



Publication Year	2018
Acceptance in OA @INAF	2020-11-19T16:04:58Z
Title	Large-scale properties of the clump mass function
Authors	OLMI, LUCA; ELIA, Davide Quintino; SCHISANO, EUGENIO; MOLINARI, Sergio
DOI	10.1093/mnras/sty1900
Handle	http://hdl.handle.net/20.500.12386/28459
Journal	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
Number	480

Large-scale properties of the clump mass function

L. Olmi,¹★ D. Elia,²★ E. Schisano² and S. Molinari²

¹INAF, Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy

²INAF, Istituto di Astrofisica e Planetologia Spaziali, Via Fosso del Cavaliere 100, I-00133 Roma, Italy

Accepted 2018 July 10. Received 2018 July 2; in original form 2018 April 13

ABSTRACT

The mass function of molecular cloud clumps and cores is widely used to compare the results of numerical simulations with the observations adopting different prescriptions for star formation. However, our ability to test different theories relies critically on our ability to measure it accurately and interpret it confidently. From an observational perspective, mass functions are subject to uncertainties due to differences in the sample selection, data reduction, and analysis techniques. To reduce these and other biasing effects, in this work we discuss methods to construct clusters or groups of sources extracted from the Hi-GAL survey, which allows us to look at large clump populations in various clouds with different physical conditions. We then construct the clump mass function of each separate group, and we fit the clump mass functions with the two most widely used functional forms, power law, and lognormal. The best-fitting parameters show no correlation with distance and positional parameters of the groups, and the mass spectra appear to have similar shapes and overall widths, though with different intrinsic mass scales. The statistical invariance of the shape of the CMF suggests that no significant differences exist in the early phases between regions dominated by low- and high-mass star formation. Finally, we found that high-mass star formation is more likely to happen in groups where the range of clump mass is shifted towards the high-mass end.

Key words: methods: data analysis – stars: formation – stars: luminosity function, mass function – ISM: structure.

1 INTRODUCTION

The origin of the stellar initial mass function (IMF) is one of the fundamental questions for the study of star formation, but it still remains poorly understood (Bastian, Covey & Meyer 2010). A critical step in understanding the origin of the IMF is to analyse whether the IMF is determined during the early phases of star formation (McKee & Ostriker 2007; Offner et al. 2014). Therefore, it is fundamental to compare the IMF to the observed mass function of young condensations like clumps (hence, clump mass function or CMF). Here, the term ‘clump’ refers to any compact density enhancement (with a size ~ 0.1 –1 pc and density $\sim 10^4$ – 10^5 cm⁻³), as identified by source-extraction algorithms. The term ‘core’ usually refers to the clump substructures (size $\lesssim 0.01$ –0.1 pc and density $\gtrsim 10^5$ cm⁻³) which, eventually, are going to form stars. It is increasingly clear that the density structure of the interstellar medium is dominated by supersonic turbulence over a wide range of scales, yielding a density distribution converging towards a lognormal probability distribution function (see e.g. McKee & Ostriker 2007). In the context of the model of gravo-turbulent fragmentation, star formation is then induced by the dissipation of large-scale turbulence to smaller scales

through radiative shocks, producing filamentary structures (Padoan & Nordlund 2002; Hopkins 2013). These shocks produce local, non-equilibrium structures with large density contrasts. Eventually, some of these dense cores collapse in gravitationally bound objects under the combined action of turbulence and gravity.

In recent years, several authors have measured the shape of the CMF (see e.g. Nutter & Ward-Thompson 2006; Reid & Wilson 2006; Sadavoy et al. 2010). In particular, Swift & Beaumont (2010) have consistently applied the same analysis to 14 data sets from the literature. Some authors have also shown that these measured CMFs have a good agreement with the shape of the stellar IMF (e.g. Alves, Lombardi & Lada 2008). Although this similarity does not imply a one-to-one relation between clumps and stars (see e.g. Offner et al. 2014; Guszejnov & Hopkins 2015), it has been suggested in the literature that the general shape of the IMF is determined in the original gaseous phase and not during the collapse process. If this tentative interpretation is true, then the initial conditions that determine the star formation process would likely originate from large-scale processes that dissipate towards smaller scales.

This scenario needs to be further tested both theoretically and observationally. As far as observations are concerned, our ability to use the CMF as a test of theories of star formation hinges crucially on our ability to measure it accurately and interpret it confidently. In fact, the observational CMFs are biased by the effects of small-

* E-mail: olmi.luca@gmail.com (LO); davide.elia@iaps.inaf.it (DE)

number statistics and certain fitting techniques. For example, the selection algorithm plays a significant role in determining the recovered distribution function. In this work, we will use existing catalogues of clumps and the discussion of the role played by the search algorithm is beyond the scope of this paper. Another important element in recovering the ‘true’ CMF is the technique of binning the data when fitting distribution functions, which may suffer from losing a lot of information. In the literature, the functional forms of the CMF most commonly employed in studies of star formation are the power-law and lognormal distributions (Reid & Wilson 2006). In this work, we will analyse and compare both functional forms.

A further step in data analysis, equally important as fitting parametrized models to the CMF, is the method used to select the region of the sky with the clumps whose mass distribution is then constructed. In general, the observed CMF is determined from the condensations found in a surveyed region of the sky. The selected region may or may not correspond to a specified volume or to an entire molecular cloud. In the assumption of a steady star formation, the CMF of a molecular cloud completely characterizes the star formation process. However, these CMFs do not take into account that clumps may have condensed locally out of their parent clouds at different epochs, and the transition of the clumps to form stars, which would bring them on the IMF while leaving the CMF, may proceed differently at different locations in a molecular cloud. Therefore, the mass function of clumps in any given region of a molecular cloud is the sum of all the local distributions, and the observed CMF may vary depending on the extent and position of the selected region (see e.g. Dib et al. 2010). In this work, we proceed differently, by selecting many clusters of clumps according to uniform criteria and analysing the resulting CMFs. Specifically, we discuss two different methods to construct groups or clusters of clumps, and apply these criteria uniformly through the Hi-GAL (*Herschel* InfraRed Galactic Plane Survey; Molinari et al. 2010) catalogue.

The outline of the paper is thus the following: in Section 2, we discuss the techniques used to construct the groups of clumps, and we discuss the general properties of these groups in Section 3. In Section 4, we then discuss different techniques for fitting parametrized models to the CMF of the selected groups. The main results from these fits are described in Section 5 and a more detailed discussion is given in Section 6. Then in Section 7 we discuss how the CMFs would change if a fragmentation model were applied, and we finally draw our conclusions in Section 8.

2 DATA SET AND GROUPING

2.1 Hi-GAL observations

Hi-GAL is an Open Time Key Project that was granted about 1000 h of observing time using the *Herschel* Space Observatory (Pilbratt et al. 2010). It delivers a complete and homogeneous survey of the Galactic plane in five continuum far-infrared (FIR) bands between 70 and 500 μm (Molinari et al. 2016). The Hi-GAL survey offers the best opportunity to look at large clump populations in various clouds with different physical conditions, while using self-consistent analysis to derive their physical parameters (see e.g. Elia et al. 2010, 2013; Olmi et al. 2013, O13 hereafter). Mass and other physical parameters of the Hi-GAL clumps were derived from a single-temperature spectral energy distribution (SED) model applied to suitable candidates in the Hi-GAL catalogue.

To extract candidate sources, a first catalogue of the inner Galaxy ($-71.0^\circ < \ell < 67.0^\circ$) based on image photometry was compiled in

each of the observed Hi-GAL fields, identifying the sources detected in the five different bands based on simple positional association. The detailed discussion on how the Hi-GAL photometric catalogue was converted to a source catalogue complete with physical properties is presented in Elia et al. (2017). These authors discuss the band-merging procedure, the search for counterparts, the assignment of distances as well as the construction, and fit of the SEDs.

2.2 Grouping of Hi-GAL sources using the clustering method

A self-consistent model should take into account the spatial and time evolution of both the CMF and IMF in a specific molecular cloud. However, when observing a population of clumps, it is very difficult to separate these multizone and multiepoch properties of the CMF, where clumps can condense out of their parent clouds over several epochs and the observed CMF may be the result of multiple local CMFs. Still, from an observational perspective, if too many selection criteria are used to choose a population of clumps, one runs immediately in the effects of small-number statistics. In this respect, the Hi-GAL survey has the advantage of many clump populations in various clouds with different physical conditions. But as we will show later, even such a large population may lead to regions with a small number of clumps.

Therefore, one of our main goals was to find an algorithm to group the many observed clumps into local clusters, each one giving rise to its own CMF independently from other physically unrelated clusters. We then try to analyse whether the selected clumps within the identified clusters are indeed physically connected or not. To this purpose we have initially tested several clustering algorithms, including the `IDL`¹ version of *k*-means clustering and `MCLUST`,² a software package for clustering analysis implementing parametrized Gaussian hierarchical clustering and the Expectation-Maximization algorithm for parametrized Gaussian mixture models (Scrucca et al. 2016). However, in the end we decided to use the customized algorithms described below because they allowed a better control over the selection parameters. In the specific case of `MCLUST`, the method becomes progressively more complicated to use and interprets the results when the number of clusters increases.

Therefore, we first describe an algorithm to search sources in the Hi-GAL catalogue and to construct clusters or groups of objects that are coherent in space. Initially, we have searched for sources that are located nearby in space by searching, for each Hi-GAL source, all related sources in the whole catalogue according to a predetermined threshold in position and distance. Associated Hi-GAL sources are added to the same group until no more sources, that lie within a maximum angular distance and velocity interval to the neighbouring sources, are found. The group is then considered to be complete and a new search around the next Hi-GAL source (not belonging to any of the already assembled groups) of the catalogue begins.

This method results in an arrangement of the Hi-GAL catalogue into groups, and sources that do not belong to any group. We have also studied the dependence of the grouping on the input parameters by varying the maximum distance and the velocity interval. As one may expect, the resulting clusters may differ as a result of these parameters and also on the exact order in the source catalogue. In particular, by increasing the angular tolerance some sources may

¹<http://www.harrisgeospatial.com/>

²<https://cran.r-project.org/web/packages/mclust/index.html>

end up in more than a single group, and we can keep track of this effect by estimating the overlapping between adjacent groups. A trade-off is finally reached between the number of clumps per group and the ‘contamination’ between different groups. However, the main problem of this algorithm is that because the threshold in the angular separation is the same for any distance to the source, the resulting groups have increasing linear sizes at larger distances.

Alternatively, we have set a threshold for the maximum linear diameter of the volume occupied by the group, and for a given reference source, we have then searched the Hi-GAL catalogue for objects (with known distance) within the spherical volume centred on the reference source and defined by the assigned linear radius. After several tests, the final selected radius was 150 pc, a few times higher than the average radius for molecular clouds found, e.g., by Rice et al. (2016) and Miville-Deschênes et al. (2017). With this choice, a total of 3610 ‘groups’ are found, including isolated clumps. This method ensures a more coherent grouping among the catalogue sources, although because of the intrinsic sensitivity and angular resolution limits of the Hi-GAL survey, the most distant groups tend to intrinsically have fewer sources and are thus more prone to statistical fluctuations. Besides to this effect, this algorithm, as well as its earlier version discussed above, does not guarantee that the groups will cover exactly the area of known giant molecular clouds (GMCs).

The top panel of Fig. 1 shows for the clustering method the resulting distribution of groups as a function of the number of sources in each group. Not surprisingly, most groups are composed of individual or few sources. With this specific selection of parameters only $\lesssim 100$ groups are composed of 30 or more sources, which clearly limits our statistical analysis despite the initially large number of sources in the Hi-GAL catalogue. In the following, we will indicate with N_{\min} the ‘acceptable’ minimum number of sources per group (which we generally set to $N_{\min} = 30$, see Section 5). We have then positionally associated the selected clusters with known GMCs from published catalogues (see also Section 2.3 and Fig. 3), and in Fig. 2, we show the number of sources in each group as a function of the total mass of the associated GMC.

2.3 Grouping using GMC catalogues

The method described in the previous section is a ‘blind’ algorithm, in the sense that the grouping of the sources does not follow a pre-determined scheme set, for example, by the density of molecular gas at a given position, by the number density of Hi-GAL sources, or other similar ancillary data. We have therefore tested a different grouping scheme based on the location of known GMCs, which are then used as ‘seeds’ for building the clusters of sources. In order to implement this method, we used two different recently published catalogues of GMCs. We first used the catalogue of Rice et al. (2016, hereafter referred to as GMC_R2016), which contains 1064 massive molecular clouds throughout the Galactic plane. Later on, during the preparation of this work, another GMC catalogue became available from Miville-Deschênes et al. (2017, hereafter referred to as GMC_M2017), containing as many as 8107 molecular clouds covering the entire Galactic plane.

Both catalogues used the ^{12}CO CfA-Chile survey by Dame, Hartmann & Thaddeus (2001) that had an angular resolution of ~ 0.25 deg, much different from that of the Hi-GAL maps, $\approx 10\text{--}35$ arcsec. The Rice et al. (2016) catalogue was obtained using a consistent dendrogram-based decomposition applied on the observed brightness, T_B , of the CO emission. In their catalogue, they were able to recover a total H_2 mass of $2.5 \times 10^8 M_{\odot}$. Miville-

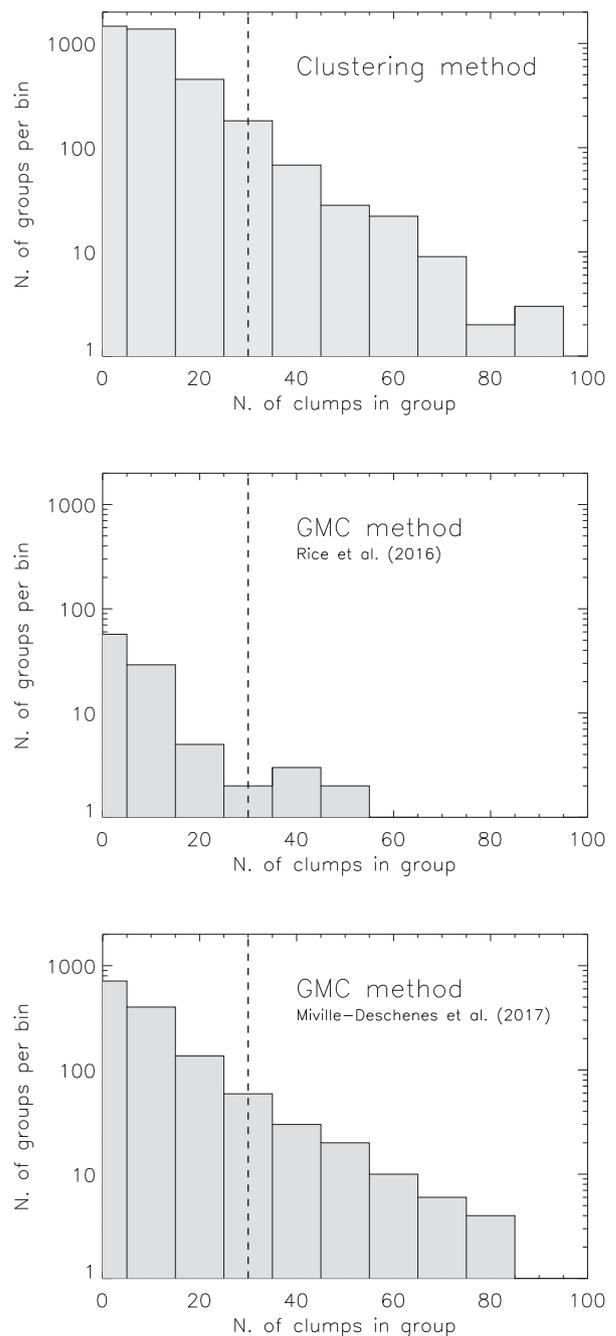


Figure 1. Histograms showing the distribution of Hi-GAL groups as a function of the number of sources in each group for the clustering method (top panel, 3610 groups in total), GMC method with the Rice et al. (2016) catalogue (middle panel, 100 groups), and GMC method with the Miville-Deschênes, Murray & Lee (2017) catalogue (bottom panel, 1383 groups). The vertical dashed line corresponds to $N_{\min} = 30$ (see the text).

Deschênes et al. (2017) also used a clustering hierarchical algorithm to identify coherent structures in a position–position–velocity cube, but the coherent structures were looked for in a more sparse description of the data resulting from a Gaussian decomposition of the CO spectra. This technique allowed them to recover a total H_2 mass of $1.2 \times 10^9 M_{\odot}$, or 98 per cent of the ^{12}CO emission observed within $b \pm 5^\circ$. Given the fragmented structure of the molecular emission, no method to identify coherent structures from the CO data gives a perfectly satisfying solution, and the identification of clouds re-

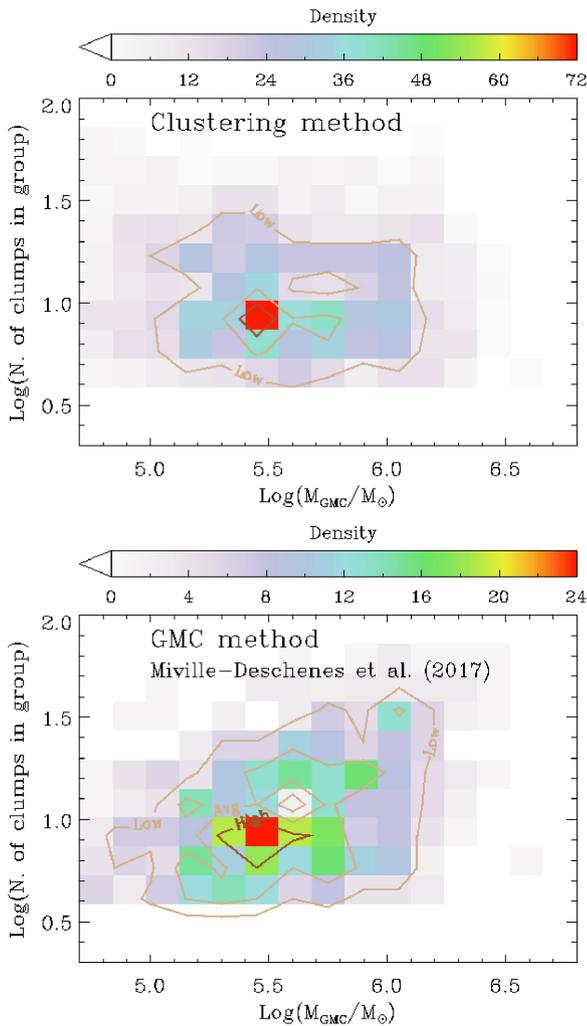


Figure 2. Density plots representing the number of clumps in each group as a function of the total mass of the associated GMC, obtained with the clustering method (*top panel*) and the GMC method (see Section 2.3) using the Miville-Deschênes et al. (2017) catalogue (*bottom panel*).

mains somewhat arbitrary. We have therefore decided to use both the GMC_R2016 and GMC_M2017 catalogues in defining coherent groups of Hi-GAL clumps, but clearly the latter is attractive because of the larger number of identified structures (1383 groups). Therefore, although we will generally show the results obtained with both catalogues, in the following our baseline results will be determined using the GMC_M2017 catalogue, whereas the GMC_R2016 should be considered as an alternative catalogue used for comparison of the main statistical results.

For both GMC catalogues, we have then adopted the central coordinates of each GMC and its measured mean size as the reference position and threshold radius, respectively, defining the spherical volume within which we have searched the Hi-GAL catalogue. This 3D search has been possible thanks to the distance assigned to the catalogue sources, as discussed by Elia et al. (2017). All sources, with known distance, identified within this volume belong to a specific GMC, and thus all groups constructed in this way are naturally associated with a known and well-defined molecular structure. For the sake of simplicity, in the current implementation of this method

the volume of the GMC has been approximated as spherical rather than ellipsoidal (as they are defined, e.g., in the catalogue by Rice et al. 2016). The bottom panel of Fig. 2 shows that the distribution of the number of clumps per GMC is quite similar to that obtained with the clustering method, and the peak occurs around GMC masses of order $3.2 \times 10^5 M_{\odot}$.

Fig. 3 shows an example of how the method described in Section 2.2 and that discussed here can yield different results. In the left-hand panel, the purple small circles show the Hi-GAL sources belonging to the group associated with the largest of the GMCs found in the $\ell = 305^{\circ}$ region at a distance of 6.82 kpc. As explained in the previous section, the groups were assembled before their association with the GMCs. The correspondence with a given GMC is then purely positional, i.e. it is determined by comparing the *centre* coordinates (ℓ , b , d), where d represents the distance, of a given group and the GMCs. The whole group and the GMC are considered to be physically associated if their centre coordinates lie to within a specified linear (and not angular) distance, which is chosen as equal to the linear diameter of the specified GMC, as reported in the catalogue. The right-hand panel of Fig. 3 shows instead the same region when the alternative grouping method described in this section is applied. The GMC_R2016 method is used, instead of the more complete catalogue by Miville-Deschênes et al. (2017), for the sake of clarity and to avoid too many overlaps on the figure. For the GMC method, the small colour-coded circles represent only those Hi-GAL sources that are found within a specified linear distance from the GMC centre, and which constitute by definition a group according to this method.

One can immediately note that the Hi-GAL groups associated with the same GMC are quite different in the two cases. Since in the clustering method the sources were aggregated according to a fixed linear scale (e.g. 150 pc), which is in general somewhat larger compared to the average molecular cloud diameter, it is not surprising that the group of purple circles in the left-hand panel of Fig. 3 is more numerous compared to that in the right-hand panel. In fact in the clustering method only if the reference coordinates of the group (i.e. the coordinates of the first source used to build a given group around it) fall within the GMC volume the group and the GMC are considered to be associated. In the left-hand panel of Fig. 3, there is only one physical association between a group and one of the GMCs in this region of the sky, but in the right-hand panel, we can see that two more sources are associated with another smaller GMC at a distance of 7.27 kpc.

The biggest advantage of the GMC method described here is that the physical association with molecular clouds is much more precise, but this leaves all other inter-cloud sources unused. In addition, this method will clearly carry with it any possible bias present in the search algorithms used by Rice et al. (2016) and Miville-Deschênes et al. (2017). On the other hand, the clustering method has the advantage of making a much more efficient use of all Hi-GAL sources, and the number of sources per group is, in general, larger than with the GMC method, as shown in Fig. 1. For this reason, in the following sections the value of N_{\min} used with the GMC_R2016 method is lower (usually, $N_{\min} = 10$ in order to obtain a statistically significant number of groups), but unless stated otherwise, N_{\min} will be the same for the clustering and GMC_M2017 methods. The disadvantage of the GMC method in terms of the lower number of sources per group is partly compensated by the fact that while with the clustering method the physical association with known GMCs is more uncertain, and therefore the resulting groups may be composed by sources that may be less uniform in terms of physical origin; in the GMC method, the a priori assignment

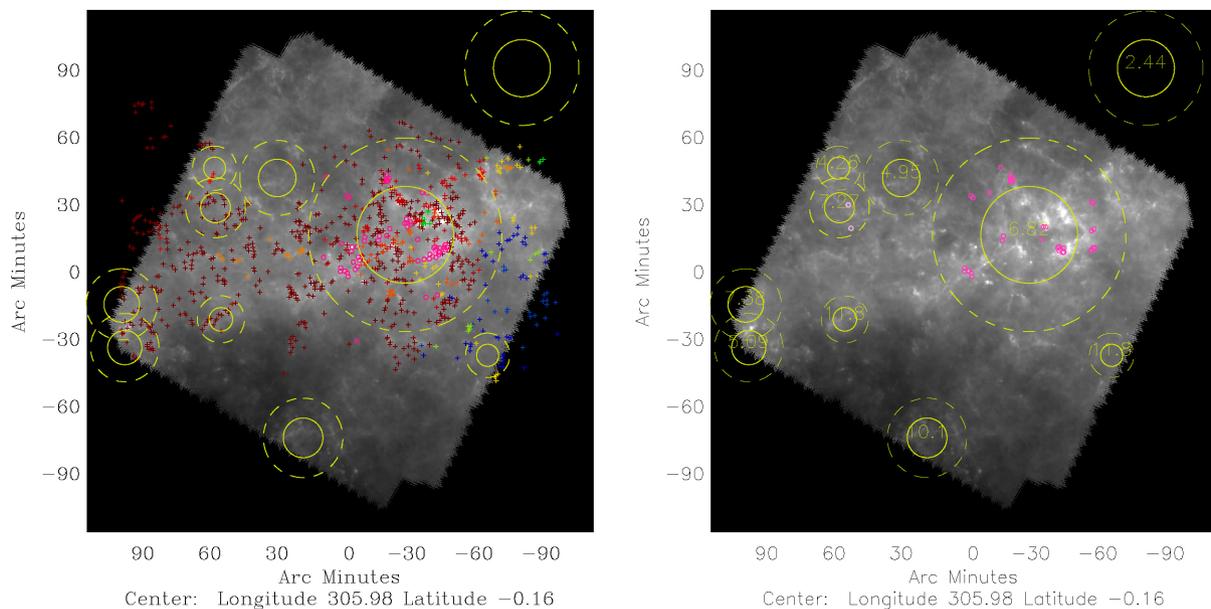


Figure 3. Example illustrating the different results obtained by applying the two grouping methods described in the text. The maps show the SPIRE 500 μm intensity observed towards the $\ell = 305^\circ$ region, with a (logarithmic) grey scale. *Left-hand panel:* clustering method. The yellow solid-line circles represent the GMCs found by Rice et al. (2016), whose radius has been estimated as \sqrt{ab} , where a and b represent the semimajor and semiminor axis lengths, while the dashed circles have twice such radius. The distance in kpc of each GMC is shown by the yellow number on the right-hand panel. The association between the GMCs and the Hi-GAL sources has been performed according to the method described in Section 2.2. The groups (shown here with small circles of a specific colour) are determined first and then the closest GMC is found. The small colour-coded crosses represent sources belonging to groups falling in the line of sight to this region but that do not belong to any of the GMCs shown here. *Right-hand panel:* GMC_R2016 method. The association between the GMCs and the Hi-GAL sources has been performed according to the method described in Section 2.3, where the groups are formed only within the GMCs of the Rice et al. (2016) catalogue. Symbols are as in the left-hand panel.

of a molecular cloud volume, where clumps must be searched, should ensure a more robust physical association with the selected clumps.

3 GALACTIC DISTRIBUTION OF THE HI-GAL GROUPS

It is of interest to analyse the distribution of the selected groups in relation to the spiral arm structure of the Milky Way as well as to the known star-forming regions. Fig. 4 shows a representation of the position of the spiral arms and we also plot the positions in the Galactic plane of the Hi-GAL groups, with groups having more sources represented by circles of increasing size. At present no distance estimates were obtained in the longitude range $-10.2^\circ \lesssim \ell \lesssim 14.0^\circ$, due to the difficulty in estimating the kinematic distances of sources in the direction of the Galactic centre (Elia et al. 2017), and therefore, the regions adjacent to the Galactic Centre are excluded from the present study, giving the appearance of a large ‘wedge’ missing between the first and fourth quadrants. The assignment of distances and the spatial distribution of the individual Hi-GAL clumps is thoroughly discussed in the papers by Elia et al. (2017) and Veneziani et al. (2017).

In Fig. 4, we show face-on views of the Milky Way representing the distribution of the selected groups with the clustering and GMC method. In every panel of this figure, each dot represents a single group, and the size of the dot is proportional to the number of sources belonging to each group. Several interesting features can

be noted. First of all, the figure shows a significant spiral structure, with many groups that fall on or near a spiral arm. However, there are also groups that do not lie near any spiral arms (more evident in the left-hand panel of Fig. 4). The clustering method also shows that the most populated groups are located within 5 kpc from the Sun. This is likely due to the fact that closer clumps are easily resolved into smaller pieces, while similar clumps that are far away are seen as a single entity.

Another interesting effect is related to the distribution of the individual clumps in the two quadrants (Elia et al. 2017). In fact, Fig. 4 shows that groups with $\gtrsim 50$ sources are more numerous in the fourth rather than in the first quadrant. Given that the number of groups (clustering method) with an estimated distance and with a number of clumps in each group $\geq N_{\min}$ in the two quadrants is approximately the same (886 and 1117 for the first and fourth quadrant, respectively), this effect appears to be real.

We also note that in the fourth quadrant there are several groups located at a Galactocentric distance ~ 10 kpc, whereas there is a lack of groups at such large distance in the first quadrant. This result is confirmed by Fig. 5, which shows the distribution of the Hi-GAL groups as a function of Galactocentric distance in the first and fourth quadrants. The overall shape of these histograms is determined by the different number of sources observed at $|\ell| \lesssim 10^\circ$, where kinematic distances are more difficult to determine (see Elia et al. 2017), and also by the location of the spiral arms (e.g. the location of the Carina-Sagittarius arm in the longitude range $289^\circ \lesssim \ell \lesssim 310^\circ$).

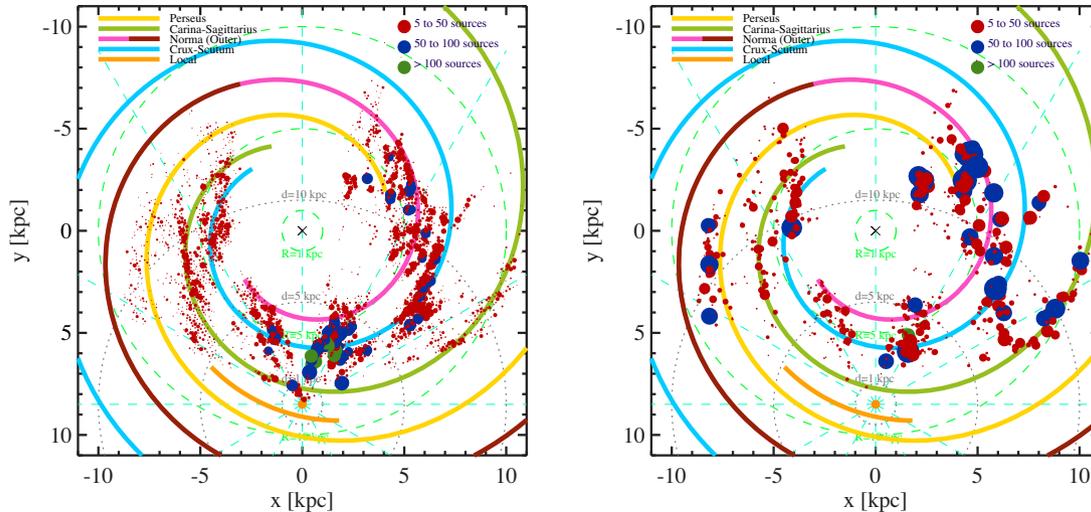


Figure 4. Plots of the position in the Galactic plane of the Hi-GAL groups selected through the clustering method (*left-hand panel*), and the GMC_M2017 (*right-hand panel*) methods. The Galactic centre is indicated with a \times symbol at coordinates $[x, y] = [0, 0]$, the Sun with an orange filled circle at coordinates $[0, 8.5]$, at the bottom of the plot. Cyan dashed lines indicate Galactic longitude in steps of 30° . R and d represent Galactocentric (light green dashed circles) and heliocentric (grey dotted circles) distances, respectively, with the following steps: 1, 5, and 10 kpc. Spiral arms, from the four-arm Milky Way prescription of Hou, Han & Shi (2009), are plotted with different colours, with the arc-colour correspondence reported in the upper-left corner of the plot. In particular, the Norma arm is represented using two colours: magenta for the inner part of the arm, and brown for the portion of it generally designated as Outer arm, whose starting point is established by comparison with Momany et al. (2006). Finally, the Local arm, not included in the model of Hou et al. (2009), is drawn, taken from Xu et al. (2016). Groups with three different ranges in terms of number of sources (5–50, 50–100, and >100) are shown as filled circles of different colours, and the number of sources in each group is proportional to the size of the circle.

4 FITTING THE MASS SPECTRA

4.1 Common functional forms of the CMF

Assuming, for the sake of simplicity, that the CMF can be approximated with a continuous function, then if dN represents the number of sources of mass M lying between M and $M + dM$, the number density distribution per mass interval (i.e. the CMF), $\xi(M) = dN/dM$, is defined through the relation (e.g. Chabrier 2003)

$$\xi(M) = \frac{dN}{dM} = \frac{\xi(\log M)}{M \ln 10} = \left(\frac{1}{M \ln 10} \right) \frac{dN}{d \log M}; \quad (1)$$

thus, $\xi(M)dM$ represents the number of objects with mass M lying in the interval $[M, M + dM]$.

The most widely used functional form for the CMF is the power law:

$$\xi_{\text{pw}}(\log M) = A_{\text{pw}} M^{-\beta}, \quad \text{or} \quad (2)$$

$$\xi_{\text{pw}}(M) = \frac{A_{\text{pw}}}{\ln 10} M^{-\alpha}, \quad (3)$$

where A_{pw} is a normalization constant (O13), and $\alpha = \beta + 1$. The original Salpeter value for the IMF is $\beta = 1.35$, $\alpha = 2.35$ (Salpeter 1955). The power-law distribution diverges as $M \rightarrow 0$ so its formal distribution cannot hold for all $M \geq 0$. We indicate the lower (or inferior) bound to the power-law behaviour with M_{inf} , i.e. the limit below which the distribution does not follow the behaviour specified by its probability density function (PDF); see O13 and Olmi et al. 2014). The top panel of Fig. 6 shows an example of a (poor) power-law fit to a group with about 80 clumps, selected using the clustering method. The dashed line represents the best fit to the mass function obtained with the PLFIT method (see Section 4.3), and *not* with a linear fit to the histogram.

Another widely used functional form for the CMF is the *log-normal*, which can be rigorously justified because the central limit theorem applied to isothermal turbulence naturally produces a log-normal PDF in density Padoan & Nordlund (2002). The continuous lognormal CMF can be written (e.g. Chabrier 2003):

$$\xi_{\ln}(\ln M) = \frac{A_{\ln}}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(\ln M - \mu)^2}{2\sigma^2} \right], \quad (4)$$

where μ and $\sigma^2 = \langle (\ln M - \langle \ln M \rangle)^2 \rangle$ denote, respectively, the mean mass and the variance in units of $\ln M$. We also define as *characteristic mass* the quantity $M_c = \exp(\mu)$.

In order to fit the lognormal functional form, we have implemented a maximum likelihood estimator (MLE)-based method similar to the PLFIT algorithm used for the power-law distribution. Therefore, we have numerically maximized the likelihood of the lognormal distribution as a function of σ and μ , using Powell's method (e.g. Press et al. 2002). An example of lognormal fit is shown in the bottom panel of Fig. 6. The sample (i.e. the clumps belonging to a specific group) is the same as that shown in the top panel, and the vertical line, representing the completeness limit, shows that the overall shape of the CMF is not affected by our sensitivity to low masses. It can be clearly seen in this case that the lognormal function gives a better fit to the data, compared to the power-law form.

In the following, we will fit both functional forms to all selected groups, for each grouping method. Therefore, we will not attempt to determine a priori the 'best' functional form, which could lead to quite arbitrary semiquantitative criteria. Instead, by fitting both functions to every CMF we will be able to evaluate a posteriori the quality of the fit using well-known statistical criteria, and thus to determine what kind of groups are best fitted by either functional form.

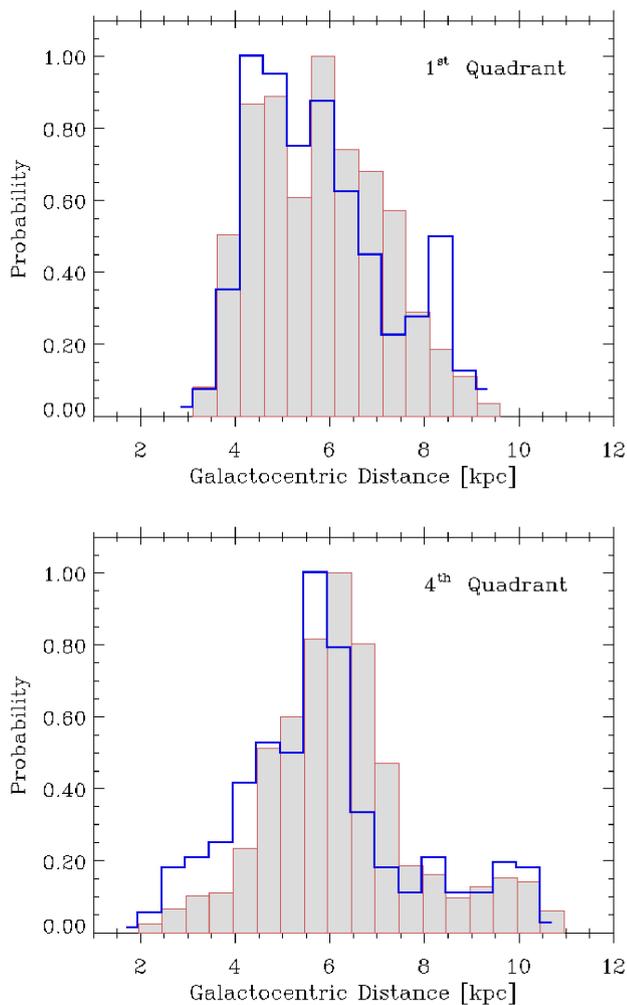


Figure 5. Histograms (normalized to the peak value) showing the distribution of Hi-GAL groups with Galactocentric radii in the first (*top panel*) and fourth (*bottom panel*) quadrants. The distribution associated with groups selected using the clustering method is represented with the filled histogram, while the outlined blue histogram represents the result obtained with the GMC_M2017 method.

4.2 Estimating the parameters of the power-law fits

In CMFs measured from real observations, the effects of incompleteness, small number statistics, and sample inhomogeneity can all lead to various degrees of uncertainty about the functional form of the CMF. In addition, the statistical analysis of the data is not always optimal. The technique of binning the data does not use every single object in the data sample, and is also very sensitive to the selected values of bin size and bin spacing (see Rosolowsky 2005; Maschberger & Kroupa 2009; and references therein).

The method of choice for fitting parametrized models such as power-law distributions to observed data is the method of maximum likelihood. In this method, the log-likelihood function for a random sample of clump masses with size N_{tot} is maximized with respect to the power-law exponent α . It is common to approximate discrete power-law behaviour with its continuous counterpart for the sake of mathematical convenience, in which case the MLE turns out to

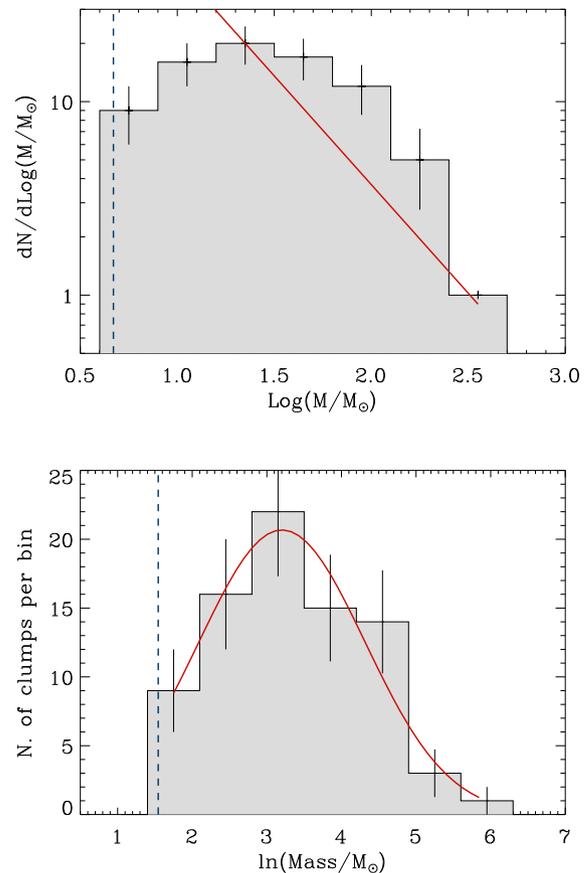


Figure 6. Examples of best fit (in red) obtained with the power-law (*top panel*) and lognormal (*bottom panel*) mass function applied to groups selected using the clustering method (Log and ln indicate the base-10 and natural logarithms, respectively). The dashed vertical lines represent the mass completeness limit at the distance of this specific group. In this example, it can be seen that the lognormal function gives a better fit to the data.

be (Clauset, Shalizi & Newman 2009):

$$\hat{\alpha} = 1 + N_{\text{inf}} \left[\sum_{i=1}^{N_{\text{inf}}} \ln \left(\frac{x_i}{x_{\text{inf}}} \right) \right]^{-1/2}, \quad (5)$$

where x_i (corresponding to the mass clump M_i in our case), $i = 1 \dots N_{\text{inf}}$, are the observed values of x such that $x_i \geq x_{\text{inf}}$ ($N_{\text{inf}} \leq N_{\text{tot}}$ should not be confused with N_{tot}). Here, $\hat{\alpha}$ denotes the estimate derived from data, whereas hatless symbols denote the true values. The parameter x_{inf} represents the lower (or inferior) bound to the data and is thus the same as M_{inf} in Section 4.1.

In order for $\hat{\alpha}$ to be a reliable estimate of α , an accurate estimate of \hat{x}_{inf} , the lower-bound, is also required. In fact, if \hat{x}_{inf} is too low then $\hat{\alpha}$ will be biased since we are attempting to fit a power-law model to non-power-law data. On the other hand, if \hat{x}_{inf} is too high then we are effectively throwing away legitimate data points $x_i < \hat{x}_{\text{inf}}$. The most common method of choosing \hat{x}_{inf} is to estimate visually the point beyond which the PDF of the distribution becomes roughly straight on a log-log plot. But this, as well as other similar approaches are clearly subjective and can be sensitive to measurement errors in M and low statistics. Alternative methods to determine $\hat{\alpha}$ based on the cumulative distribution function (CDF) are not immune to the challenge of determining \hat{x}_{inf} .

Olmi et al. (2014) applied Bayesian techniques to the statistical analysis of the CMF towards two Hi-GAL fields previously analysed by O13. Specifically, they studied the problem of estimating parameters, such as M_{inf} , that limit the range of the distribution. Olmi et al. (2014) showed that Bayesian inference may be computationally intensive and that the results may be quite sensitive to the prior type and range. As a consequence, we concluded that the extra computing time required by Bayesian techniques for the analysis of the whole Hi-GAL catalogue would not be justifiable in terms of possibly more robust statistical results. In the next section, we discuss alternative methods to approach this problem.

4.3 Comparison of methods to implement the power-law fits

The discussion of Section 4.2 shows that a more objective and principled approach is desirable for computing the best-fitting parameters of the power-law functional form. Several methods can be found in the literature that implement MLE techniques to determine $\hat{\alpha}$ and, in some cases, \hat{x}_{inf} . Here, we analyse three such methods, by Clauset et al. (2009), Maschberger & Kroupa (2009) and Rosolowsky (2005). In the work by Clauset et al. (2009), the authors discuss both the discrete and continuous MLEs for fitting the power-law distribution to data, along with a goodness-of-fit based approach to estimating the lower cut-off, \hat{x}_{inf} , of the data. Hereafter, we will refer to this procedure simply as ‘PLFIT’, from the name of the main MATLAB³ function performing the aforementioned statistical operations.

Then, as mentioned earlier, Maschberger & Kroupa (2009) analyse a variety of techniques to estimate the parameters, exponent, and lower cut-off of a power-law distribution. Rosolowsky (2005) uses the ‘errors-in-variables’ regression, rather than the ordinary least squares regression where the independent variable is assumed to be measured without error. This method maximizes the likelihood that a set of data (x, y) , with associated uncertainties (σ_x, σ_y) , can be drawn from a power-law distribution.

These three different methods are compared in Appendix A, where we conclude that the best approach is to use the MSPECFIT method, but with the value of M_{inf} determined by PLFIT. Therefore, in the following sections we will use this hybrid method, which we will refer to as MSPEC-PLFIT, to estimate the power-law index, unless stated otherwise. However, M_{inf} and the log-likelihood for the power-law fit will still be determined using PLFIT.

5 RESULTS FROM FITS TO THE OBSERVED CLUMP MASS FUNCTIONS

5.1 Power-law fits

We have applied the power-law and lognormal fitting procedure to all selected groups obtained using both the clustering and GMC method, and with a number of clumps per group higher than a minimum threshold. Unless stated otherwise, for the clustering method $N_{\text{min}} = 30$ (where N_{min} is minimum number of sources per group, see Section 2.2). The individual clumps belonging to each group were further selected based on their estimated mass. The maximum allowed mass was fixed to $\sim 7 \times 10^3 M_{\odot}$ to avoid including small-mass GMCs or even clusters of unresolved clumps. The minimum mass was selected according to the completeness limit that was estimated as described in Elia et al. (2017), using the clump

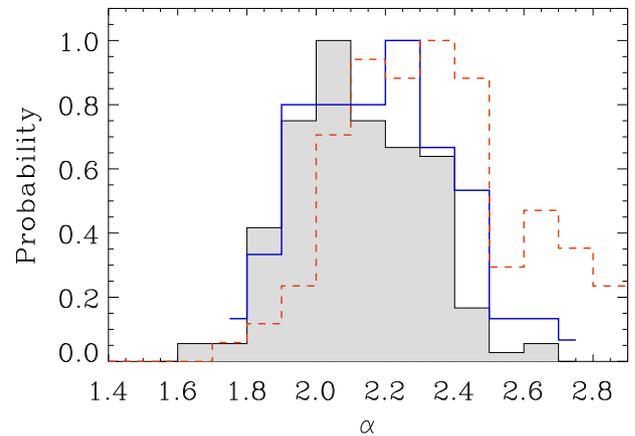


Figure 7. Histogram (normalized to the peak value) showing the distribution of the power-law best-fitting parameter α , obtained with the MSPEC-PLFIT method applied to the groups selected with the clustering (filled histogram) and GMC_M2017 (outlined blue histogram) methods. The red, dashed curve represents the distribution obtained with a different selection of the mass of the clumps within the groups obtained with the clustering method (see the text).

distance and a lower limit $S_{350} = 1$ Jy for the source flux density at $350 \mu\text{m}$. The distribution of the α best-fitting parameters, obtained with the MSPEC-PLFIT method discussed in Section 4.3, is shown in Fig. 7 for the clustering and GMC_M2017 methods. One can see that the two distributions are quite similar, despite the different number of groups found by the two methods (see Fig. 1). This similarity, like the one shown in Fig. 5, suggests that the clustering and GMC_M2017 methods manage to ‘catch’ similar clusters of clumps that are spatially (if not physically) associated.

The similarity in Fig. 7 is confirmed by the average values that are $\langle \alpha_{\text{cls}} \rangle = 2.12 \pm 0.15$ and $\langle \alpha_{\text{gmc}} \rangle = 2.12 \pm 0.16$ for the clustering and GMC methods, respectively. We note that the width of the two histograms are $\simeq 0.15$ in both cases. In order to test whether the width of the distribution gets smaller when only distributions ‘closer’ to a power law are selected, we have repeated the computation for groups with a log-likelihood less than a maximum value (specifically, $\log\text{-likelihood} \leq -400$), finding $\langle \alpha_{\text{gmc}} \rangle = 2.28 \pm 0.15$ (GMC_M2017 method). We note that while $\langle \alpha_{\text{gmc}} \rangle$ is somewhat higher compared to the previous estimates, the width of the distribution is the same. Although all of these estimates are very similar and are consistent, within the errors, with the Salpeter value, they also show a trend to be systematically lower than the Salpeter value. Therefore, our analysis confirm the general similarity between the power-law indices of the clump and stellar mass functions. Whether this similarity reinforces the idea that the stellar IMF originates in the mass function of clumps in a fragmented molecular cloud is still a matter of discussion (see e.g. Offner et al. 2014).

As an additional (semiquantitative) method to test how much the complex shape of the measured CMF may affect our previous MSPEC-PLFIT results, we have also adopted a different criterion to select the clumps within each group that will be used to apply the power-law fit. In this test, only clumps with mass $M > M_c + \frac{1}{2} \sigma$ have been selected to compute the fit in each group of clumps. As previously defined, σ and M_c represent the width and the characteristic mass, respectively, of the corresponding lognormal distribution (see Section 4.1), and in this test, we have assumed that the difference between the minimum mass M and M_c is equal to half the width of the distribution. This specific criterion should avoid to in-

³<http://www.mathworks.it/>

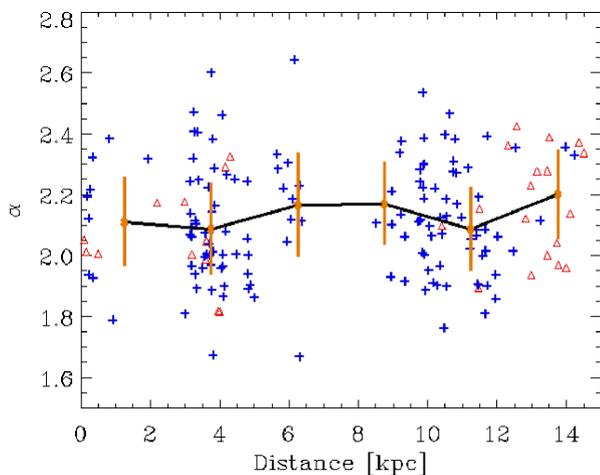


Figure 8. Plots of the best-fitting parameter α , obtained with the MSPEC-PLFIT procedure applied to the groups selected with the clustering method (and a number of clumps per group $\geq N_{\min}$), as a function of the heliocentric distance. The red triangles and the blue ‘+’ signs are associated with clumps in the first and fourth quadrants, respectively. The orange points and error bars represent the median values calculated in bins of the same width along the horizontal axis and the corresponding standard deviations.

clude in the power-law fit mass values too close to the curvature near the peak of the CMF, as determined by the lognormal method (see also Section 6.1). The red, dashed curve in Fig. 7 shows the resulting distribution obtained using the groups selected with the clustering method. Similar to the case of groups selected by their log-likelihood, this histogram is also shifted towards higher α values, as confirmed by the corresponding average $\langle\alpha_{\text{cls}}\rangle = 2.32 \pm 0.29$. All estimated values are consistent within the errors, and the last estimate is much closer to the Salpeter value, although the width of the distribution is higher.

We find no detectable trend of either α or M_{inf} with distance or Galactic longitude. As an example Fig. 8 shows the slope of the mass function for the selected groups as a function of the heliocentric distance, which in fact does not suggest any specific trend. The distance from the Sun, instead of the Galactocentric distance, has been chosen to double-check whether any bias effect due to increasing distance was present, since at larger distances one might expect to be biased towards larger (merged) clumps. We have also searched for a possible correlation of the slope, α , with the average temperature of the clumps belonging to a given group, and also with the number of sources in each group, but we found none. The lack of correlation of α and M_{inf} with the most significant physical parameters is an interesting result in its own right and suggests that the high-mass end of the CMF is on average the same irrespectively of the observed conditions in the group.

5.2 Lognormal fits

We have then applied our best-fitting procedure using the lognormal functional form to all selected groups, and we have searched for possible correlations of the best-fitting parameters, M_c and σ , with various physical parameters. Fig. 9 shows the distributions of these parameters as obtained with the clustering (filled histogram) and GMC_M2017 (blue, outlined histogram) methods. While M_c shows quite a wide distribution, $\langle M_c \rangle = 556 \pm 412 M_{\odot}$, the parameter σ has a relatively narrow distribution, $\langle \sigma / [\ln M_{\odot}] \rangle = 1.05 \pm 0.16$ (both values refer to the clustering method). In fact, this value

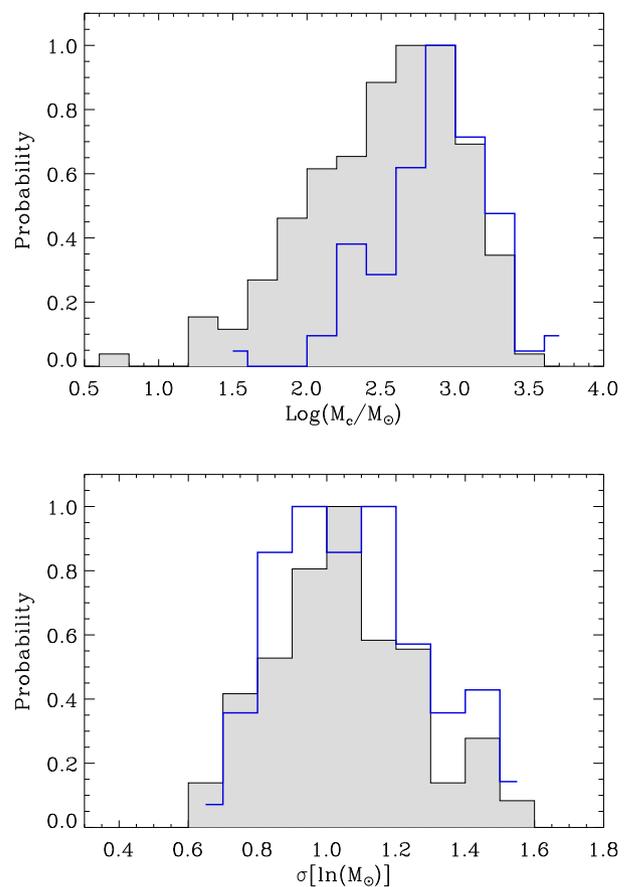


Figure 9. Same as Fig. 7 for the M_c (top panel) and σ (bottom panel) best-fitting parameters used in the lognormal fits.

corresponds to a maximum variation (at the level of one standard deviation) of the mass distribution of less than 50 per cent on a linear scale. These results suggest that, despite the very different mass scales, the shape of the CMF does not change significantly from one group to another (see also Section 5.3), as already reported by O13. We also note that the shape of the M_c distributions obtained with the clustering and GMC_M2017 methods are quite similar and the peak values coincide. The distributions of the parameter σ show some differences but the peak and width values are still quite similar.

It is important to note that no correlation is found between M_c and σ , implying that there is no bias in identifying narrow or wide distributions. Also, no correlation is found between the best-fitting parameters M_c and σ with either Galactic longitude or Galactocentric distance, as in the case of the power-law functional form. As an example of this lack of correlation in Fig. 10, we show the plot of M_c and σ (as determined from the clustering method) as a function of the heliocentric and Galactocentric distance, respectively. The characteristic mass has been plotted versus the heliocentric distance in order to analyse whether the increasing distance from the observer and the limited angular resolution may cause an artificial increase of M_c at larger distances, which is not found in Fig. 10. Likewise, using the GMC_M2017 method no correlation is found either. The best-fitting parameter σ has been plotted instead as a function of the Galactocentric distance to determine whether the typical width of the CMF may have some trend depending on the location within the Galactic plane. Even in this case, we find no specific correlation. Further associations with the physical parameters of the parent GMC are discussed in Section 6.4.

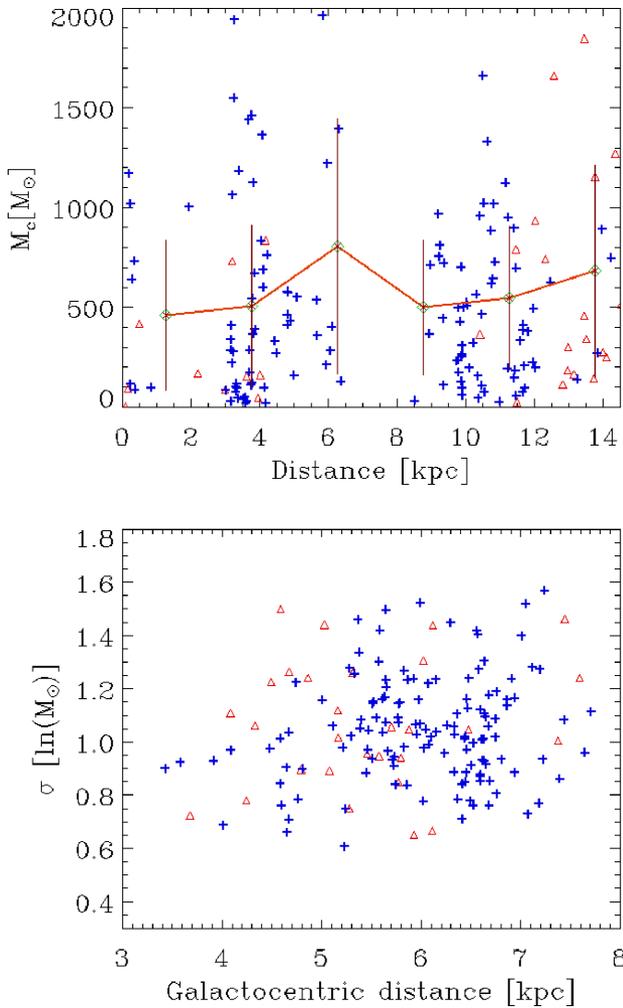


Figure 10. Plot of the best-fitting parameters M_c and σ (clustering method) as a function of the heliocentric and Galactocentric distance, respectively. Only groups with enough clumps (i.e. $\geq N_{\min}$) and which resulted in a χ^2 lower than a specified value are shown in this plot. Symbols and error bars in the top panel are same as in Fig. 8.

It is also of interest to determine whether there may be any systematic difference between the pre- and proto-stellar populations. We therefore run the MSPEC-PLFIT method on these two types of sources, decreasing the parameter $N_{\min} = 20$ to increase the number of groups selected by the clustering method for each population. We thus obtained $\langle \sigma_{\text{pre}} / [\ln M_\odot] \rangle = 0.97 \pm 0.19$ and $\langle \sigma_{\text{proto}} / [\ln M_\odot] \rangle = 1.15 \pm 0.20$ for the pre- and proto-stellar populations, respectively. Although $\langle \sigma_{\text{pre}} \rangle$ is slightly lower than $\langle \sigma_{\text{proto}} \rangle$, they are still consistent within the errors, as also already found by O13. However, we note that using the GMC_M2017 method we also obtain similar results, namely $\langle \sigma_{\text{pre}} / [\ln M_\odot] \rangle = 0.91 \pm 0.17$ and $\langle \sigma_{\text{proto}} / [\ln M_\odot] \rangle = 1.19 \pm 0.18$ for the pre- and proto-stellar populations, respectively (also with $N_{\min} = 20$). In this case, the difference is even more significant, and the fact that two independent methods yield the same trend further supports this tentative result. This trend could be the consequence of further fragmentation occurring during the passage from pre- to proto-stellar clumps, leading to a variation of the width of the distribution of the final masses. However, our data cannot be used reliably to test this scenario.

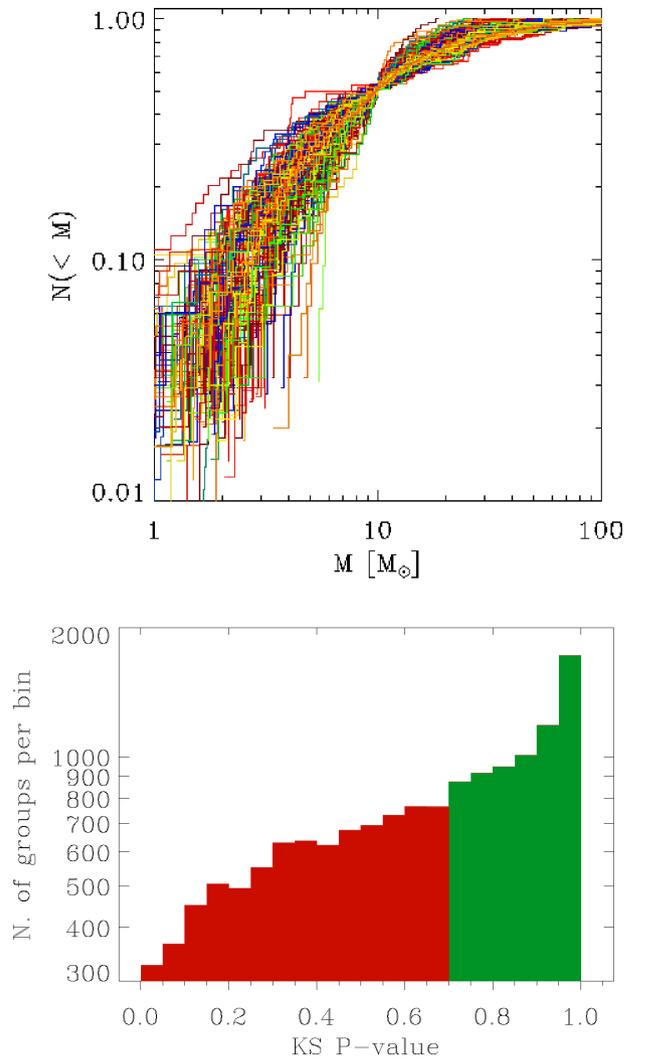


Figure 11. *Top panel:* plot of the CDF for all selected groups with the clustering method. The masses of the clumps in each CDF have been rescaled such that all groups have a common median mass ($10 M_\odot$ in this case). The new mass scale is arbitrary. *Bottom panel:* histogram showing the distribution of the Kolmogorov–Smirnov P_{ks} values for mass function comparisons. Red bins indicate those values that fall below the threshold of statistical significance ($P_{\text{ks}} = 0.72$, see the text).

5.3 Variations in the CMF

We previously mentioned in Section 5.1 the lack of correlation of the power-law fit parameter α with the positional parameters of each group of sources. Then, in Section 5.2 we also found that the lognormal fit parameter σ has a relatively narrow distribution. This apparent little region-to-region variation of both α and σ must be further analysed. If there were systematic variations in the CMF (but little variations in the observed IMF), for example between low- and high-mass star-forming regions, it would be evidence that the mechanisms for the formation of low- and high-mass stars are different. In order to facilitate the comparison between the CMF of the selected groups, in the top panel of Fig. 11 we plot the CDF of all groups (with a number of clumps $> N_{\min}$) selected with the clustering method, instead of the CMFs. To further facilitate comparisons among the different CDFs, they have all been renormalized such that they share a common (arbitrary) median clump mass, thus preserving the original shape of each CDF.

In order to more precisely quantify the extent to which the shape of the CMF varies from region to region we also used the Kolmogorov–Smirnov (KS) test, which measures the probability, P_{ks} , that two sets of clump masses represent random samplings from the same parent distribution. We compared the mass function of each group, selected with the clustering method, to the CMF of all the other groups using the KS test. The bottom panel of Fig. 11 shows the distribution of the P_{ks} values for pairwise KS test comparisons of these CMFs. To enable a comparison of the *shapes* of any two mass functions, regardless of whether they correspond to a low- or high-mass star-forming region, we renormalize each pair of CMFs as we already did for the CDFs in the top panel of Fig. 11. In the bottom panel, we plot the histogram of the P_{ks} values for a total of 14 878 inter-comparisons. As a reference P_{ks} value, we chose $P_{\text{ks}} = 0.72$ to allow a direct comparison with the analysis of Reid & Wilson (2006). We then find that in 6327 cases ($\simeq 43$ per cent) the P values are higher than 0.72 implying that there is a significant likelihood that the normalized mass functions represent random samples from the same clump mass distribution.

The differences between the computed P_{ks} values for pairwise KS test comparisons of all the selected mass functions are likely to reflect differences introduced during our data reduction and analysis (e.g. distance and position in the Galactic plane), besides to intrinsic region-to-region variations. Though we have not analysed in detail how major modifications to our procedure may finally affect the distribution of P_{ks} values, we have noted that changing the minimum number of sources per group, N_{min} , or even using the GMC method for grouping clumps, do not significantly affect the relative distribution of P_{ks} values below and above the selected threshold. The consistency of the distribution further supports the idea that the CMFs of the selected groups all appear to have very similar shapes, despite the differences in their intrinsic mass scales. We note that this similarity is intrinsic to the clump mass distributions, without any reference to the functional form of the CMF. Thus, our conclusion confirms and expands a similar result previously found by Reid & Wilson (2006) using a much smaller sample, and also mentioned by O13 using data from the science demonstration phase (SDP) of the *Herschel* satellite.

6 INTERPRETING THE FITS TO THE CMF

6.1 Correlations between power-law and lognormal fit parameters

Reid & Wilson (2006), in their study of 11 low- and high-mass star-forming regions, found that in regions where the median clump mass is less than a few solar masses, the CMF is best fitted by a lognormal distribution. In more massive regions, a double power-law provides a better fit, and at intermediate masses, both the lognormal and the double power law adequately describe the CMF. Using Bayesian inference techniques Olmi et al. (2014) found that the lognormal model better describes the CMF measured towards the two Hi-GAL regions studied for the *Herschel* SDP, though their conclusion was very much dependent on the choice of the parameter priors. Clearly, the mass function is unlikely to be a pure single power law, so the potential to be misled when fitting the CMF is significant.

Because we have analysed each group CMF with both lognormal and power-law distributions, it is therefore of interest to investigate whether any correlation exist between the best-fitting parameters of the two functional forms. As far as the power-law slopes are concerned, in their previous work Reid & Wilson (2006) found no trend in the power-law slopes with the median clump mass in

each of their regions. The top panel of Fig. C1 shows α versus the characteristic mass, M_c , which is not the same as the median clump mass. All the same we find no obvious correlation of α with M_c , supporting the original result by Reid & Wilson (2006). However, the same plot obtained with the GMC_M2017 method shows instead a weak correlation between the two parameters. Because of the low likelihood of this correlation, we cannot yet conclude that the power-law slopes derived from fits to the mass functions of submillimetre continuum clumps offer a way to distinguish between low-mass regions and very high-mass regions, but our result hints none the less to a possibility worth further investigating.

An interesting question is whether a general trend in the fits to the CMF can be observed. For example, we would like to know whether the lognormal and power-law functional forms show some kind of trend with the mass range of the groups and, more specifically, if either functional form shows better fits among the low- or high-mass star-forming regions. In order to analyse this issue, for each selected group we have constructed and fit, with the lognormal method, 10^4 Monte Carlo realizations of the CMF data. The resulting distribution of χ^2 values is then used to compute the probability, P_χ , of obtaining by chance a χ^2 value poorer than that from the lognormal best fit to the original CMF, with higher P_χ values indicating a better quality of fit. An example of CDF for the χ^2 distribution obtained with the procedure outlined above for a specific group selected with the clustering method is shown in the top panel of Fig. 12. The figure also shows the best-fitting χ^2 value corresponding to the lognormal fit of the original CMF of the selected group. Then, in this particular case the probability of obtaining by chance a χ^2 value larger than that from the lognormal best fit is $P_\chi \simeq 0.65$.

The bottom panel of Fig. 12 then shows the P_χ value as a function of the M_c of each group selected with the clustering method. Contrary to the result reported by Reid & Wilson (2006) (who used only 11 groups in their analysis), we do not find that there is a tendency towards better (higher P_χ) lognormal fits to the CMFs in less massive star-forming regions (i.e. lower M_c), since the distribution of the P_χ values appears quite uniform across the almost two orders of magnitude of the explored mass range. To further investigate this issue, we have also increased the threshold, N_{min} , for the selected groups, from 30 to 50, but without noticing any change in the overall distribution. Finally, we have repeated the procedure using the GMC_M2017 method, but even in this case, we find no apparent trend of P_χ value as a function of M_c . We are then led to the conclusion that the lognormal fit to the CMF is, in general, satisfactory in both low- and high-mass star-forming regions.

6.2 Average stability of clumps

In this section, we analyse whether any relation exists between our best-fitting parameters to the CMF and two typical masses that are frequently associated with the fragmentation and stability properties of the host GMC, i.e. the Jeans mass and Larson’s mass.

In the turbulent fragmentation scenario, star-forming clumps can no longer be regarded as equilibrium configurations and there is a distribution of local Jeans masses determined by the (lognormal) probability distribution function of the gas density. During the dissipation of supersonic turbulence, which is responsible for the fragmentation of molecular clouds, a higher fragmentation length (and thus a higher M_J) would imply that only objects with a mass $\gtrsim M_J$ would have a significant probability to form and, eventually, to collapse into stars under the action of self-gravity. In this scenario, higher Jeans masses would presumably drive the observable M_c towards higher values (see McKee & Ostriker 2007 and refer-

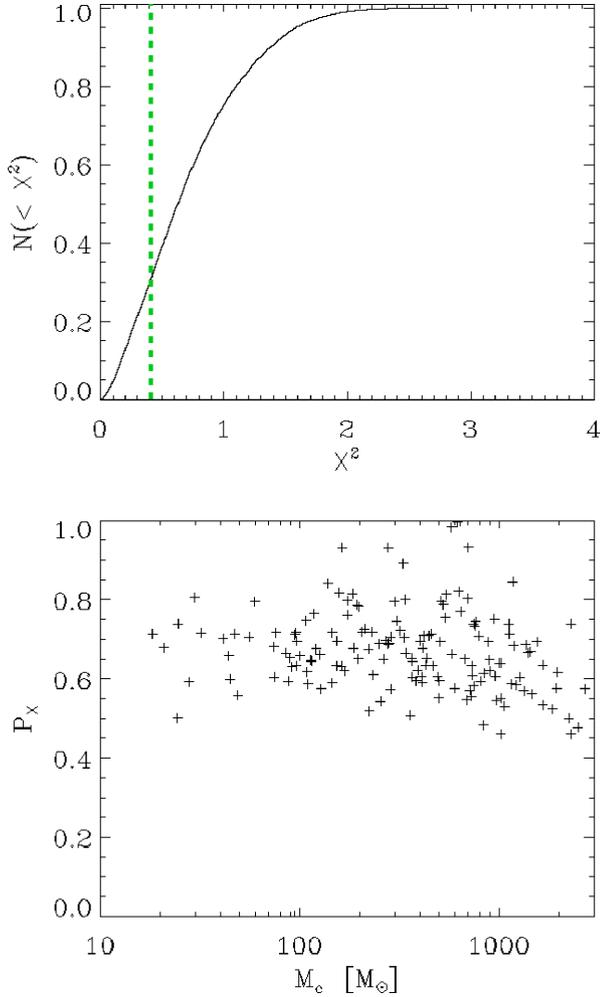


Figure 12. *Top panel:* example of CDF for the χ^2 values obtained from 10^4 Monte Carlo realizations of the CMF data (see the text). The green, dashed line marks the best-fitting χ^2 value obtained with a lognormal fit to the original CMF of the selected group. In this particular case, $P_\chi \simeq 0.65$. *Bottom panel:* plot of the P_χ values versus M_c , for groups selected using the clustering method.

ences therein). Indeed, we have found a trend of larger M_c with increasing Jeans masses when using the clustering method but not with the GMC_M2017 method. A similar problem arises with the temperature, as discussed in Section 6.3.

Fig. 13 shows instead that a correlation clearly exists between the characteristic mass M_c and the so-called Larson’s mass, M_{lars} . Following Elia et al. (2017), clumps of radius R and mass over the threshold $M_{\text{lars}} = 460 M_\odot (R/1 \text{ pc})^{1.9}$ are considered gravitationally bound. Similarly to $\langle M_J \rangle$, also $\langle M_{\text{lars}} \rangle$ in Fig. 13 represents the average of all M_{lars} calculated for the clumps belonging to a given group. We note that since the individual Hi-GAL clump masses, as estimated from the SED fitting (see Elia et al. 2017), are $M_{\text{clump}} \propto R^2$, it is possible that the good correlation between M_c and $\langle M_{\text{lars}} \rangle$ is partly the result of $M_{\text{clump}} \propto M_{\text{lars}} \sim R^2$. However, M_c is derived from a lognormal fit to a range of clump masses, under conditions where the lognormal fit is or is not the best-fitting function (see Section 6.1). The fact that the same correlation between M_c and $\langle M_{\text{lars}} \rangle$ is also observed when the GMC method is used, and in particular, the fact that $M_c > \langle M_{\text{lars}} \rangle$ constitutes further evidence that

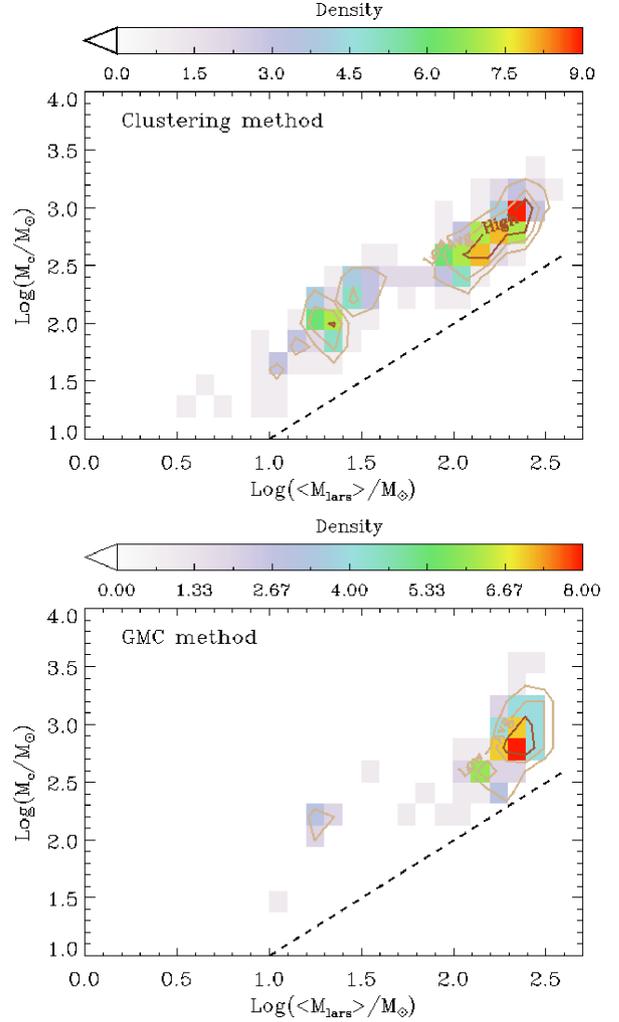


Figure 13. Plot of M_c versus the group-averaged Larson mass obtained with the clustering (*top panel*) and GMC_M2017 (*bottom panel*) methods. The lines represent the $M_c = \langle M_{\text{lars}} \rangle$ locus. Symbols are same as in Fig. 2.

the selected groups contain mostly clumps that are gravitationally bound.

6.3 Effects of the average clump temperature

It is not clear, from the different models of early star formation, whether and how the temperature of either the hosting GMC or the temperature of the individual clumps, should affect the shape and/or evolution of the CMF. It was therefore of interest to analyse whether any correlation of both power law and lognormal best-fitting parameters exist with the clump temperature. We found that when the lognormal fit is applied only to the *proto-stellar* sources within each group (selected with the clustering method), then M_c shows a good correlation with the group average temperature (Spearman’s rank correlation $\rho = 0.6$ with $N_{\text{min}} = 20$; see the top panel of Fig. C2), which is estimated by averaging the temperature of all the clumps in a given group.

However, as with the case of the Jeans mass discussed in Section 6.2, when the GMC_M2017 method is used (see the bottom panel of Fig. C2), we find no significant correlation between the parameters. This is not totally surprising, since $M_J \sim T^{3/2}$, and thus the lack of correlation of M_c with temperature for the GMC_M2017

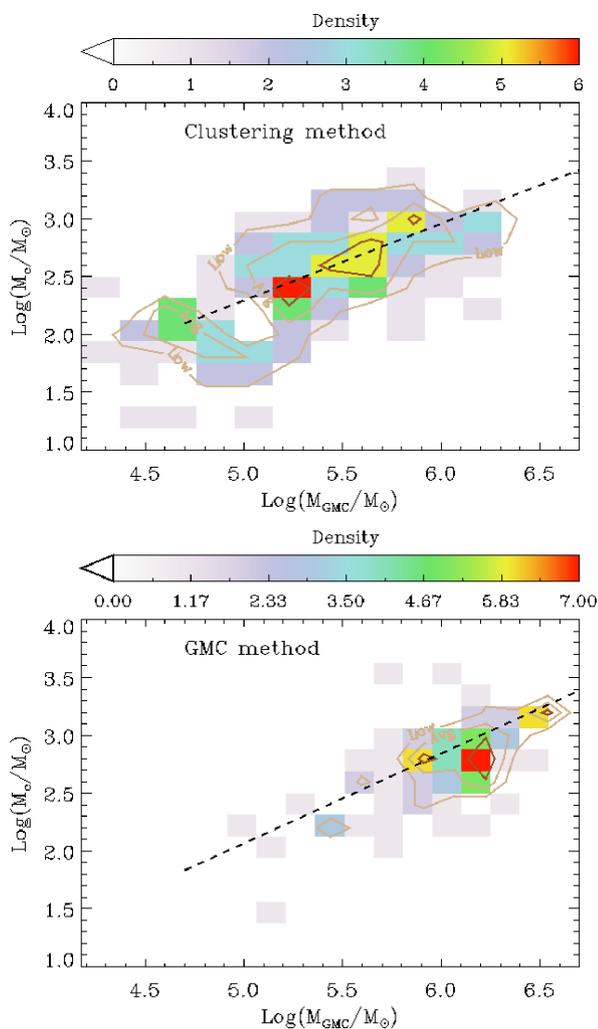


Figure 14. *Top panel:* plot of M_c as a function of the total mass of the associated GMC, using the clustering method. *Bottom panel:* same as above but with the GMC_M2017 method ($N_{\min} = 20$). The dashed line represents a linear fit to the data.

method would also likely affect the relationship with M_J . However, it is not clear what kind of selection effect, if any, is causing this discrepancy between the clustering and GMC methods. One possible explanation relies on the different composition of groups (as selected through the two methods) in terms of pre- and proto-stellar clumps.

6.4 Association with parameters of GMCs

In this section, we analyse the trend of the power law and lognormal best-fitting parameters with some of the physical properties of the associated GMC to each group of clumps. We begin discussing Fig. 14 that shows a trend of increasing M_c with larger masses, M_{GMC} , of the associated GMC (as listed in the catalogues by Miville-Deschênes et al. 2017 and Rice et al. 2016). The top panel shows the result obtained using the clustering method, characterized by a significant scatter. However, the bottom panel shows the same plot using the M_c and M_{GMC} obtained with the GMC method, and we can note that despite the lower number of data points, the scatter is much reduced. This result indicates that the mass of the ‘characteristic’ clump being formed within each GMC tends to increase if the total mass of the host molecular cloud is higher. Hence, if indeed high-

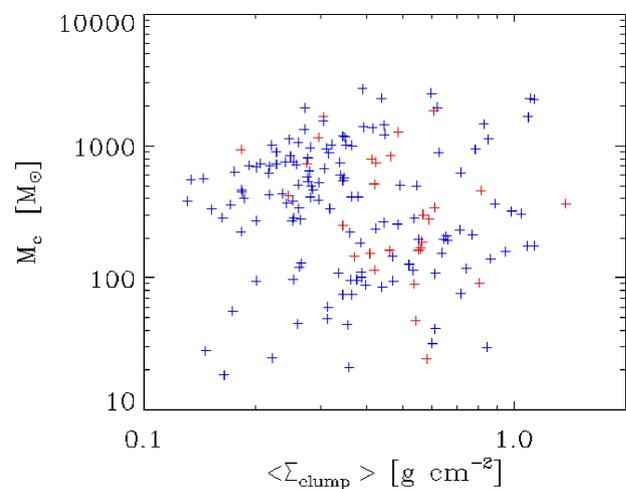


Figure 15. Plot of M_c as a function of the surface density averaged for all clumps in a given group (clustering method). Symbols and error bars are same as in Fig. 8.

mass stars are formed more likely within more massive clumps (McKee & Ostriker 2007), then this result suggests that massive stars are also more likely to form in massive GMCs. We find no significant correlation of the parameter σ with the mass of the associated GMC. Likewise, for the case of the power-law functional form, the best-fitting slope, α , does not appear to correlate with M_{GMC} .

Given the correlation between M_c and M_{GMC} it would be of interest to determine what is actually driving this relation. We have therefore analysed our data for a possible correlation between M_c and the average surface mass density of all the clumps in a given group, $\langle \Sigma_{\text{clump}} \rangle$, shown in Fig. 15. However, we find no correlation between the two parameters, suggesting that the observed trend of increasing M_c with larger M_{GMC} has no relation with the local density.

In terms of kinematic parameters, Fig. 16 shows that, despite a large scatter, a trend is observed between M_c and the CO rms velocity of the host GMC, as reported by Miville-Deschênes et al. (2017). The correlation is clearly observed using both the clustering and GMC_M2017 methods.⁴ Figs 14 and 16 suggest that, on average, more massive clumps tend to form in more massive and/or more turbulent GMCs. This is consistent with larger and more massive GMC inheriting higher levels of internal turbulence (i.e. larger rms velocities), hence causing fragmentation into a range of masses shifted towards a higher characteristic mass (see e.g. McKee & Ostriker 2007).

6.5 Association of groups with star-forming regions

In this section, we analyse the correlation between selected catalogues containing tracers of star formation with our selected groups. Specifically, we have chosen the Lockman catalogue of radio recombination lines (Lockman 1989), the CORNISH radio continuum catalogue (Purcell et al. 2013), and the Avedisova catalogue of star-forming regions (Avedisova 2000). The Lockman catalogue is a low-angular resolution (≈ 3 arcsec) survey containing 462 H II regions, mostly concentrated in the first quadrant. The CORNISH radio continuum project also covers part of the first quadrant and

⁴The fit is performed on the data points and not on the density contours.

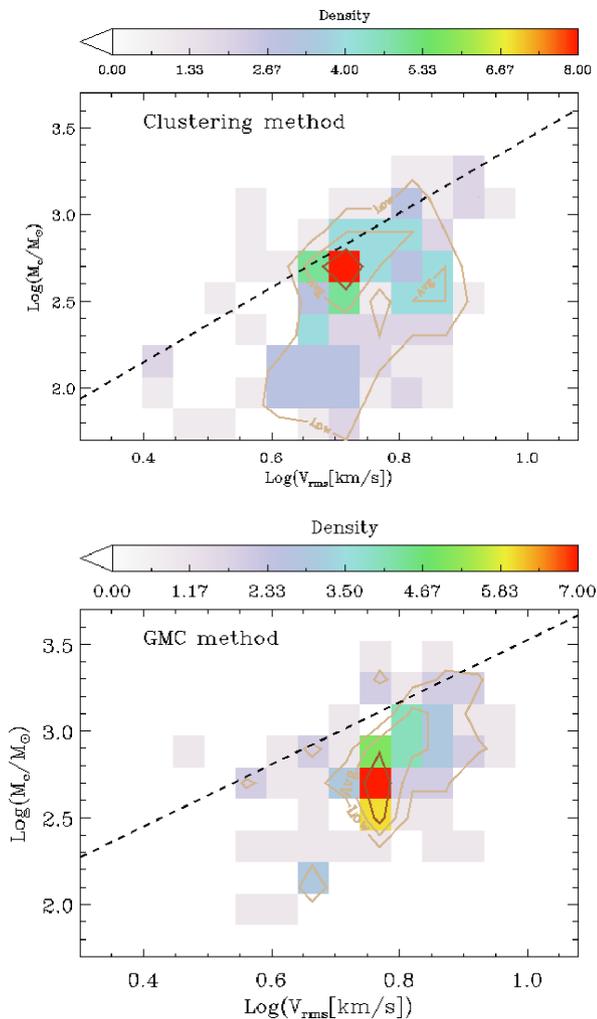


Figure 16. *Top panel:* plot of M_c as a function of the CO rms velocity, using the clustering method. *Bottom panel:* same as above but with the GMC_M2017 method ($N_{\min} = 20$). The dashed line represents a linear fit to the data.

is instead a high-angular resolution (~ 1.5 arcsec), sensitive (rms noise level < 0.4 mJy beam $^{-1}$) survey that contains 3062 compact radio sources. The Avedisova catalogue contains various types of sources that are considered tracers of star formation, and it covers all Galactic plane.

Fig. 17 shows, for each of these three catalogues, the number of sources associated with each group of clumps, as a function of M_c . We note two main features in this plot: first, most of the groups, with associated tracers of star formation, are concentrated in the mass range $M_c \gtrsim 100 M_\odot$. Secondly, the number of sources per group decreases at larger M_c values. We think that the latter is likely not a real effect, but a consequence of $M_{\text{clump}} \propto d^2 \Omega$, where d is the distance to the clump and Ω is the solid angle subtended by the source. Therefore, higher clump masses (and thus higher M_c) are biased towards larger distances where fewer associated star-forming tracers are likely to have been detected.

However, the fact that most groups with signs of star formation are found at $M_c \gtrsim 100 M_\odot$ is a real and important effect, since the tracers of star formation in the CORNISH and Lockman catalogues are tracers of high-mass star formation (HMSF). Therefore, this would suggest that HMSF is more likely to happen in groups where the range of clump mass is shifted towards the high-mass end. Al-

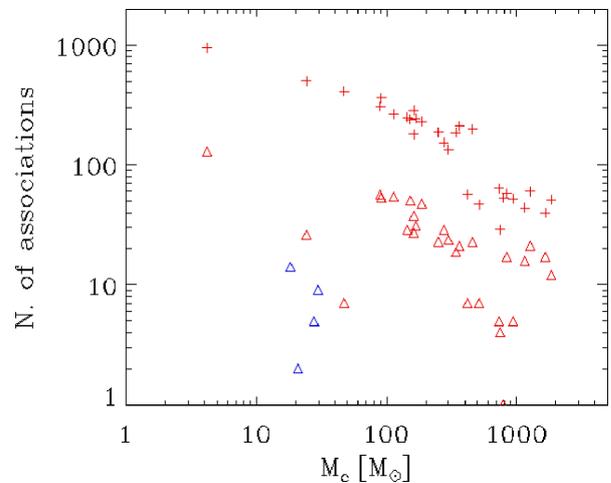


Figure 17. Plot of the number of sources, from the CORNISH ('+' signs), Lockman (open triangles) and Avedisova (open squares) catalogues, associated with each group of clumps as a function of M_c . Colour codes are same as in Fig. 8.

though this result does not directly support the turbulent-core model for HMSF (McKee & Tan 2002), where high-mass cores would collapse largely monolithically rather than fragmenting to small stars, over alternative models, such as for example the competitive accretion model (Bonnell et al. 1997), it contributes to individuate the most favourable set of initial conditions for HMSF to happen (see e.g. the discussion in Schilke 2017).

7 GENERATING MASS FUNCTIONS FROM FRAGMENTATION

Gravo-turbulent fragmentation plays a key role in determining the physical conditions that allow the CMF to evolve towards the IMF. Our observed CMFs not only show intrinsically different mass ranges but are also likely to span a wide range of evolutionary conditions and, more specifically, may represent and/or include different collapse and fragmentation histories. However, in Section 5.3 we found that the measured CMFs appear to have similar shapes, despite all these likely differences. This result supports one of the conclusions by Guszejnov & Hopkins (2015) that turbulent fragmentation occurs in a top-down cascade as large clouds fragment into clouds of smaller sizes, which then undergo fragmentation again, such that this cascade is scale free. In particular, the simulations by Guszejnov & Hopkins (2015) show that fragmentation at the high-mass end is close to scale free, i.e. the slope of the IMF is always a power law, which is systematically steeper than the CMF by a small, approximately fixed amount.

In Section B, we discuss a simple method to analyse whether our measured distribution of the power-law fit parameter α is consistent with this idea of a scale-free fragmentation cascade. This method uses some of the basic ideas discussed by Larson (1973). Evidently, these simple numerical simulations constitute a crude oversimplification of the real fragmentation process, but they lend support to the idea that not *any* kind of fragmentation process can reproduce the observed distribution of the power-law spectral indices of the high-mass end of the CMF. The consistent results obtained using two different methods for the selection of the groups provides additional support to this simple test.

8 CONCLUSIONS

In this paper, we have discussed two procedures to construct ‘coherent’ cluster of clumps from submillimetre continuum observations of Galactic star-forming regions, and then we have examined the functional form of the CMF derived from these clusters, or groups, of clumps. This work has taken advantage of the Hi-GAL source catalogue based on image photometry, which was compiled in each of the observed Hi-GAL fields, identifying the sources detected in the five different bands based on simple positional association (Elia et al. 2017). We have searched sources in the Hi-GAL catalogue to construct groups of objects that are coherent in space and velocity. These groups have been generated using both unbiased and biased methods. In the former case, the grouping of the sources does not follow a pre-determined scheme set, whereas in the latter we use the location of known GMCs as ‘seeds’ for building the clusters of sources. Most of the resulting groups are composed by $\lesssim 10$ –20 clumps; however, for our analysis, we have selected only groups with ≥ 30 sources.

We have then fitted the mass distribution of each individual group, and since the CMF is unlikely to be a pure single power law we have also used the lognormal functional form. We have demonstrated that even in the case of a strict single power-law mass function, great care must be taken in choosing methods for fitting parametrized models and that this may significantly affect the measurement of the exponent of the power law. We have shown that much of these problems derive from the uncertainty in estimating the lower bound to the power-law behaviour, M_{inf} , i.e. the limit below which the distribution does not follow the behaviour specified by its PDF. The average value of the best-fitting power-law exponent of the CMF is found to be $\langle \alpha \rangle = 2.12 \pm 0.15$, consistent with the most recent numerical simulations (e.g. Guszejnov & Hopkins 2015) and close, within the errors, to the Salpeter value.

The shape of the CMF does not appear to be a strong function of the intrinsic mass scale. In fact, while the characteristic mass, M_c , of the selected groups, as derived from the lognormal fits, spans more than three orders of magnitude, the *width* of the lognormal distribution varies collectively by less than a factor of 3. We compared the mass function of each group, selected with the clustering method, to the CMF of all the other groups using the KS test, which confirmed the similarities among the shapes of the normalized CMFs. We also found that most groups with signposts of (high-mass) star formation are associated with clusters having $M_c \gtrsim 100 M_\odot$. Since M_c shows no evident correlation with the distance, this is unlikely to be an observational effect and it indicates that HMSF is more likely to happen in groups hosting more massive clumps. Furthermore, the correlation of M_c with the mass, M_{GMC} , of the associated GMC suggest that more massive clumps tend to form in more massive GMCs.

ACKNOWLEDGEMENTS

LO thanks J. Rodriguez Santiago at the University of Puerto Rico for helping with the initial data analysis. E.S.’s research activity is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 607380 (VIALACTEA: the Milky Way as a star formation engine). We also wish to thank M.-A. Miville-Deschênes for providing us directly with the GMC data.

REFERENCES

- Alves J., Lombardi M., Lada C. J., 2008, *A&A*, 462, L17–L21
 Anathpindika S., 2013, *New A*, 18, 6
 Avedisova V., 2000, *Baltic Astron.*, 9, 569
 Bastian N., Covey K. R., Meyer M. R., 2010, *ARA&A*, 48, 339
 Bonnell I. A., Bate M. R., Clarke C. J., Pringle J. E., 1997, *MNRAS*, 285, 201
 Chabrier G., 2003, *PASP*, 115, 763
 Clauset A., Shalizi C. R., Newman M. E. J., 2009, *SIAM Rev.*, 51, 661
 Dame T. M., Hartmann D., Thaddeus P., 2001, *ApJ*, 547, 792
 Dib S., Shadmehri M., Padoan P., Maheswar G., Ojha D. K., Khajenabi F., 2010, *MNRAS*, 405, 401
 Elia D. et al., 2010, *A&A*, 518, L97
 Elia D. et al., 2013, *ApJ*, 772, 45
 Elia D. et al., 2017, *MNRAS*, 471, 100
 Guszejnov D., Hopkins P. F., 2015, *MNRAS*, 450, 4137
 Hopkins P. F., 2013, *MNRAS*, 430, 1653
 Hou L. G., Han J. L., Shi W. B., 2009, *A&A*, 499, 473
 Larson R. B., 1973, *MNRAS*, 161, 133
 Lockman F. J., 1989, *ApJS*, 71, 469
 Maschberger T., Kroupa P., 2009, *MNRAS*, 395, 931
 McKee C. F., Ostriker E. C., 2007, *ARA&A*, 45, 565
 McKee C. F., Tan J. C., 2002, *Nature*, 416, 59
 Miville-Deschênes M.-A., Murray N., Lee E. J., 2017, *ApJ*, 834, 57
 Molinari S. et al., 2010, *PASP*, 122, 314
 Molinari S. et al., 2016, *A&A*, 591, A149
 Momany Y., Zaggia S., Gilmore G., Piotto G., Carraro G., Bedin L. R., de Angeli F., 2006, *A&A*, 451, 515
 Nutter D., Ward-Thompson D., 2006, *MNRAS*, 368, 1833
 Offner S. S. R., Clark P. C., Hennebelle P., Bastian N., Bate M. R., Hopkins P. F., Moraux E., Whitworth A. P., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, *Protostars and Planets VI*. Univ. Arizona Press, Tucson, AZ p. 53
 Olmi L. et al., 2013, *A&A*, 551, A111
 Olmi L. et al., 2014, *A&A*, 564, A87
 Padoan P., Nordlund Å., 2002, *ApJ*, 576, 870
 Pilbratt G. L. et al., 2010, *A&A*, 518, L1
 Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2002, *Numerical recipes in C++ : The Art of Scientific Computing*, Cambridge University Press, New York
 Purcell C. R. et al., 2013, *ApJS*, 205, 1
 Reid M. A., Wilson C. D., 2006, *ApJ*, 650, 970
 Rice T. S., Goodman A. A., Bergin E. A., Beaumont C., Dame T. M., 2016, *ApJ*, 822, 52
 Rosolowsky E., 2005, *PASP*, 117, 1403
 Sadavoy S. I. et al., 2010, *ApJ*, 710, 1247
 Salpeter E. E., 1955, *ApJ*, 121, 161
 Schilke P., 2017, preprint ([arXiv:1712.05281](https://arxiv.org/abs/1712.05281))
 Scrucca L., Fop M., Murphy T. B., Raftery A. E., 2016, *R Journal*, 8, 289
 Swift J. J., Beaumont C. N., 2010, *PASP*, 122, 224
 Veneziani M. et al., 2017, *A&A*, 599, A7
 Xu Y. et al., 2016, *Sci. Adv.*, 2, e1600878

APPENDIX A: COMPARISON BETWEEN METHODS TO FIT THE POWER-LAW DISTRIBUTION

In this appendix, we discuss and compare several methods to implement the power-law fits, as described in Section 4.3. Clauset et al. (2009) discuss both the discrete and continuous MLEs for fitting the power-law distribution to data. Maschberger & Kroupa (2009) analyse a variety of methods to estimate the parameters, exponent, and lower cut-off of a power-law distribution. Their MLE estimate

is given as

$$\hat{\alpha} = 1 + \frac{N_{\text{tot}}}{T - N_{\text{tot}} \ln Y} \quad (\text{A1})$$

with $Y = \min\{x_i\} = x_{\text{min}}$, or the given lower limit (it must not be confused with x_{inf} ; $x_{\text{min}} = \min\{x_i\} \leq x_{\text{inf}}$), and $T = \sum_{i=1}^{N_{\text{tot}}} \ln x_i$. Equation (A1) can be easily shown to be the same as equation (5) if $N_{\text{tot}} = N_{\text{inf}}$ and $Y = x_{\text{inf}}$. Both Clauset et al. (2009) and Maschberger & Kroupa (2009) then use the same unbiased estimator:

$$\hat{\alpha}' = 1 + (\hat{\alpha} - 1) \frac{N_{\text{tot}} - 1}{N_{\text{tot}}}. \quad (\text{A2})$$

However, the analysis of Maschberger & Kroupa (2009) is more focused on the estimates of α and no recipe for the objective determination of x_{inf} is given. In the following, we will refer to this method as ‘MK’.

Rosolowsky (2005) uses the ‘errors-in-variables’ regression, rather than the ordinary least squares regression where the independent variable is assumed to be measured without error. This method maximizes the likelihood that a set of data (x, y) , with associated uncertainties (σ_x, σ_y) , can be drawn from a power-law distribution. One of the main advantages of this approach is that it includes the errors in the mass determination of the clumps, which can be substantial. In addition, similar to PLFIT, the method used by Rosolowsky (2005) also allows to estimate the number of clumps more massive than the point where the distribution shows a significant deviation from a power law. We will refer to this procedure as ‘MSPECFIT⁵’, from the name of the main IDL function implementing the method.

We have run a series of tests to compare the performance of the MK, MSPECFIT, and PLFIT methods to fit a power-law functional form to the CMF, and the problems that we have found raise concern about the statistical analysis of similar sets of data. The first of such tests regards the simulation of fragmentation within our clumps. In Section 7, we have described a method to simulate the fragmentation of the observed (‘parent’) clumps into a number of ‘daughter’ clumps, according to the recipe of gravo-turbulent fragmentation. Fig. A1 shows some of the problems we have found by running these fragmentation tests. We plot both the originally measured values of α and those obtained after our simulation of gravo-turbulent fragmentation is applied, named α_{frag} . The top panel shows the results of the MK method when *all* clumps of a given group are considered for the estimate of α , i.e. including also those with $M < M_{\text{inf}}$ (all groups refer to the clustering method). A clear correlation is observed between α and α_{frag} for which we have no immediate physical explanation. We think it is unlikely this is a consequence of the algorithm used to simulate the fragmentation of the clumps (Section 7) since two different methods lead to similar correlations. However, it could be an artifact caused by the specific MLE method used to determine the values of the power-law indices. In fact, also MSPECFIT shows a similar correlation between α and α_{frag} , though with different ranges of values.

The bottom panel of Fig. A1 shows instead the same plot, also obtained with the MK method, but when only clumps with $M > M_{\text{inf}}$ are used. In this case, the value of M_{inf} has been determined through the PLFIT method, which only runs under this mode, and then it has been applied to both MK and MSPECFIT. Given the similarities between the MK and PLFIT methods mentioned in Section 4.2, it is not surprising that PLFIT also shows a similar distribution, where

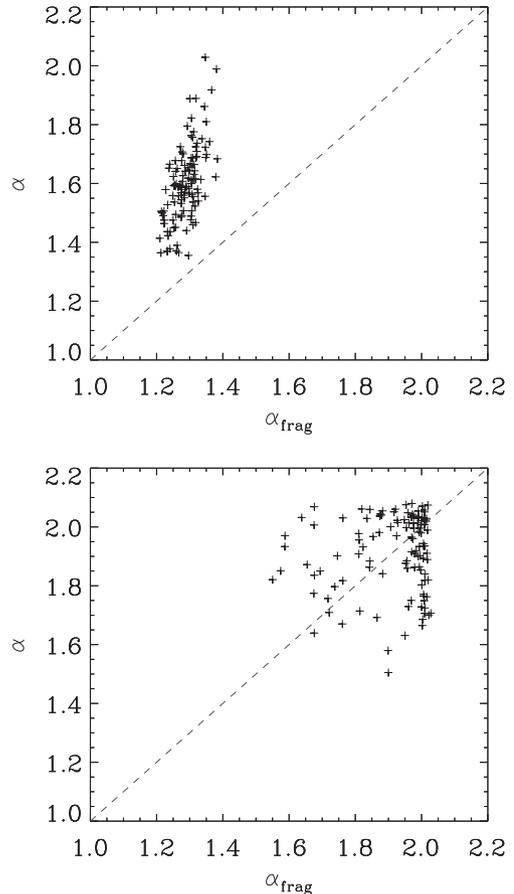


Figure A1. *Top panel.* Plot of the values of α for a given group versus the values of α_{frag} for the ‘fragmented’ groups (see the text), using *all* available clumps in each group and the MK method. The dashed line represents the identity line. *Bottom panel.* Same as above but using only clumps with $M > M_{\text{inf}}$ (all groups refer to the clustering method).

both α and α_{frag} show an upper limit $\simeq 2$. This raises the suspicion that equation (5), or (A1), may be responsible for this bias in the values of α . If one sets $\alpha \leq 2$ in equation (5) it is easy to show that this condition is equivalent to $\sum_{i=1}^{N_{\text{inf}}} \ln[x_i / (e x_{\text{inf}})] \geq 0$, which is always satisfied if $x_i \geq e x_{\text{inf}}$ for all values of i (‘ e ’ represents Euler’s number). The fact that there may be clumps with mass $M_{\text{inf}} \leq M \leq e M_{\text{inf}}$ and also that both the PLFIT and MK methods use the unbiased estimator $\hat{\alpha}'$ in equation (A2), may account for $\alpha \lesssim 2$ as observed in the bottom panel of Fig. A1.

Contrary to MK and PLFIT, the MSPECFIT method does not appear to have this problem, and both α and α_{frag} may take also values > 2 . As mentioned earlier MSPECFIT, like PLFIT, has built in the capability to estimate the value of the lower cut-off, M_{inf} . However, during our tests the values of α estimated by MSPECFIT with this option were systematically lower compared to those obtained with the other two methods. Therefore, we preferred to use MSPECFIT without the option to calculate M_{inf} . Instead, we first determined the value of M_{inf} by PLFIT and we then fed this parameter back to MSPECFIT. We refer to this hybrid method as MSPEC-PLFIT.

⁵<https://github.com/low-sky/idl-low-sky/>

APPENDIX B: SIMULATION OF FRAGMENTATION

In this section, we attempt to analyse whether our measured distribution of the power-law fit parameter α is consistent with the basic model of a scale-free fragmentation cascade. This is equivalent to assume the coexistence of multiple fragmentation stages in our observed CMFs. Then, if the present observing epoch does not represent a privileged time in the history of the continuing Galactic star formation (assumed to be stationary), the distribution of the parameters characterizing the CMFs should be time invariant. We can check this argument by simulating, for example, an additional step in the fragmentation of the observed clumps (corresponding to a later arbitrary time in the evolution of the cluster), in which case one would expect the ‘daughter’ distribution of the α parameter to be mostly unchanged compared to the ‘parent’ distribution.

We begin our analysis writing the power spectrum of turbulence that in the inertial range may be assumed to be a power law:

$$E(k) \propto k^{-x}, \quad (\text{B1})$$

where k is the wavenumber and x the spectral index, which can vary from cloud to cloud but it is normally found in the range $1 \lesssim x < 2.9$ (Padoan & Nordlund 2002). Following Anathpindika (2013), we assume that turbulence is driven over spatial scales between 10 and 200 pc, corresponding to wavenumbers between $k_{\min} = 1$ and $k_{\max} = 8$. Anathpindika (2013) also shows that the mass of the generic clump (which we call a ‘daughter’ clump), M_d , condensing out of the original (‘parent’) clump, M_p , can be written as

$$M_d(k) \propto k^{3(1-x)/2} \quad (\text{B2})$$

and the number of clumps, N_d , condensing out is M_p/M_d , so that the daughter clump-mass function is

$$\frac{dN_d}{dM_d} \propto k^{3(1-x)}. \quad (\text{B3})$$

Since we are not attempting to reproduce the temporal evolution of the turbulent fragmentation process, we will only induce and follow the fragmentation of the current population of clumps, while ignoring further fragmentation that might be originating at the same time in the parent GMC. Our simulation thus proceeds as follows:

(a) For the selected group, the first clump (parent) mass is read. Only parent masses $M_p \geq 10 M_J$ are subject to further fragmentation.

(b) A wavenumber, k , is randomly chosen in the interval $[k_{\min}, k_{\max}]$.

(c) The number of daughter clumps, N_d , is calculated using equation (B3).

(d) The mass of each M_d is calculated as $M_d = M_p k^{3(1-x)/2}$, where k is again randomly chosen. If $M_d \geq 5 M_J$, this step is repeated (i.e. the procedure does not accept too small fragmented masses). We also used $x = 1.7$, an intermediate value among those tested by Anathpindika (2013).

(e) The daughter clump masses are checked to satisfy the mass conservation condition $\sum M_d \simeq M_p$.

At the end of this procedure each group, i , with an original number of clumps N_{i0} is fragmented into $N_{i1} > N_{i0}$ less massive clumps. The procedure could be repeated n times so that $N_{in} > N_{i(n-1)} > \dots > N_{i2} > N_{i1} > N_{i0}$; however, we stopped the process of fragmentation after the first step. We then analysed the resulting CMF of each fragmented group i , having N_{i1} clumps, and applied the method MSPEC-PLFIT described in Section 4.3, yielding a power-law slope, α_{frag} , for each fragmented CMF. The final distribution

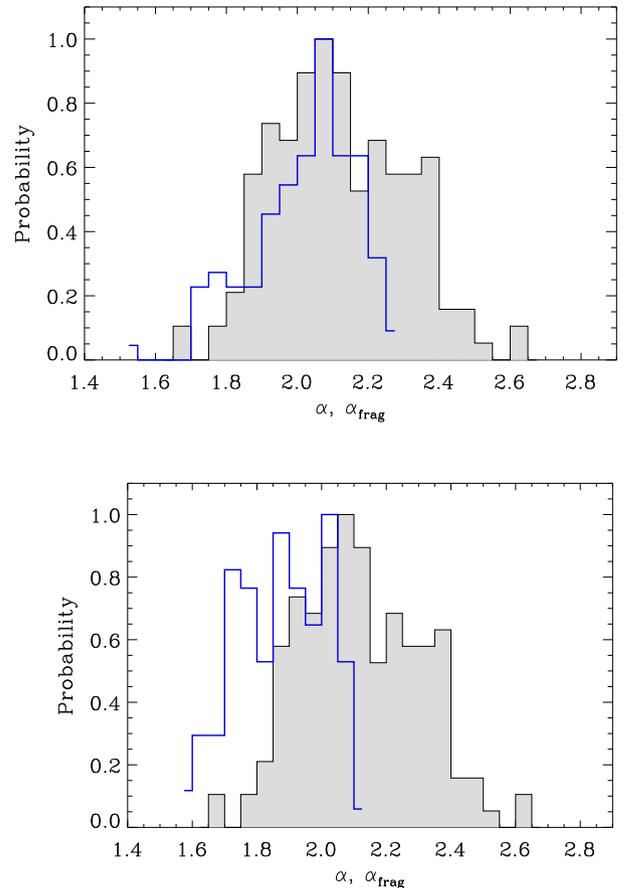


Figure B1. *Top panel:* histograms of the distribution of α (grey histogram) and α_{frag} (blue line) values, obtained with the MSPEC-PLFIT method, and applied to the groups obtained with the clustering method. The fragmented clumps were obtained using the Gravo-turbulent fragmentation model (see the text). *Bottom panel:* same as above but the fragmented clumps were obtained through a completely random process.

of the α_{frag} values is shown in the top panel of Fig. B1, when the clustering method is used for the selection of the groups, and in Fig. B2 for the GMC_M2017 method. We note that the histograms of the original α and α_{frag} values are quite consistent, as expected, though the distribution of α_{frag} is somewhat narrower. We also find that the average value of M_{inf} is shifted towards lower values after the fragmentation process.

In order to test whether this is actually a consequence of the turbulent fragmentation model, we run a different simulation where fragmentation is assumed to be a completely random process and independent from the power spectrum of turbulence. We thus repeated the steps described earlier, but replacing step (c) by choosing $N_d = r$, where r is a random number drawn uniformly from the (arbitrary) interval $[1, 8]$. The mass, M_d , of each fragment, i.e. step (d) above, was then determined by once again drawing a random number in the same arbitrary range $r' \in [1, 8]$ and then calculating M_d by scaling M_p according to $M_d = M_p (r'/8)$. The rest of the procedure was the same as discussed earlier. The distribution of the new values for α_{frag} is shown in the bottom panels of Figs B1 and B2. We can note that the main difference with respect to the histograms obtained applying the turbulent fragmentation is the offset with respect to the histogram of the original distribution of the α values. Furthermore, in the case of the clustering method the new distribution does not have a well-defined peak.

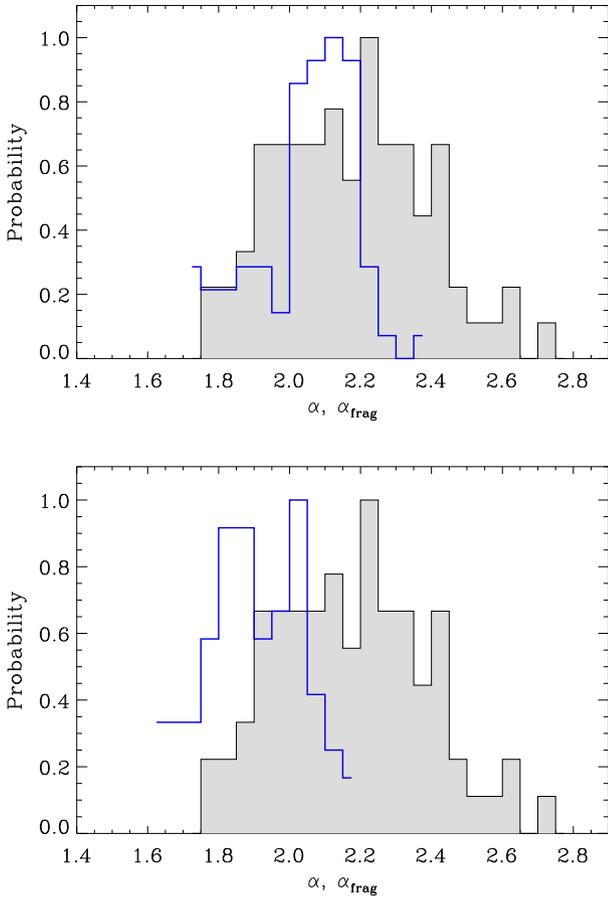


Figure B2. Same as Fig. B1 for the GMC_M2017 method.

APPENDIX C: ADDITIONAL FIGURES

See Figs C1 and C2.

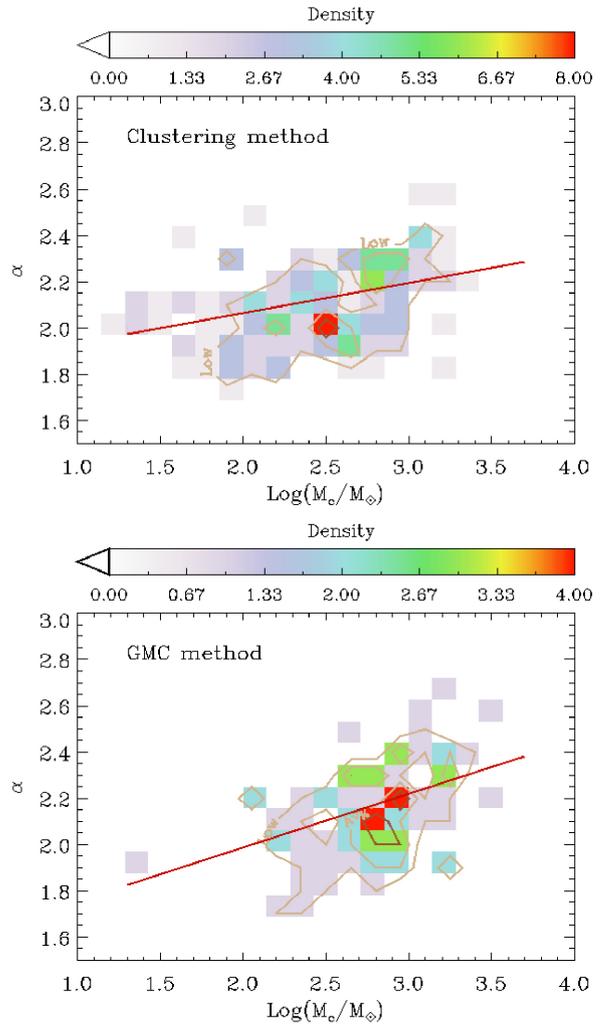


Figure C1. Density plots of the power-law slope α versus M_c obtained with the clustering (*top panel*) and GMC_M2017 (*bottom panel*) methods. The best-fitting lines have been superimposed (Spearman's rank correlation $\rho = 0.25$, top, and 0.31 , bottom).

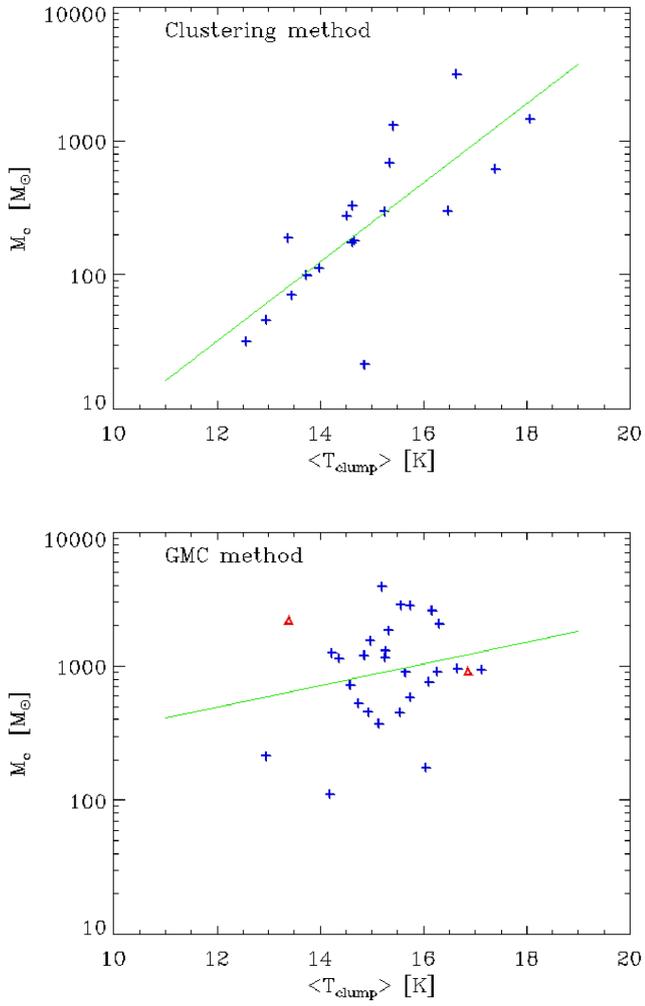


Figure C2. Plot of the best-fitting parameter M_c as a function of the average clump temperature in each group with *only* proto-stellar sources, as obtained with the clustering ($N_{\min} = 30$, *top panel*) and GMC_M2017 ($N_{\min} = 20$, *bottom panel*) methods ($N_{\min} = 20$). The lines represent linear fits to the data. Symbols are as in Fig. 8.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.