.

The most important factor hampering the identification method was found to be the mean separation of the cores within the cloud. By comparing the input and output CMFs, as well as input and output masses of individual cores, we concluded that the observed CMF recovers the underlying CMF adequately if the cores are on average smaller than their separations ($f \gtrsim 1$).

This is not a totally unexpected result, considering that the identification method is based on contouring and is, therefore, unable to deal with any kind of core overlaps. The statistical uncertainty in the masses of detected cores was found to be 25−30%, provided that $f \gtrsim 1.5$. However, the core masses are systematically underestimated by about 15% on average. This is due to the low column density outer-edges of the cores that remain undetected because the extraction algorithm artificially truncates the cores according to the lowest contour level. We note the 15% systematic error represents the mean ratio of the observed and input masses in the simulations (see also Fig. 4, left panel), and may depend on the specifics of the adopted core profile.

In real clouds the core crowding could be so significant that the use of this method may be impaired. In the Pipe nebula, which is relatively quiescent and sparse, the ratio is $f \approx 2.0$. In complexes where star formation is more prominent, the cores may be located in a substantially smaller volume than they are in the Pipe nebula. For example, the active central cluster of Rho Ophiuchi covers an area of about 1200 arcmin^2 (defined roughly as a region where $A_V > 10$ mag, according to the extinction map by Lombardi et al. 2008), and contains ~100 cores (estimated from the sub-mm study of Stanke et al. 2006). We can calculate a first-order estimate for the core crowding by assuming that the region is a projection of a spherical volume inside which the cores are located. If 100 cores are distributed randomly into the volume, the sampling function of the projected nearest neighbor distribution has a mean of about 2′. If the radii of the cores are similar to the cores of the Pipe nebula, i.e. $R \approx 5′$, the ratio $f \approx 0.4$ follows. If 200 cores are distributed into the volume, following the approximate number of stars in the cluster (Wilking et al. 2005), the ratio $f \approx 0.3$ follows. This implies that the method outlined in this paper would produce unreliable results in Rho Ophiuchi using similar dust column density data as was used for the Pipe nebula by ALL07, unless the cores are on average about 3 times smaller than the cores in the Pipe nebula.

When working with data from real clouds, the determination of the CMF is affected also by another important feature specific to the extinction data. Due to the strong and variable background extinction component, there is always confusion between real cores and projected spurious core-like structures of the background component. Even though the large-scale, smoothly varying component is filtered out by the wavelet decomposition, projected structures at the scale of real cores may pass the wavelet decomposition and be included in the cores-only map. This is not a problem with high-mass cores, because they have high density contrasts with the local background. Therefore, having projected features mimicking large cores is very unlikely. In contrast, the local background may exceed the average extinction of low-mass cores by factors of several, making the occurrence of spurious structures more likely. It is not trivial to estimate the mass range where these false detections become important, but our experience with the Pipe nebula suggests that measuring the CMF accurately below $M \lesssim 0.5\ M_\odot$ would require a further analysis of this effect. As a consequence, when working with real data, one may be forced to compromise between extraction parameters yielding lower fidelity limits and those parameters that introduce a significant number of spurious cores. The confusion may also be solved by imposing some more sophisticated prerequisites, such as the detection of any associated dense gas traced by molecular lines or multiwavelength continuum observations, for a column density peak to be counted as a real core.

Finally, we comment on the applicability of our simulations. First, the background extinction field used in the simulations was always the same, i.e. the wavelet subtracted extinction map of the Pipe nebula. This field has its specific features, characterized by, for example, the frequency distribution of the extinction pixel values. Although we expect that the results of the simulations are generally applicable to other clouds, some caution is appropriate especially if the characteristics of the extended extinction component are rigorously different from those of the Pipe nebula. Second, in the simulations all the massive cores are single, well defined objects. In a real cloud, cores whose masses are significantly in excess of the Jeans' or Bonnor-Ebert critical mass are likely to exhibit substructure and may be fragmented. Under these circumstances it is not clear how to incorporate such cores into a CMF. Should these cores be counted as a single high mass core or broken up and counted as a number of lower mass cores? The strategy one adopts to deal with this issue will necessarily influence the choice of input parameters for the core identification algorithm. For example, low detection thresholds and narrow contour level spacing (e.g., contour levels #9) will tend to break up massive cores with substructure into individual components, while high detection thresholds and wide intervals (e.g., contour levels #3) will suppress substructure and identify such substructured cores as single massive objects.

Third, even though the shape of the observed CMF above the fidelity limit suggested by the simulations is not affected by the identification method, the shape is still subject to the normal counting errors. In the traditional scheme where a power-law is fitted to the logarithmically binned differential CMF, it would require $N > 500$ cores to reduce the actual error of the fitted power-law index below 0.1 (e.g. Rosolowsky 2005). In our simulations the counting errors of the CMFs were deliberately reduced, because our purpose was to examine uncertainties originating from the identification method only.

5. Conclusions

We performed simulations where we examined the recoverability of core masses and the accuracy and completeness limit of the CMF derived for a molecular cloud using dust column density data. In particular, we studied a method where the cores are extracted from the background extinction using a wavelet decomposition and a 2-dimensional thresholding routine, clfind2d. In the simulations, we embedded a population of artificial cores into a variable field of background extinction. We extracted the cores and compared the mass function constructed from them to the input mass function. We also studied how the degree of core crowding affects the fidelity limits and the accuracy of the derived core masses. The main conclusions of our work are as follows.

1. The selection of parameters for the core extraction algorithm has an effect on the accuracy and on the completeness limit of the derived CMF. However, a more important factor is the crowding of the cores. We find that the CMF and the core masses can be reliably determined if the ratio of the mean separation to the mean diameter (f) of the cores is larger than one, i.e. if the cores are on average smaller than their separations.

2. Provided that $f > 1$, the derived CMF resembles well the underlying mass function above $M > 0.8−1.5\ M_\odot$, depending on the selection of the core extraction parameters. In the simulations the lowest fidelity limit was achieved with a relatively low threshold level and narrow level spacings, namely $A_V^{th} = 1$ or $2\sigma^{rms}$, and $\delta A_V = 3\sigma^{rms}$. Choosing higher threshold and wider level spacings results in higher fidelity limits. If $f < 1$, the observed CMF may not represent the

underlying mass function, and thus the CMF cannot be reliably measured using the method studied in this paper.

3. The masses of individual cores are recovered with a typical uncertainty of $25-30\%$, provided that $f \gtrsim 1.5$. In the range $1.5 > f > 1.0$ the uncertainty is about 60%. If $f < 1$ the mass determination of individual cores is very uncertain.

References

Alves, J., Lombardi, M., & Lada, C. J. 2007, A&A, 462, L17 (ALL07)
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Enoch, M. L., Evans, N. J., II, Sargent, A. I., et al. 2008, ApJ, 684, 1240
Johnstone, D., & Bally, J. 2006, ApJ, 653, 383
Johnstone, D., Wilson, C. D., Moriarty-Schieven, G., et al. 2000, ApJ, 545, 327
Johnstone, D., Fich, M., Mitchell, G. F., & Moriarty-Schieven, G. 2001, ApJ, 559, 307
Johnstone, D., Matthews, H., & Mitchell, G. F. 2006, ApJ, 639, 259
Kroupa, P. 2002, Science, 295, 82
Lada, C. J., Muench, A. A., Rathborne, J., Alves, J. F., & Lombardi, M. 2008, ApJ, 672, 410
Lombardi, M. 2009, A&A, 493, 735
Lombardi, M., & Alves, J. 2001, A&A, 377, 1023
Lombardi, M., Alves, J., & Lada, C. J. 2006, A&A, 454, 781 (LAL06)
Lombardi, M., Lada, C. J., & Alves, J. 2008, A&A, 480, 785
Mathis, J. S. 1990, ARA&A, 28, 37
Motte, F., Andre, P., & Neri, R. 1998, A&A, 336, 150
Motte, F., André, P., Ward-Thompson, D., & Bontemps, S. 2001, A&A, 372, L41
Nutter, D., & Ward-Thompson, D. 2007, MNRAS, 374, 1413
Reid, M. A., & Wilson, C. D. 2006a, ApJ, 650, 970
Reid, M. A., & Wilson, C. D. 2006b, ApJ, 644, 990
Rosolowsky, E. 2005, PASP, 117, 1403
Stanke, T., Smith, M. D., Gredel, R., & Khanzadyan, T. 2006, A&A, 447, 609
Salpeter, E. E. 1955, ApJ, 121, 161
Simpson, R. J., Nutter, D., & Ward-Thompson, D. 2008, MNARS, 391, 205
Testi, L., & Sargent, A. I. 1998, ApJ, 508, L91
Young, K. E., Enoch, M. L., Evans, N. J. II, et al. 2006, ApJ, 644, 326
Wilking, B. A., Meyer, M. R., Robinson, J. G., & Greene, T. P. 2005, AJ, 130, 1733
Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693