

# ALMAGAL

## VI. The spatial distribution of dense cores during the evolution of cluster-forming massive clumps

E. Schisano<sup>1,\*</sup>, S. Molinari<sup>1</sup>, A. Coletta<sup>1,2</sup>, D. Elia<sup>1</sup>, P. Schilke<sup>3</sup>, A. Traficante<sup>1</sup>, Á. Sánchez-Monge<sup>4,5</sup>, H. Beuther<sup>6</sup>, M. Benedettini<sup>1</sup>, C. Mininni<sup>1</sup>, R. S. Klessen<sup>7,8,9,10</sup>, J. D. Soler<sup>1</sup>, A. Nucara<sup>1,11</sup>, S. Pezzuto<sup>1</sup>, F. van der Tak<sup>12,13</sup>, P. Hennebelle<sup>14</sup>, M. T. Beltrán<sup>15</sup>, L. Moscadelli<sup>15</sup>, K. L. J. Rygl<sup>16</sup>, P. Sanhueza<sup>17</sup>, P. M. Koch<sup>18</sup>, D. C. Lis<sup>19</sup>, R. Kuiper<sup>20</sup>, G. A. Fuller<sup>3,21</sup>, A. Avison<sup>22</sup>, L. Bronfman<sup>23</sup>, U. Lebreuilly<sup>14</sup>, T. Möller<sup>3</sup>, T. Liu<sup>24</sup>, V.-M. Pelkonen<sup>1</sup>, L. Testi<sup>25</sup>, Q. Zhang<sup>9</sup>, T. Zhang<sup>26,3</sup>, A. Ahmadi<sup>27</sup>, J. Allande<sup>15,28</sup>, C. Battersby<sup>29</sup>, J. Wallace<sup>29</sup>, C. L. Brogan<sup>30</sup>, S. Clarke<sup>3,18</sup>, F. De Angelis<sup>1</sup>, F. Fontani<sup>15,31,32</sup>, P. T. P. Ho<sup>18,33</sup>, T. R. Hunter<sup>30</sup>, B. Jones<sup>3</sup>, K. G. Johnston<sup>34</sup>, P. D. Klaassen<sup>35</sup>, S. J. Liu<sup>1</sup>, S.-Y. Liu<sup>18</sup>, Y. Maruccia<sup>36</sup>, A. J. Rigby<sup>37</sup>, Y.-N. Su<sup>18</sup>, Y.-W. Tang<sup>18</sup>, S. Walch<sup>3,38</sup>, and H. Zinnecker<sup>39</sup>

(Affiliations can be found after the references)

Received 21 May 2025 / Accepted 25 November 2025

### ABSTRACT

**Context.** High-mass stars and star clusters form from the fragmentation of massive dense clumps driven by gravity, turbulence, and magnetic fields. The extent to which each of these agents impacts the fragmentation depending on the clump mass, density, and evolutionary stage is still largely unknown.

**Aims.** The ALMA evolutionary study of high-mass protocluster formation in the GALaxy (ALMAGAL) project, with ~1000 clumps observed at ~1000 au resolution, allows a statistically significant characterization of the fragmentation process over a large range of clump physical parameters and evolutionary stages. Our goal is to characterize where and how the dense cores revealed by ALMA are distributed in massive potentially cluster-forming clumps to trace how fragmentation is initially set and how it proceeds before gas dispersal due to stellar feedback.

**Methods.** We characterized the spatial distribution of dense cores in the 514 ALMAGAL clumps that host at least four cores, using a set of quantitative descriptors that we evaluated against the clump bolometric luminosity-to-mass ratio, which we adopted as an indicator of the evolution of the system. We measured the separations between cores with the minimum spanning tree (MST) method, which we compared with the predictions of gravitational fragmentation from Jeans theory. We investigated whether cores have specific arrangements using the  $Q$  parameter or variations due to their masses with the mass segregation ratio,  $\Lambda_{MSR}$ .

**Results.** ALMAGAL cores are distributed throughout the entire area of the clump, usually arranged in elliptical groups with an axis ratio  $e \sim 2.2$ , although high values with  $e \geq 5$  are also observed. We found a single characteristic core separation per clump in ~76% of cases, suggesting that multiple fragmentation lengths may be frequently present. Typical core separations are compatible with the clump-averaged thermal Jeans length,  $\lambda_J^h$ . However, we found an additional population of cores, typical of low-fragmented and young clumps, which are on average more widely separated with  $l \approx 3 \times \lambda_J^h$ . By stacking the distributions of the core separations in clumps of similar evolutionary stage, we also found that the separation decreases on average from  $l \sim 22\,000$  au in younger systems to  $l \sim 7\,000$  au in more evolved ones. The ALMAGAL cores are typically distributed in fractal-type subclusters, while centrally concentrated patterns appear only at later stages, but we do not observe a progressive transition between these configurations with evolution. Finally, we also found 110 ALMAGAL systems with a signature of mass segregation, with an occurrence that increases with evolution.

**Key words.** stars: formation – stars: massive – stars: protostars – ISM: clouds – submillimeter: ISM

### 1. Introduction

At least half of all stars, including our Sun and almost all massive stars, formed in clustered systems of hundreds to thousands of objects (Lada & Lada 2003; Zinnecker & Yorke 2007; Bressert et al. 2010; Adams 2010; Megeath et al. 2016, 2022; Adamo et al. 2020). The prevalence of clustered star formation is a direct consequence of the hierarchical nature of the interstellar medium (ISM) from which stars form (Elmegreen 2010). The ISM shows substructures on all spatial scales (Scalo 1985; Elmegreen & Falgarone 1996; Williams et al. 2000; Molinari et al. 2010);

molecular clouds ( $\geq 1$  pc), filaments ( $\sim 0.1$ – $1$  pc), clumps ( $\sim 0.3$ – $1$  pc), and cores ( $\lesssim 0.1$  pc). The fragmentation of these structures leads to the observed distribution of young stars, which are typically arranged in groups or clumpy subclusters (Kuhn et al. 2014; Zhou et al. 2024b). Only a fraction of these young systems ultimately remain as stellar bounded clusters after they merge into smoother distributions (Kruijssen 2012; Grudić et al. 2018; Li et al. 2019; Krumholz & McKee 2020; Krause et al. 2020; Zhou et al. 2025). Star clusters form in dense clumps that are typically located at the intersection of multiple filaments (Myers 2009; Schneider et al. 2012; Motte et al. 2018; Zhou et al. 2022). These dense clusters contain one or more high-mass stars ( $M \geq 8 M_{\odot}$ ),

\* Corresponding author: eugenio.schisano@inaf.it

whose stellar winds, ionization, heating, radiation pressure, and ultimately supernova explosions play a fundamental role in shaping the evolution of the interstellar medium, from planetary systems to the entire galactic ecosystem (Portegies Zwart et al. 2010).

The formation of dense clusters is a complex theoretical problem. It is governed by the interplay of gravitational collapse, turbulence, and magnetic fields in an environment that is continuously altered by radiative and mechanical feedback from the embedded high-mass stars still in formation (Mac Low & Klessen 2004; Dale & Bonnell 2011; Krumholz et al. 2014; Krumholz & McKee 2020; Li et al. 2014; Girichidis et al. 2020; Padoan et al. 2020; Grudić et al. 2021; Guszejnov et al. 2021; Vázquez-Semadeni et al. 2019, 2025). In addition to these processes, the final structure and properties of clusters are influenced by other factors. These include the initial cloud conditions and morphology (Girichidis et al. 2011), which may also be related to the local galactic environment (Smith et al. 2020); the possible coalescence of protostellar cores in the early evolutionary phases (Dib 2023); and the hierarchical merging where distinct groups of pre-existing star-forming cores or young subclusters assemble (Maschberger et al. 2010).

Current models of high-mass stars and cluster formation are divided into categories that mostly differ by the dynamical state of the parent structure that fragments and where the material that ends up in the protostars comes from. One category is composed of static models that assume that high-mass star formation starts in substructures created by the supersonic turbulence present in molecular clouds (McKee & Tan 2003), with possible contributions from the magnetic field. The material is initially gathered in these substructures and builds up the entire mass reservoir for the star formation process. In this category, there are different scenarios distinguished by the structure type that is initially assembled: cores for the core-fed models, for example the turbulent model of McKee & Tan (2003), or the entire clump for the in situ formation scenario introduced by Longmore et al. (2014). In contrast, another category posits a more dynamical scenario where star formation is coeval to a further accumulation of gas in the high-density structures: cores and clumps. This additional contribution increases the reservoir of material available for accretion onto young protostars. Longmore et al. (2014) refer to this category as conveyor belt mode to recall that the mass of star-forming material is never assembled at a single time (see also the discussion in Krumholz & McKee 2020), but is continuously transferred. The competitive accretion model introduced by Bonnell et al. (2001) falls into this category since it prescribes that cores gather their masses from the clump local environment, with a mass growth that is more efficient for objects close to the bottom of the gravitational well (Bonnell & Bate 2006). While many models agree that flows funnel material from low to high densities to assemble massive stars, they differ on what drives the local accretion flow. For example, it is driven by gravity in the global hierarchical collapse (GHC) model (Ballesteros-Paredes et al. 2018; Vázquez-Semadeni et al. 2019, 2025), while by the large-scale random (turbulent) velocity field in the inertial flow scenario of Padoan et al. (2020).

Over the years, several studies have been dedicated to characterizing the initial conditions of star formation by observing young stellar clusters (Gutermuth et al. 2009; Kuhn et al. 2014; Parker & Alves de Oliveira 2017; Parker 2018; Dib et al. 2018). Although they are considerably young, these systems are already influenced by their dynamical evolution, which does not allow one to trace back where and under which initial conditions the stars originally formed. This leaves severe uncertainties about

how the time-dependent process of fragmentation started and how it proceeds during the early evolution of the system, before most of the gas is dispersed, questions that can be answered only by observing the distribution of star-forming cores.

Modern millimeter interferometers, such as the Atacama Large Millimeter/submillimeter Array (ALMA), the Northern Extended Millimeter Array (NOEMA), and the Submillimeter Array (SMA) have achieved the angular resolution and sensitivity required to unveil the fragmentation properties in the dense clumps, which are the crowded and heavily extincted sites where high-mass stars and clusters form, typically located a few kiloparsec from the Sun (Zhang et al. 2014, 2015; Csengeri et al. 2017; Sanhueza et al. 2017; Beuther et al. 2018; Fontani et al. 2018; Svoboda et al. 2019; Sadaghiani et al. 2020; Moscadelli et al. 2021; Palau et al. 2021; Liu et al. 2020, 2022; Morii et al. 2021, 2023; Traficante et al. 2023; Avison et al. 2023; Ishihara et al. 2024; Liu et al. 2024; Zhou et al. 2024a). In recent years, an important observational effort has been carried out to study these massive clumps, identifying and characterizing the embedded population of star-forming fragments in order to provide important observational constraints to the models of high-mass stars and cluster formation. Several surveys focus on clumps in specific evolutionary phases, either very young and quiescent systems embedded in infrared dark clouds (Zhang et al. 2014; Sanhueza et al. 2017; Svoboda et al. 2019; Sanhueza et al. 2019; Anderson et al. 2021; Morii et al. 2023; Rigby et al. 2024) or systems in more advanced stages that are associated with infrared bright emission (Beuther et al. 2018; Liu et al. 2020, 2024; Xu et al. 2024). Other surveys selected their targets to cover a wide range of evolutionary phases (Avison et al. 2023; Traficante et al. 2023; Ishihara et al. 2024), but observed only a limited number of systems. The ALMA-IMF Large Program (Motte et al. 2022) surveyed 15 massive Galactic protoclusters near the Sun ( $d < 5.5$  kpc) for a detailed analysis of the core mass function in different environments and evolutionary stages. These observations have built up a rich, but rather heterogeneous dataset of observed properties of the fragments hosted in about two hundred massive clumps.

Although observed core masses are an important reference point for star formation models, the spatial distribution of cores offers an additional and robust diagnostic with critical information on how clusters assemble and the relevance of various physical mechanisms in their formation process. Fragmentation leaves a different signature imprinted on the separations of cores in a turbulent or gravity-dominated medium, which is assumed in different categories of models, characterized by different values for the Jeans length (Mac Low & Klessen 2004; Vázquez-Semadeni et al. 2025). Several observations have reported separations that are close to or smaller than the average thermal Jeans length of the hosting structure (clump or cloud; Palau et al. 2015; Kainulainen et al. 2017; Beuther et al. 2018; Sanhueza et al. 2019; Morii et al. 2024; Ishihara et al. 2024; Beuther et al. 2025). However, turbulent fragmentation is not completely ruled out as an additional contribution from turbulent support is required to explain observations in other systems (Rebolledo et al. 2020; Avison et al. 2023) and in nearby clouds (Ishihara et al. 2025). In general, distinct fragmentation patterns with strong core alignments, clustered arrangements with cores randomly distributed, or low fragmentation and rather symmetrical spatial distributions are observed, whether the dominant energetic contribution in the system is the magnetic field, turbulence, or gravity, respectively (Tang et al. 2019). Anisotropic accretion flows, whose presence may be traced by small-scale subfilaments, provide additional fragmentation, increasing the number of cores and

their location in the system. In addition, dynamical effects, such as the merging of subclusters (Bonnell et al. 2003), clump global gravitational collapse (Vázquez-Semadeni et al. 2017), and tidal interactions between the cores (Zhou et al. 2024a; Li 2024), modify the overall structure and spatial distribution of young systems. Consequently, variations in these structural properties across different evolutionary stages can be used to determine the presence and impact of such dynamics. Observations suggest that the typical separation between cores is shorter in more evolved systems (Beuther et al. 2021; Traficante et al. 2023; Ishihara et al. 2024). Moreover, the core distribution is also found to be more compact in the later stages (Xu et al. 2024) and in regions with higher star formation rates (Dib & Henning 2019). These two pieces of evidence both support the presence of a strong gravitational pull, a signature that a global collapse, as predicted by the GHC theory (Vázquez-Semadeni et al. 2025), is active in massive clumps.

Additional information is provided by looking at the location of the massive members of the clustered system. Although young stellar clusters often show the presence of mass segregation (MS), with high-mass objects that are most likely to lie close to the cluster center (Allison et al. 2009; Parker & Alves de Oliveira 2017; Parker 2018; Dib & Henning 2019; Jia et al. 2025), it is unclear whether those patterns are connected to their initial formation or are due to their subsequent dynamical evolution (see also Portegies Zwart et al. 2010). A similar statistical analysis carried out on the dense core population shows that MS may be present since the early stages (Sanhueza et al. 2019; Sadaghiani et al. 2020; Nony et al. 2021; Morii et al. 2024; Xu et al. 2024), which is indicative that competitive accretion may be a dominant path of their formation (Bonnell & Bate 2006).

Although several indications were drawn from the available dataset, the diverse selection criteria of the targets, the different spatial resolutions and sensitivity levels achieved by the various surveys, and the low number of observed clumps limit the statistical analysis and direct comparison of the observed fragmentation population. The ALMA Cycle 7 large program ALMA evolutionary study of high-mass protocluster formation in the GALaxy (ALMAGAL) (2019.1.00195L, PI: S. Molinari, P. Schilke, C. Battersby, and P. Ho) was designed to mitigate these problems by observing an unprecedented large sample of 1013 clumps covering a wide range of physical conditions and Galactic locations (Molinari et al. 2025). The ALMAGAL program provides an extended and homogeneous dataset to study the fragmentation process in dense and massive clumps. In this paper we analyzed the spatial distribution of the cores identified in the ALMAGAL fields by Coletta et al. (2025) to provide a robust characterization of where the cores are located in massive and potentially cluster-forming clumps, what their typical separation is, how they are distributed, and how these properties vary across the early evolutionary phases. This work is a complement to the comparison between the physical properties of the ALMAGAL core population, specifically the level of fragmentation, core masses, and efficiency, and the large-scale clump parameters presented by Elia et al. (2026).

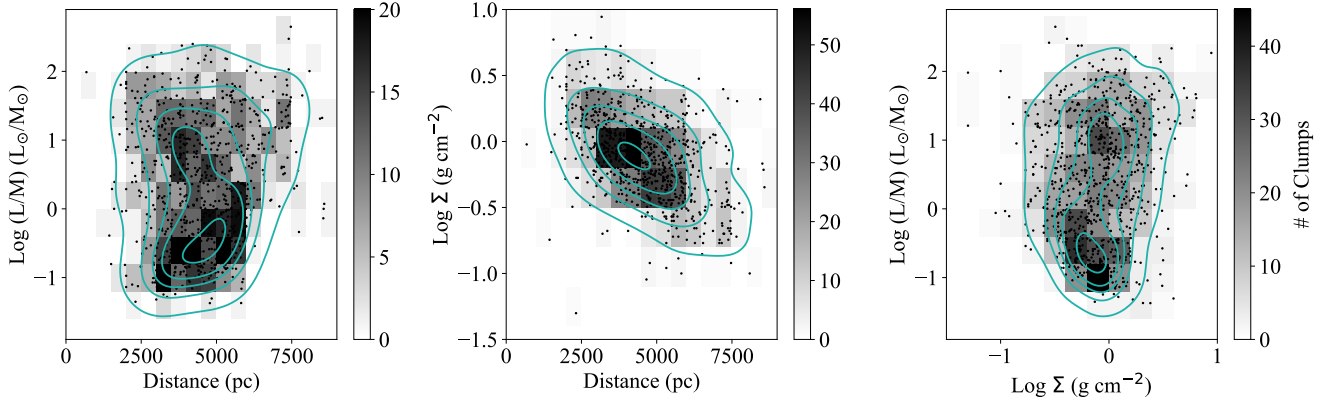
The structure of this paper is as follows. We discuss the details of the ALMAGAL program in Sect. 2, summarizing the observational strategy for the survey and the properties of the ALMAGAL clumps and the extracted core population used in the analysis. We present the results of the analysis of the spatial distribution of ALMAGAL cores in Sect. 3. Specifically, in Sect. 3.1 we describe the general morphological properties of the groups of cores, defining where they are distributed. In Sect. 3.2, we determine the separations between the cores and compare

them with the clump averaged properties in Sect. 3.3, focusing on how the separations vary during different evolutionary phases. We discuss how they relate to the expectation of the Jeans fragmentation theory in Sect. 3.4. In Sect. 3.5, we present additional statistical descriptors to illustrate how cores are distributed and whether mass segregation is present in systems, checking if they show a variation across the different evolutionary phases. Finally, we discuss our results within the framework of the generally adopted models in Sect. 4.

## 2. Data

### 2.1. The ALMAGAL clump sample

The ALMAGAL large program (2019.1.00195L, PI: Molinari, Schilke, Battersby, Ho) observed with ALMA in Band 6 ( $\sim 220$  GHz,  $\sim 1.4$  mm) 1013 dense massive clumps located close to the Galactic Plane, and selected from the Hi-GAL (Molinari et al. 2010, 2016b; Elia et al. 2017, 2021) and RMS surveys (Lumsden et al. 2013), to investigate their fragmentation properties. The criteria adopted for the target selection are described in Molinari et al. (2025) and are based on thresholds on the *Herschel* continuum emission, the clump mass,  $M$ , and the average surface density,  $\Sigma_{cl}$ , determined from the *Herschel* Hi-GAL data (Molinari et al. 2016b; Elia et al. 2017, 2021). An initial analysis of the ALMA ACA data provided new estimates for the clump heliocentric distances, derived from the radial velocities measured from the spectral lines of multiple molecular tracers (Benedettini et al., in prep.). The ALMAGAL clumps, with physical properties updated thanks to these new measurements, have masses  $M$  between  $\sim 100$  and  $12\,000 M_{\odot}$ , bolometric luminosity,  $L$ , between  $\sim 50$  and  $\sim 5 \times 10^5 L_{\odot}$ , and are located in the Galaxy at heliocentric distances,  $d_{cl}$ , from  $\sim 1$  up to 14 kpc, with only one clump lying closer than 1 kpc (Molinari et al. 2025; Elia et al. 2026). Notwithstanding these updates, the observed ALMAGAL clumps effectively sample the entire evolutionary path of massive Galactic clumps that are likely to host the formation of high-mass stars and clusters. In fact, they have clump-averaged surface densities,  $\Sigma_{cl}$ , ranging from  $\sim 0.1$  up to  $\sim 10 \text{ g cm}^{-2}$ , which covers the thresholds suggested by theoretical modeling and observational constraints for the formation of high-mass stars (Krumholz & McKee 2008; Kauffmann & Pillai 2010; Baldeschi et al. 2017, see also the discussion in Beuther et al. 2025). Moreover, we note that the ALMAGAL sample includes clumps with a ratio of their bolometric luminosity to their mass,  $L/M$ , which is uniformly distributed between  $\sim 0.5$  and  $\sim 50 L_{\odot}/M_{\odot}$ , and includes cases with a  $L/M$  down to  $\sim 0.05$  and up to  $\sim 400 L_{\odot}/M_{\odot}$ , although highly evolved phases are less sampled, as there are fewer systems with  $L/M > 100 L_{\odot}/M_{\odot}$  compared to the other evolutionary stages. Since the  $L/M$  ratio is a well-established indicator of the evolutionary stage of clumps (Molinari et al. 2016a), the ALMAGAL clumps provide a complete and rather homogeneous sampling of the evolution of massive clumps with active star formation. We present in Fig. 1 a thorough comparison between the  $L/M$  ratio, the surface density,  $\Sigma_{cl}$ , and the heliocentric distance,  $d_{cl}$ , of the ALMAGAL targets. There are 25 clumps located at distances greater than 9 kpc after updating their distances that are not shown in the left and central panels for clarity reasons, but the distribution of their properties is similar to the ones presented. Figure 1 shows that there is no relationship between the ratio  $L/M$  and either  $\Sigma_{cl}$  or  $d_{cl}$ , indicating that target selection did not introduce any strong bias due to an eventual correlation between these two quantities. In contrast, the ALMAGAL sample presents a relation between  $\Sigma_{cl}$  and



**Fig. 1.** Distribution of the  $L/M$  ratio and average surface density,  $\Sigma_{cl}$ , as function of the clump heliocentric distance for the ALMAGAL sample, and comparison of these two properties. Each panel includes the 2D density distribution computed dividing the intervals of  $L/M$ ,  $\Sigma_{cl}$ , and  $d_{cl}$  into bins of width 0.3 dex, 0.2 dex, and 500 pc, respectively. The solid light gray lines indicate the contour levels of the 2D density distribution derived with kernel density estimation (KDE) corresponding to 10, 30, 50, 70, 95% of the sample.

$d_{cl}$ , introduced as a consequence of the selection criteria on the clump mass.

## 2.2. ALMAGAL observation strategy and continuum images

The ALMAGAL program adopted an observational strategy of single pointings centered on the emission peaks identified in the *Herschel* Hi-GAL 250  $\mu\text{m}$  images (Molinari et al. 2016b; Elia et al. 2017). Each ALMAGAL clump has been observed with a single pointing observation with ALMA, both with the ACA 7 m and the main 12 m antenna arrays. Observations have been carried out using two different configurations for the main array to image all ALMAGAL clumps with a similar physical resolution of  $\sim 1000$  au (Sánchez-Monge et al. 2025), although they are located over a wide range of heliocentric distances (Molinari et al. 2025). The C2+C5 configuration has been adopted to observe clumps with  $d_{cl} \lesssim 4.7$  kpc, achieving a resolution of  $\sim 0.3''$ , while the C3+C6 configuration for clumps with  $d_{cl} \gtrsim 4.7$  kpc for a resolution of  $\sim 0.15''$ . These configurations cover the long and short baselines observed by the main array which, combined with the ACA compact array, allowed the recovery of the flux over all spatial scales up to  $\sim 29''$ , corresponding to the maximum recoverable scale of the ACA array in band 6. In the following discussion, we refer to the group of clumps observed with the two configurations as near and far groups for the targets observed with C2+C5 and C3+C6 configurations, respectively. Although the two groups were well separated in  $d_{cl}$  when the initial target selection was made, this distinction has been lost with the update of the kinematic distances (see Benedettini et al., in prep.). However, this revision has a marginal impact on the statistical results, as most of the targets in each group are still located in a limited interval of  $d_{cl}$  equal to  $2 \leq d_{cl} \leq 4.7$  kpc and  $4.7 \leq d_{cl} \leq 7.5$  kpc, for near and far, respectively (see also Sect. 3.2.2).

The detailed description of data reduction, calibration, and imaging is presented in Sánchez-Monge et al. (2025). The continuum images obtained from the combination of the data from all the arrays have a circular field of view (FoV) with a diameter of  $\sim 36''$ , limited by the primary beam size of the 12m antenna and the cut to an attenuation of 0.3 applied during the image processing (Sánchez-Monge et al. 2025). In each image, the noise level is rather homogeneous in the central region of the FoV with an average value of  $\sim 0.05$  mJy/beam (Sánchez-Monge et al. 2025), but naturally increases towards the edges due

to the antenna primary beam attenuation that causes a decrease in sensitivity by a factor of two at angular distances greater than  $\sim 13.8''$  from the phase center.

## 2.3. ALMAGAL core catalog

Coletta et al. (2025) extracted compact sources in ALMAGAL continuum images using a modified version of the CuTex photometric package (Molinari et al. 2011). The entire catalog of identified compact sources is composed of 6348 objects detected at  $5\sigma$  in 844 clumps. The nature of these sources can be assessed by their photometric and physical properties. All of these sources have a centrally peaked intensity profile, reflecting the high detection efficiency of the CuTex algorithm for this morphology and its low sensitivity to extended or flat-profile emission. These sources span sizes between  $\sim 500$  and  $\sim 5000$  au and have mean densities greater than  $10^6$  part  $\text{cm}^{-3}$  (Coletta et al. 2025). However, most of them have sizes between 1000 and 2500 au (see Fig. 13 in Coletta et al. 2025) and quite high densities between  $1.4$  and  $8 \cdot 10^7$  part  $\text{cm}^{-3}$ . These properties are characteristic of centrally condensed star-forming cores whose innermost regions have undergone collapse. Hence, it is probable that a large fraction of the ALMAGAL sources already host at least one protostellar object. Following this interpretation, we refer to all sources in the ALMAGAL catalog as cores throughout this paper, with the understanding that they likely represent the inner, high-density regions of larger, more extended structures.

We adopted the positions provided by Coletta et al. (2025) to analyze how the cores are distributed in the fields and to measure their relative separations, with the aim of characterizing the fragmentation process in high-mass clumps. Of the 844 systems, 130 clumps host only one core and are not suitable for our analysis. We recognize that the reliability of our analysis and the diagnostics adopted for the spatial distributions are highly dependent on the number of cores. We consider as our full sample all the systems with  $N_{core} \geq 4$ , excluding the 200 clumps with two or three hosted cores, which are 110 and 90 clumps, respectively (Coletta et al. 2025). The inclusion of these systems has a negligible impact on the results discussed in the paper, and the properties of their separations are consistent with the systems with  $4 \leq N_{cores} \leq 6$ . We adopted the full sample to discuss the distribution of core separation and the comparison with the Jeans fragmentation theory, as systems with a low number of cores

**Table 1.** Size of statistical subsamples (number of clumps,  $N_{clump}$ , and cores,  $N_{cores}$ ) studied in this work, selected with different cuts on the degree of fragmentation.

	Total	$N_{clump}$ near	$N_{clump}$ far	$N_{cores}$ near	$N_{cores}$ far
$N_{cores} \geq 4$	514	330	184	4012	1716
$N_{cores} \geq 8$	296	218	78	3429	1158
$N_{cores} \geq 10$	226	172	54	3033	956

**Notes.** Statistics from the ALMAGAL catalog by Coletta et al. (2025). The subsamples are divided into near and far group, referring to systems with different heliocentric distances, which are observed with different array configurations (see text and Molinari et al. 2025).

may correspond to phases when fragmentation is at the beginning. In contrast, to study the global structural properties of the spatial distribution of the cores, i.e., the type of distribution, mass segregation, we restricted our analysis only to systems with a high degree of fragmentation ( $N_{core} \geq 8$  or higher).

In summary, the analyzed dataset comprises 514 clumps ( $N_{core} \geq 4$ ), containing a total of 5728 cores, greatly increasing the statistics available in the literature. However, we identify subsamples with different cuts on the minimum fragmentation level for the various aspects of the analysis of the core spatial distribution. Table 1 summarizes the sample sizes of these subsamples.

## 3. Results

### 3.1. The spatial distribution of ALMAGAL cores

The FoV of the ALMAGAL continuum images allows us to probe the fragmentation of massive clumps over circular regions with linear diameters of  $\sim 0.3$ – $1.4$  pc in clumps located between 2 and 8 kpc, which are the bulk of the ALMAGAL sample. Visual inspection reveals that the identified cores are arranged in highly diverse configurations. The observed patterns depend on three main factors: the degree of fragmentation (number of cores), the presence (or absence) of extended continuum emission, and the morphology of that emission. The variety in morphologies disclosed by the ALMAGAL data is partially illustrated by the examples presented in Figs. 2 and 3.

We observe a difference between systems where multiple cores are detected and those that are poorly fragmented. In the first case, cores are often grouped together and located in a restricted area of the observed field, typically found on a major patch of extended continuum emission with an irregular morphological appearance. Although there are cases where the emission has a circular shape or modest elongation, it is not uncommon to find cases in which it appears quite elongated (see discussion in Molinari et al. 2025). The distribution of cores typically follows the most prominent continuum emission features, but they are also found outside those boundaries. Cores are also found lying in substructures with a highly filamentary appearance, typical of molecular clouds and present at all spatial scales, as highlighted by *Herschel* observations (André et al. 2014; Schisano et al. 2014, 2020; Könyves et al. 2015; Dib et al. 2020; Schneider et al. 2022; Hacar et al. 2023). Occasionally, these subparsec filaments are found converging toward the main emission patch as in hub-filament systems (Myers 2009; Hacar et al. 2018; Motte et al. 2018) (see the bottom left panel of Fig. 2). In other fields, the extended continuum emission splits into several well-separated

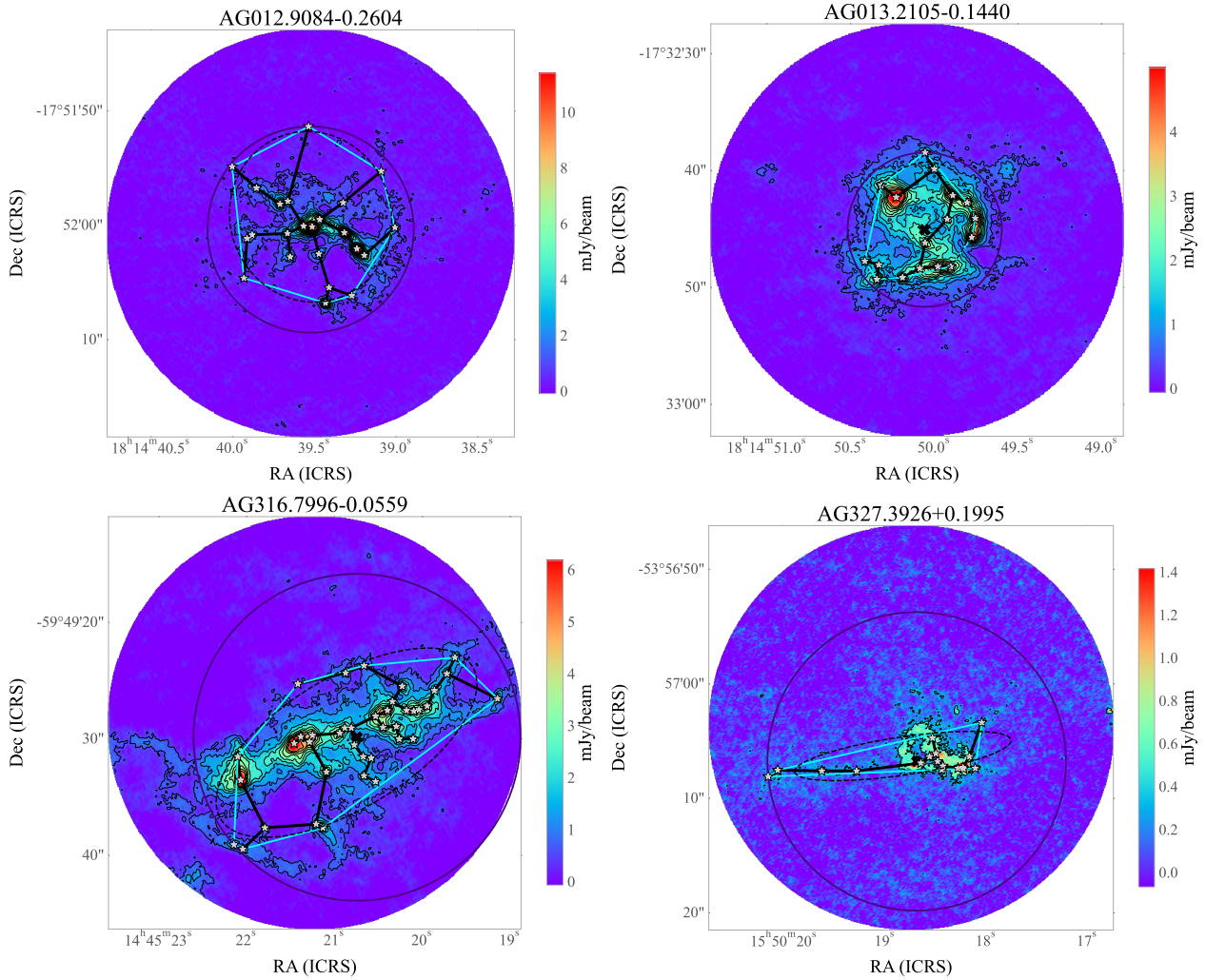
regular patches, each of which hosts a group of associated cores. This is indicative that massive clumps can have an additional higher hierarchical level of substructuring (see the lower panels of Fig. 3), whose presence introduces a subclustering of the core population.

In contrast, systems characterized by a low level of fragmentation do not show a well-defined pattern in the core distribution. These cores are typically sparse in the observed field and do not present a strong signature of grouping or clustering. However, we identified cases where the cores are aligned along specific directions, which extend over several arcseconds. These aligned patterns sometimes match one or a few extended filaments crossing a fraction of the FoV. Even in the cases lacking the detection of extended emission, it may be possible that the hosting filament is undetected for sensitivity reasons and that a signature of its presence is only revealed by the alignment of the cores.

In general, ALMAGAL observations show that the cores are associated with almost every extended region of continuum emission. The high angular resolution reveals that these regions have a wide variety of appearances, as discussed in Molinari et al. (2025). As a consequence of the association, the spatial distribution of star-forming cores shows a similar diversity, likely due to the interplay of different physical factors, such as gravitational force, turbulence, magnetic fields, and feedback mechanisms (Mac Low & Klessen 2004; Krumholz 2006; Li et al. 2014; Shima et al. 2017; Tang et al. 2019; Sanhueza et al. 2025). All of these effects contribute to shape the fragmentation in clumps, affecting both the number and the distribution of cores during their evolution. In this paper, we refer to any group of cores identified in the ALMAGAL clumps with  $N_{cores} \geq 4$  as a “cluster”, even when they are limited in number, smaller than the members of typical embedded star clusters.

In the following sections, we introduce quantitative metrics and diagnostic tools to characterize the variety of observed patterns described above. Our aim was to answer the questions of where the cores are formed, how they are distributed, and what clues of the fragmentation process are left imprinted in their spatial distribution. To this end, we analyzed the size and morphology of the observed core distributions, and then focused on the mutual separation between cores. This is an important observable of the fragmentation process, which we related to the clump-averaged properties and compared to the Jeans fragmentation theory. Finally, we evaluated additional diagnostics capable of defining how cores are distributed and how such a distribution changes as a function of the evolutionary phase of the entire system. In addition, we note that a first rough classification of the observed variety of ALMAGAL clusters can be drawn using the convex hull polygon<sup>1</sup> (CH) applied to the cores. A simple classification derives from the comparison of the number of vertices of the CH,  $N_v^{hull}$ , with the number of cores located well inside the polygon boundaries,  $N_{in}^{hull}$  (see Figs. 2 and 3 for reference). About  $\sim 37\%$  of the ALMAGAL clumps studied in this work (190 out of 514) have more cores within the CH boundaries than at the edges ( $N_{in}^{hull} > N_v^{hull}$ ), a pattern that defines a well-localized group of objects with a surrounding sparse population, which is similar to a star cluster. On the contrary, there are 64 clumps ( $\sim 12\%$ ) where all detected cores form the CH polygon ( $N_{in}^{hull} = 0$ ). These systems have a small number of cores, from four to six sparse cores, and typically lack substantial associated extended continuum emission that would be expected if a large amount of high-density material is present. However, most

<sup>1</sup> The convex hull polygon is the smallest polygon that encloses all the cores and has all interior angles smaller than  $\pi$ .



**Fig. 2.** Examples of the distribution of cores found in four ALMAGAL clumps overlapping on the dust continuum map at 1.4 mm from which they were extracted. These fields show examples of clustered systems with circular and elliptical patterns and cases of aligned cores. The contour levels correspond to  $4n \times \sigma$  level and  $\sigma$  equal to the noise level measured on the image. The positions of the cores from the ALMAGAL catalog extracted by Coletta et al. (2025) are shown with gray star markers. In each field we indicate the cluster geometrical center and its radius (see the definition in the text) with a black cross and a dark blue circle, respectively. We also show the convex hull polygon (cyan segments) and the corresponding best-fitting ellipse (black dashed line) adopted for the morphological characterization of the system. The solid thick black segments are the MST edges determined with Prim’s algorithm (Prim 1957), connecting all the cores in the field.

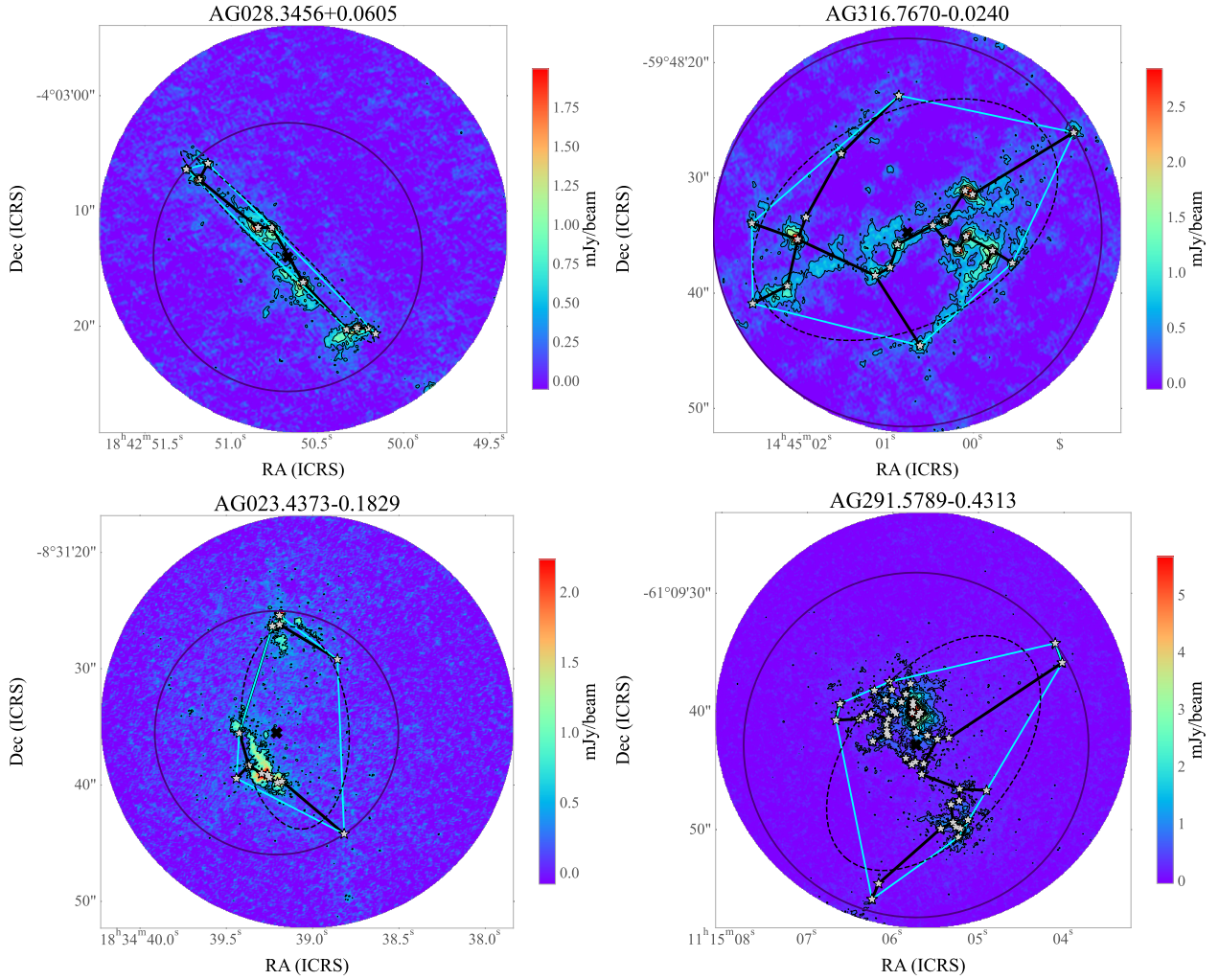
ALMAGAL systems ( $\sim 51\%$ ) are between these two extreme patterns, since they show only a few cores inside the CH polygon. These systems can be young precursors of a clustered system where fragmentation is still at the beginning, or they can be a population of sparse objects derived from the local fragmentation of the hosting molecular cloud. Nonetheless, it is worth to evaluate the core distribution even in this latter case, as clusters are also recognized to form through mergers of group of cores initially distributed throughout the entire cloud (Chen et al. 2021; Dobbs et al. 2022).

### 3.1.1. Center and size of the clusters of ALMAGAL cores

In general, the ALMAGAL cores are located in a portion of the observed FoV, although there are several cases where they spread over the entire, with cores that are also found near the edges. The centers of the core clusters closely match the center of the observed ALMA fields. We adopted two methods to determine the cluster center: (i) the geometrical and (ii) the mass-weighted

average of the core positions. These methods produce similar statistical results, although these positions may differ for individual systems. The ALMAGAL cluster centers are mostly located within  $3''$  from the center of the ALMA observed fields, which corresponds to the photometric center of the Hi-GAL clumps as traced in the *Herschel*  $250 \mu\text{m}$  maps (Molinari et al. 2016b; Elia et al. 2021). The offset between these positions is significant only in a small number of systems: as a reference, it exceeds  $6''$ , which is equal to  $\sim 1/3$  of the FoV, only in 35 fields ( $\sim 7\%$  of the investigated sample).

We estimated the radius of the cluster,  $R_{cluster}$ , as the distance of the farthest core from the cluster center (Cartwright & Whitworth 2004). These radii are smaller than  $13''$  in 384 systems, i.e.,  $75\%$  of the investigated sample. This region, corresponding to  $\sim 0.72 \times \text{FoV}$ , is roughly coincident with the fraction of the images where the noise level is more homogeneous (Sánchez-Monge et al. 2025). However, there are several systems where cores are also detected close to the FoV’s edge; most of those clumps are systems with  $d_{cl} < 4 \text{ kpc}$ . It is likely that



**Fig. 3.** As in Fig. 2. These fields show examples of cores distributed over filamentary features and in well-separated local substructures, both patterns that are sometimes found in ALMAGAL continuum images.

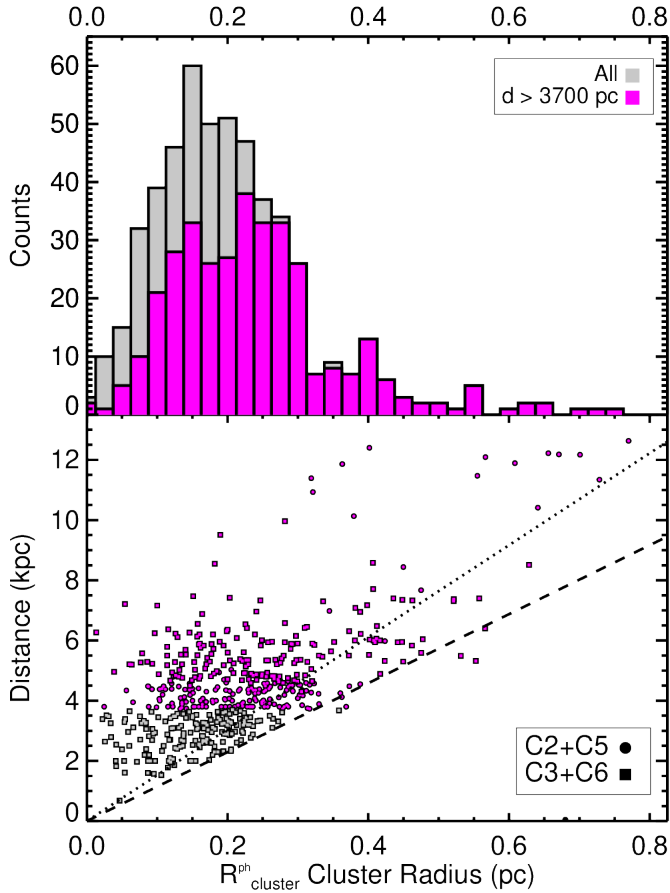
for these systems, the single pointing mode of the ALMAGAL observations is sampling only the inner region of the clump. This is a critical selection effect that must be taken into account.

The top panel of Fig. 4 shows the distribution of the cluster radius measured by converting  $R_{cluster}$  into a linear size,  $R_{cluster}^{ph}$ , using the updated heliocentric distances (Molinari et al. 2025, Benedettini et al., in prep.). The distribution for the full sample shows two key features: a peak close to the median value of  $\sim 0.2$  pc, and a notable drop at  $R_{cluster}^{ph} \approx 0.32$  pc. To determine the impact of the FoV bias on the true physical distribution, we selected the 347 clumps with  $d_{cl} \geq 3.7$  kpc. As shown in the bottom panel of Fig. 4, for those systems, the observed area is sufficiently large to identify clusters with radius  $R_{cluster}^{ph} \geq 0.32$  pc. In this sample, the peak at  $\sim 0.2$  pc disappears, revealing a distribution that is rather uniform in the interval  $0.1 \leq R_{cluster}^{ph} \leq 0.32$  pc. However, the cut-off at 0.32 pc remains even in this selected sample, suggesting that it is a true physical upper limit to the cluster size. We find only 36 systems with  $R_{cluster}^{ph} \geq 0.32$  pc (out of the 302 clumps with  $3.7 \leq d_{cl} \leq 6.5$  kpc), a number that is sufficiently small to not represent a significant statistic. Our conclusion is that the clumps with  $3.7 \leq d_{cl} \leq 6.5$  kpc provide the most robust representation of the true cluster size distribution in a generic massive dense

clump. As a consequence, it is quite likely that there would be cores located outside the observed area in the 168 fields with  $N_{core} \geq 4$  and  $d_{cl} \leq 3.7$  kpc, meaning that the measurements of the core separations can be potentially incomplete in these systems, while they offer a better snapshot of the central region of the clump.

Our estimates of  $R_{cluster}^{ph}$  can be affected by uncertainties. First, this definition of cluster radius depends on the effective membership of the farthest core, which is not always reliably confirmed. Secondly, there is a decrease in sensitivity towards the image edges, meaning that cores with low fluxes located at the margin of the FoV may remain undetected. It is interesting to compare these estimates with the size of their hosting clump, which defines where there is an overdensity of material that is potentially available for star formation.

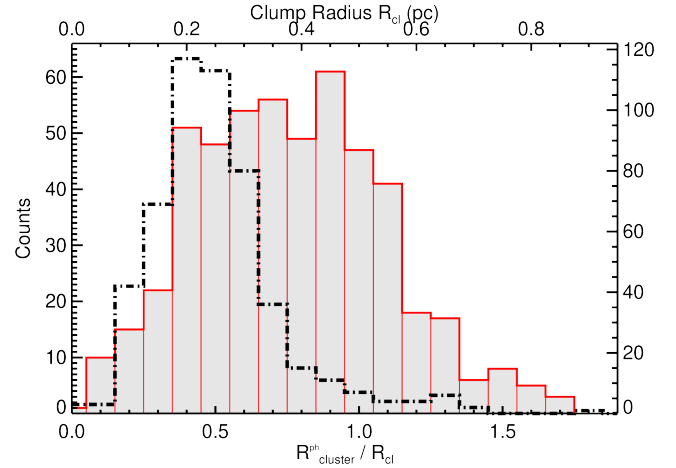
Ideally, it would be useful to compare  $R_{cluster}^{ph}$  with the extension of the continuum emission measured on the same ALMAGAL data. However, as discussed by Molinari et al. (2025), ALMAGAL interferometric data are affected by spatial filtering, and in several fields, they may miss a significant flux fraction. This makes ALMAGAL data unsuitable for reliably tracing the full, extended emission of the clump, even if they reveal the complexity of the clump inner structure. For this reason, we adopted the clump sizes,  $R_{cl}$ , determined from *Herschel* 250  $\mu$ m data and



**Fig. 4.** *Top panel:* distribution of the cluster radius measured from the average core positions for the sample of 514 ALMAGAL clumps with  $N_{core} \geq 4$  (in gray) and the subsample composed 347 clumps with  $d_{cl} \geq 3.7$  kpc (magenta). *Bottom panel:* clump heliocentric distance vs. measured cluster radius. The dashed line shows the physical size of the ALMAGAL FoV, while the dotted line indicates the area where the sensitivity is constant within a factor of two, i.e., angular sizes  $\sim 13.5''$ . Cluster radii exceeding the FoV sizes are possible in systems with a substantial offset between the cluster and the image center.

defined as the circularized half-width at half-maximum of the elliptical Gaussian fit with CuTex (Molinari et al. 2016b; Elia et al. 2021). This definition is certainly limited and a simplification of the complexity of the clump structure. However, our goal here is not to define an exact physical boundary, which is also not well defined in molecular clouds. Rather, our aim is to use a consistent and robust proxy for the region containing the bulk of the clump mass. We note that ALMAGAL single pointings are sufficiently wide to probe a wider area than the *Herschel* beam at  $250 \mu\text{m}$  (with a full-width half maximum of  $\sim 18''$ ).

The size of the clumps ranges from  $\sim 0.1$  to  $\sim 0.9$  pc, and most of them have  $R_{cl} \leq 0.35$  pc, as shown in Fig. 5, which is comparable to our measurements of the cluster radius. The distribution of the ratio of  $R_{cluster}^{ph}$  to  $R_{cl}$  reported in Fig. 5 indicates that the cores are typically located between  $\sim 0.3$  and  $\sim 1.2 \times R_{cl}$ . Hence, the formation of cores takes place in any fraction of the hosting clump, and may even extend to its close outskirts, but rarely beyond the overdensity that has formed in the molecular cloud. Although it is possible that additional cores may reside farther away from the center of the clump, they must have low mass to remain undetected towards the edge of the FoV. As the sensitivity in this area drops by a factor of  $\sim 3$ , these undetected cores



**Fig. 5.** Distribution of the ratio of the radius of the core cluster to the clump radius (gray histogram). The dot-dashed line, which refers to the top and right axes, shows the distribution of the clump radius derived from *Herschel* Hi-GAL data at  $250 \mu\text{m}$  from Elia et al. (2021).

would have masses below  $\leq 0.6 M_{\odot}$ , adopting the completeness limit indicated by Coletta et al. (2025).

Finally, we found that our measurements of  $R_{cluster}^{ph}$  do not show any correlation with the clump surface density  $\Sigma_{cl}$  or the  $L/M$  ratio as shown by Fig. 6. A similar result is found for the ratio  $R_{cluster}^{ph}/R_{cl}$ .

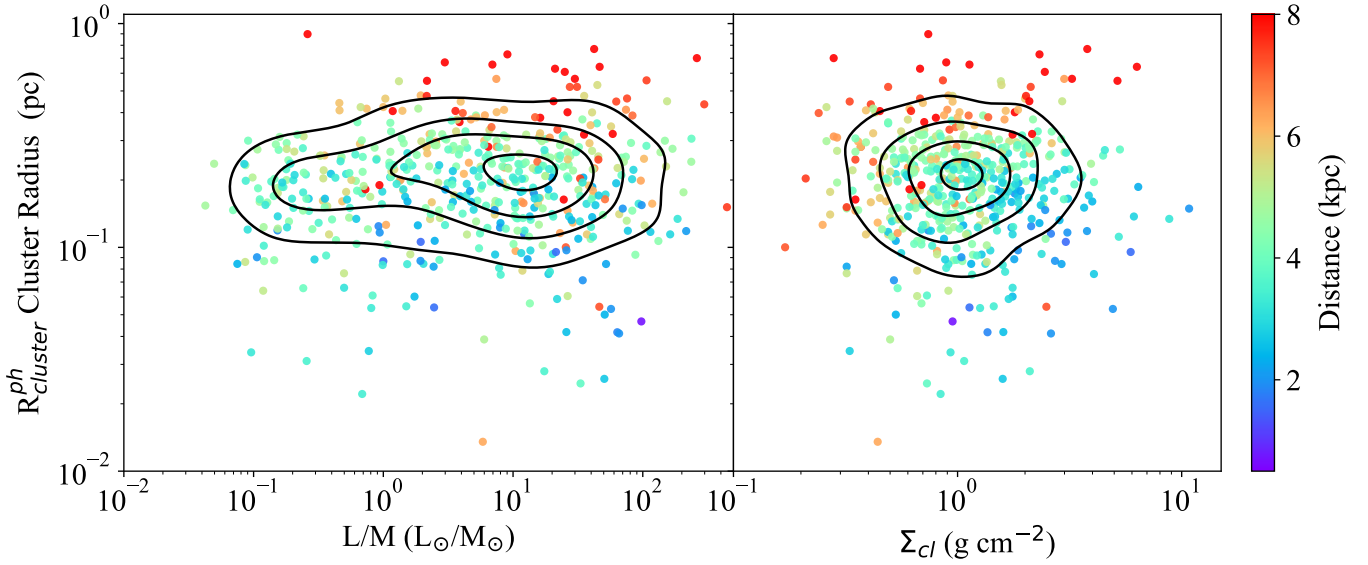
### 3.1.2. Morphology of the clusters of cores

The definition of cluster radius introduced in the previous section provides a robust measurement for spherical systems, for which the projected circular area  $A = \pi R_{cluster}^2$  is a valid approximation for the size of the entire cluster. Nonetheless, cores are often distributed over elongated or irregular regions, for which the circular area gives a considerable overestimate of the cluster size (see, e.g., the bottom panels of Fig. 2). The CH polygon allows us to better estimate the cluster area and determine the occurrence of spherical, elongated, and irregular clusters. This method is widely adopted in studies of embedded clusters in star-forming regions (Schmeja & Klessen 2006; Gutermuth et al. 2009; Dib & Henning 2019) and it was recently adopted in Molinari et al. (2025) to also characterize the extended continuum emission in the ALMAGAL fields that shows significant morphological diversity.

We note that the observed variety in the spatial distribution of cores is not unexpected, since similar results are also found when observing stellar systems, of which the ALMAGAL cores are the precursors. In fact, Kuhn et al. (2014) observed different appearances in young star systems, with the most recurrent morphologies represented by chains of subclusters, clumpy ones, and isolated groups. The elongation of the cluster is a simple metric that can identify these possible patterns. To measure the cluster elongation, we fitted an ellipse to the area covered by the CH, measuring its orientation,  $PA_{CH}$ , and the ratio between the major and minor axes,  $e_{CH}$ . A different method was proposed by Schmeja & Klessen (2006), who introduced the quantity

$$\xi = \frac{R_{cluster}}{R_{CH}}, \quad (1)$$

where  $R_{cluster}$  is the circular cluster radius and  $R_{CH}$  is the convex hull radius, which is defined as the radius of the circle



**Fig. 6.** Cluster radius as a function of the clump-averaged properties:  $L/M$  (left panel) and surface density  $\Sigma_{cl}$  (right panel), color-coded by the clump distance. The black lines indicate the contours of the 2D probability density distribution corresponding to 10, 25, 50, 75, and 95% levels.

**Table 2.** Metrics of the spatial distribution of the ALMAGAL cores: position, sizes, and shape of the distribution.

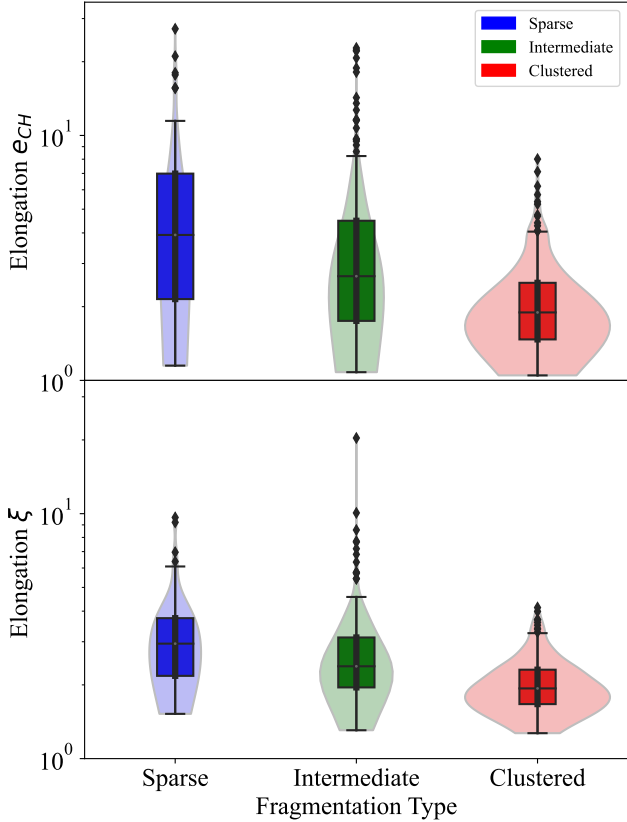
ALMAGAL_ID Field	$N_{cores}$ –	RA <sup>a</sup> deg	Dec <sup>a</sup> deg	Offset <sup>a</sup> arcsec	$R_{cluster}^b$ arcsec	$R_{cluster}^{ph}{}^b$ pc	$A^{CH}{}^c$ arcsec <sup>2</sup>	$A_{norm}^{CH}{}^c$ arcsec <sup>2</sup>	$N_v^{Hull}{}^c$ –	$N_{in}^{Hull}{}^c$ –	$\xi^d$ –	$e_{CH}{}^d$ –	PA <sup>d</sup> degree
AG010.2283–0.2064	9	272.25125	–20.19142	3.10	13.79	1.30E-01	51.67	93.01	4	5	3.40	5.24	–110
AG010.3226–0.1599	11	272.25671	–20.08653	3.96	13.34	2.26E-01	130.42	204.94	4	7	2.07	1.37	–31
AG010.6185–0.0314	6	272.28967	–19.76494	2.68	5.96	1.52E-01	7.17	21.52	4	2	3.95	5.28	–27
AG010.6691–0.2196	5	272.49161	–19.81147	4.96	9.11	1.54E-01	35.82	179.11	4	1	2.70	2.33	–27
AG010.8851+0.1225	4	272.28314	–19.45659	0.43	2.02	2.64E-02	2.55	10.22	3	1	2.24	1.95	–12
AG011.0829–0.5311	6	272.99283	–19.59885	3.12	7.44	1.19E-01	34.28	102.84	4	2	2.25	2.86	56
AG011.3614+0.8017	5	271.89975	–18.71168	3.80	13.76	3.67E-01	112.73	563.64	4	1	2.30	4.30	–266

**Notes.** <sup>(a)</sup> Geometric center of the group of cores present in the ALMAGAL core catalog (Coletta et al. 2025) and its offset from the image center, which corresponds to the clump center measured in the Hi-GAL 250  $\mu$ m maps. <sup>(b)</sup> Cluster angular radius,  $R_{cluster}$  and corresponding linear size,  $R_{cluster}^{ph}$ . <sup>(c)</sup> Parameters of the convex hull (CH) polygon computed from the core positions: area,  $A^{CH}$ , normalized area according to  $A_{norm}^{CH} = A^{CH}/(1 - N_v^{hull}/N_{cores})$  from Schmeja & Klessen (2006), number of cores at the CH vertices,  $N_v^{hull}$ , and inside the polygon,  $N_{in}^{hull}$ . <sup>(d)</sup> Elongation of the cluster: ratio of cluster radius to CH equivalent radius,  $\xi$ , and ellipticity  $e$  and position angle PA of the best fitting ellipse to the convex hull. The complete version of this table is available at the CDS.

with an area equivalent to CH,  $A^{CH}$ . We note that Schmeja & Klessen (2006) adopted an additional correction factor  $1/(1 - N_v^{hull}/N_{cores})$  to statistically recover the real area on which the sources are distributed (Hoffman & Jain 1983) (see also the discussion in Appendix A of Parker (2018)). However, such a factor was found to be reliable only when  $\sim 200$  points are available (Ripley & Rasson 1977), while in our case it would introduce large discrepancies since  $N_v^{hull}$  may be very close to  $N_{cores}$ . Therefore, we computed  $R^{CH}$  without applying this factor, although we also report the CH area normalized by it in Table 2. The distributions of the two estimators  $e_{CH}$  and  $\xi$  are similar, with similar average values. The distribution of  $\xi$  is narrower as  $\xi$  has a smaller dynamic despite the irregular shape of the CH. The most common shape of the cluster is a slightly elongated ellipse, with a median  $e_{CH} \sim 2.2$  [ $\xi \sim 2.3$ ]. While we found systems in which the cluster has a rather circular shape ( $e_{CH} \leq 1.3$ ), they represent a minority with only 46 cases ( $\sim 9\%$ ) of 514. Other systems have strong elongations as high as  $\approx 20$ – $30$ , depending on the adopted estimator.

Approximately circular shapes may result from a distribution where cores are uniformly distributed in the system, but also from patterns where a few filamentary features are

converging towards a central hub and are symmetrically distributed over different directions, as in the top left panel of Fig. 2. Figure 7 shows the distribution of the measured elongations for the clusters classified into the three groups described in Sect. 3.1. Similar results are obtained if the classes are defined in terms of the fragmentation level. High elongations are measured in systems with a limited number of cores and include the cases where they align along specific directions or are distributed over extended filamentary structures. Clustered systems typically have smaller elongation, with a median and interquartile range equal to  $\xi = 1.9_{-0.3}^{+0.4}$  ( $e_{CH} = 1.9_{-0.4}^{+0.6}$ ). As they are representative of systems with a high degree of fragmentation, this means that their cores are typically distributed over half of the circular area ascribed to the cluster, in slightly flattened structures like the ones shown in the bottom left panel of Fig. 2. However, it is worth noticing that a fraction of  $\sim 7$ – $16\%$  (depending on the adopted estimator) of the 190 systems classified as clustered have elongations greater than three. While these high values suggest rather elongated morphologies, they also derive from peculiar distributions of the cores with highly asymmetric patterns. An example is shown in the bottom right panel of Fig. 2, where a group of cores is aligned along a direction converging towards an



**Fig. 7.** Whisker plots showing the distribution of the two measured estimates of the cluster elongation for the different classes of ALMAGAL clumps identified by the number of cores inside the CH polygon: sparse, where no further core lies inside it; intermediate, where there are fewer cores than the ones defining the polygon itself; clustered, where there are more cores than the ones at polygon vertices. The left panel presents the elongation  $e_{CH}$ , defined as the ratio of the major to the minor axis of the best fitting ellipse to the CH polygon, while the right panel refers to the ratio  $\xi$  of the circular area of the cluster to the area of the CH polygon. The first and third quartiles, and the 10th and 90th percentiles of the distribution are indicated by the boxes and the segment extremes, respectively, while the central segment corresponds to the median.

extended and rather circular emission causing a large elongation due to their separation from the main system.

### 3.2. Core separations

#### 3.2.1. Methods

Several mathematical tools are available to characterize the spatial distribution of a set of localized points, which are representative of the members of a cluster. In particular, an important metric to measure is the mutual separations between those points, whose distribution may be related both to the fragmentation scales and to the degree of clustering (Cartwright & Whitworth 2004; Schmeja et al. 2008; Sánchez & Alfaro 2009; Allison et al. 2009; Parker 2018; Dib & Henning 2019). Those tools are based on the computation of the distance matrix  $\delta(i, j)$ , a symmetric matrix whose elements are the angular distances between the points taken pairwise. Given  $N$  sources, the distance matrix  $\delta(i, j)$  includes all possible  $\frac{(N-1) \times N}{2}$  connections and fully characterizes the relative separations for all elements of the cluster. In this work, we do not require the complete information provided by  $\delta(i, j)$ , as, to describe the length of fragmentation, it

is sufficient to relate each core to its direct neighbor. Those separations are usually computed with the minimum spanning tree (MST) method.

The MST is a graph that connects all cores of the system and corresponds, among all the possible connecting paths, to the set of segments with the shortest total length. It is composed of  $(N - 1)$  segments, named MST edges, that are a subset of the  $\delta(i, j)$ . A simple way to estimate the typical separations between the cores is the statistical analysis of the distribution of the MST edges.

We computed the MST with an IDL implementation of the Prim's algorithm (Prim 1957), identifying in the distance matrix  $\delta(i, j)$  the edges  $l_i$  that compose the MST. Prim's algorithm starts from a single position, adopted as a vertex, and identifies the shortest among all the possible connections with the adjacent sources, then proceeds iteratively until all the sources are finally connected, avoiding to form closed loops in the path. Examples of the computed MSTs are presented in Figs. 2 and 3. We identified a total of 5214 edges,  $l_i$ , measured in all 514 ALMAGAL fields analyzed in this study.

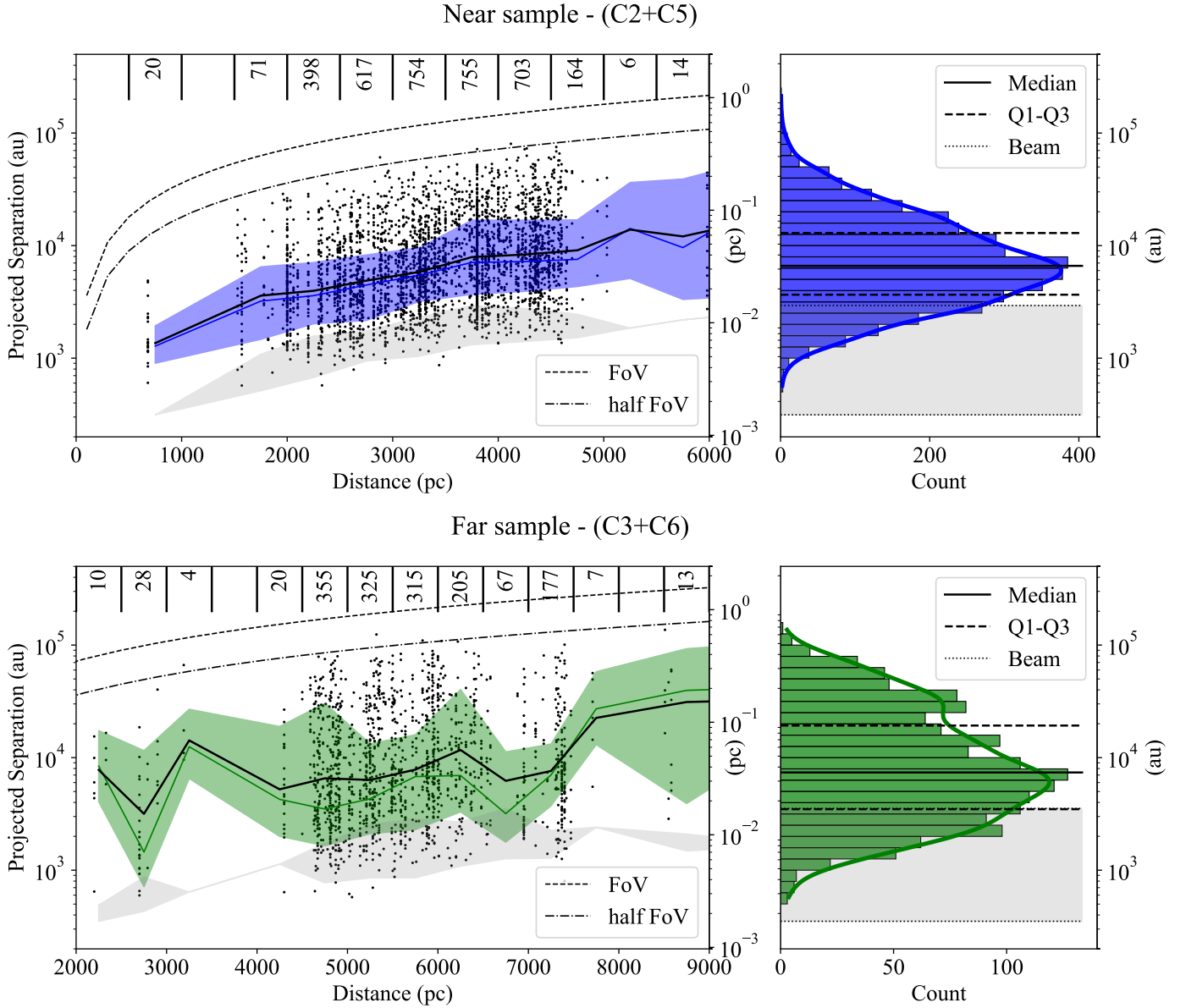
We followed two different approaches in the analysis of such a large ensemble<sup>2</sup> of separations. On the one hand, we independently analyzed the distribution of the separations in each clump. We measured statistical quantities, such as the median and mode of distribution, to provide an estimator for the characteristic core separation in each system. On the other hand, we consider the entire ensemble as a diagnostic of the distributions of core separations, relating all the measured separations to the clump physical properties. In such a case, the statistical estimators are determined over subsamples obtained by stacking all the separations measured in clumps with similar values of a specific property, such as the  $L/M$  ratio, ignoring that those clumps are distinct structures where possible different external conditions may be present. These two approaches allow for a well-rounded and complete description of the statistical information present in our measurements and to determine the possible link to the clump-averaged properties.

With the first approach, we provided estimates for the core separations that are clump-averaged, similar to other physical properties provided for clumps, such as density or temperature. These estimates also highlight the information provided by poorly fragmented systems, which would be hidden in the entire ensemble, as they are numerically outweighed by the contribution of clumps with several cores. Instead, with the entire ensemble, we provide an overall snapshot of the fragmentation process and relax the assumption of a unique characteristic core separation for clump. This analysis method provides a better statistical characterization thanks to the size of the sample and a more robust handling of outlier cores, whose presence would affect the clump-averaged estimators, especially for systems with a low degree of fragmentation.

#### 3.2.2. The distribution of core separations

We divided the entire ensemble according to the set of array configurations adopted for the ALMAGAL observations, i.e., the near and far groups, respectively. The reason for this choice is twofold. First, the FoV is resolved by these two configurations

<sup>2</sup> We use the term ensemble as in statistical mechanics to highlight the implicit assumption that the observed systems are statistical realizations of the same fragmentation process active in massive clumps. With this in mind, each observed pattern is a possible state undergone by this process.



**Fig. 8.** *Left panels:* linear separation between ALMAGAL cores as a function of the distance of the ALMAGAL clump for near (*top panel*) and far configuration sample (*bottom panel*). The gray shaded areas indicate the ranges of the limiting sizes set by the beams of the ALMAGAL images. The black and colored solid lines (blue and green) indicate the median and the mode of the distribution of the separations, respectively, while the colored area traces the distribution width, defined at the 68% peak intensity. The dashed and dotted lines indicates one-half of the observed field and the full field, respectively. The vertical lines show the adopted distance bins with the relative number of the separations. The 20 separations lying below the spatial resolution are found where the beam is strongly elongated, and are larger than the beam minor axis. In addition, a limited number of separations are not included in the figure as they exceed the selected intervals: 180 in the near sample (162 of which have  $d_{cl} \geq 9$  kpc), and six in the far sample. *Right panels:* overall distribution of the projected linear separations in the near (*top*) and far (*bottom*) sample, respectively. The thick colored lines (blue and green) show the probability distribution function estimated with the KDE method. The gray shaded area indicates the interval spanned by the circularized beams in the corresponding ALMAGAL images.

in a different number of synthesized beams, meaning that these two datasets could potentially provide a substantially different density of point-like sources. In more detail, the FoV of the combined ALMAGAL continuum images is covered by  $\sim 2000$  and  $\sim 8100$  synthesized beams with the C2+C5 and C3+C6 configurations, respectively. Secondly, this division allows us to control the impact of the clump distance, as these two groups are composed of clumps located mainly in two different ranges of heliocentric distances. The two subsamples are composed of 4012 and 1716 cores, identified in 330 and 184 clumps, and provide 3682 and 1532 separations, respectively, in the near and far

groups. The discrepancy in the number of cores is mainly due to the sensitivity limit of the survey as discussed by Coletta et al. (2025).

The right panels of Fig. 8 show the distributions of the projected linear separations between the cores given by the MST edges,  $l_i$ , obtained with the revised heliocentric distances for the ALMAGAL clumps. Most of these separations ( $\sim 70\%$ ) exceed the spatial resolution of the data by a factor of  $\sim 3$ . In general, they range from  $\lesssim 1000$  to  $\sim 100\,000$  au (corresponding to  $\lesssim 0.005$ – $0.5$  pc), with the median of the two distributions equal to  $\sim 6600$  and  $\sim 7400$  au for near and far sample,

respectively. Although the two distributions have a similar median and first quartile, they do not derive from the same underlying distribution, as indicated by the Kolmogorov–Smirnov (KS) test rejecting this null hypothesis with a  $p \lesssim 10^{-5}$ . The major difference between the two distributions is given by the large tail present in the far sample due to an increase in the number of separations with  $l_i > 15\,000$  au, caused by the wider physical area covered by the FoV in those clumps. Those separations produce a secondary bump in the distribution at  $\sim 30\,000$  au.

We present the projected linear separations as a function of the clump distance in the left panels of Fig. 8 to compare them with the spatial resolution and the FoV of the observations. Clumps with heliocentric distances greater than 9 kpc are omitted in the plot for better visual clarity, but are included in the analysis. We found separations between cores ranging from the beam up to half the FoV’s size, indicating that the enlargement of the observed area includes additional cores located at the outskirts of the system that are more widely separated. The inclusion of these cores introduces a weak trend in the median and mode of the distribution, calculated in bins of width 500 pc, as also shown in Fig. 8. The mode increases steadily with the distance for both samples and varies from  $\sim 3000$ – $3500$  au at a distance of 2 kpc up to  $\sim 7000$ – $7500$  au for  $d_{cl} \approx 4$  kpc. A similar trend, but with slightly larger values, is found for the median of  $l_i$ , which varies from  $\sim 4000$  to  $\sim 9000$  au and  $\sim 7000$  to  $\sim 8000$  au for the near and far group, respectively. The median and mode of the distributions of  $l_i$  have the largest variations for distance  $d_{cl} < 3.7$  kpc, where we most likely observed only the inner region of the cluster, as discussed in Sect. 3.1.1. This suggests that typical separations in the central portion of these clusters may be shorter, with differences in the estimated characteristic lengths in the densest regions of the clump. The gradual shift of the median towards larger values is ascribed to the inclusion of the cores that are more widely separated. The variations of the mode are explained by the fact that the inclusion of cores with larger separations extends the interval of the data analyzed with the KDE method. As the interval increases, the KDE computation adopts a wider bandwidth, with the effect of shifting the position of the peak of the continuous distribution to larger values. Finally, the distributions are affected by the sensitivity limit of the observations, since in the clumps located closer to the Sun it is possible to identify fainter cores. The ALMAGAL catalog is complete for objects  $M \gtrsim 0.2 M_\odot$  (Coletta et al. 2025), but less massive cores are regularly detected in clumps with  $d_{cl} \leq 4$  kpc. To deal with this effect, we discuss separately the near and far groups. Moreover, we also checked whether our results holds when selecting only the clumps located in narrow distance intervals.

### 3.2.3. Deprojection of the measured separations

The measured separations correspond to the projection in the plane of the sky of the intrinsic 3D separations between cores hosted in the clump. The distribution of 3D separations can be statistically derived by evaluating the effect of the projection on any generic vector randomly oriented in the 3D space. The details of the calculations are reported in Appendix A. On average, the projection along the line of sight of a generic vector oriented in the 3D space reduces its intrinsic length by a factor equal to  $\pi/4$  (see Eq. (A.4)). As a first approximation, we corrected the measured separation between each core pair by this factor to scale to an average deprojected value, which we

consider as the intrinsic separation  $l_i^{depr}$ . In such a case, the distributions of  $l_i^{depr}$  retain the shape of the projected separations as we simply applied a scaling factor between the two quantities.

### 3.2.4. Characteristic separations between cores

The variety of observed patterns is reflected in different distributions of the separations between the hosted cores in each ALMAGAL field. To provide a representative characterization of these distributions, we determined the minimum, maximum and median of the separation  $l_i$ , which we report in Table 3. We also included the mode of the distribution to indicate the most frequently observed separation, which is particularly relevant for systems where one or more groups of cores are surrounded by a more widespread population. In these systems the median would provide a bias estimator, which may not be representative of the typical observed separation. In each clump we estimated the continuous probability density function  $P^k(l_i^k)$  of the separations of the hosted cores using the Gaussian KDE method implemented in the *scipy* library. We identified the peak position of  $P^k(l_i^k)$  to determine the mode,  $l_{mode}$ , and the spread of the distribution, the latter defined as the width of  $P^k(l_i)$  at 68% of the intensity peak, avoiding any assumption of the shape of the PDFs, which we observed to be often quite asymmetric. These quantities are measured for the 437 clumps hosting more than 5 cores, which provide at least four measurements for the core separations, and are included in Table 3. Statistically deprojected values in the  $k$ -th clump,  $l_i^{depr,k}$ , can be obtained by scaling by the constant factor  $4/\pi$  (see the discussion in Sect 3.2.3).

Our estimates of the characteristic, clump-averaged, separation are shown in Fig. 9. They vary from  $\sim 1500$  to  $\sim 60\,000$  au, and values larger than  $\sim 15\,000$  au are found mainly in clumps with a low level of fragmentation. A decrease in the average separations in highly fragmented systems is expected for simple geometrical reasons, but we verified that the observed trend is not caused by the reduction of the available area per source. In fact, we compared with the expected decrease in separations when sources are uniformly distributed over a field with a fixed angular size and found that the observed averages are sensibly smaller than these expectations.

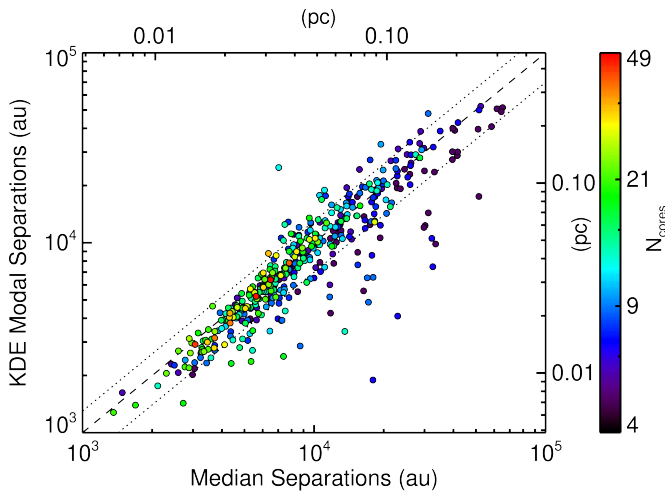
However, we note that the clump-averaged estimates provide limited information about the distribution of separations in a clump. Individual distributions  $P^k(l_i^k)$  have a diversity of shapes, often presenting multi-modal patterns, features that are not easily captured by a single averaged quantity. A thorough statistical analysis and classification of the distribution  $P^k(l_i^k)$  will be conducted in a future study. Here, we compared  $l_{median,k}$  and  $l_{mode,k}$  in Fig. 9 to report the occurrence of asymmetric distributions. These two quantities agree within 25% for 316 systems ( $\sim 72\%$ ), and represent cases where  $P^k(l_i^k)$  is rather symmetric around a central value equal to the measurements provided. However, noticeable differences between the two estimators are not rare. In systems where they differ, the distributions of core separations  $P^k(l_i^{depr,k})$  are sensitively skewed and asymmetric. Quantitatively, the median separation is larger than the mode in 96 systems ( $\sim 22\%$  of the sample), with such a pattern found in both near and far systems, with 59 and 37 cases, respectively. The opposite case, with a median smaller than the mode, is rarer and is found in only 25 systems ( $\leq 5\%$  of the sample).

Asymmetric distributions can be caused by the complex morphology of the 3D clump structure and the spatial distribution of cores. However, it is important to note that it is extremely unlikely to obtain multi-modal distributions for the projected

**Table 3.** Clump-averaged thermal Jeans parameters and reference values of the distribution of projected linear core separations from MST analysis.

ALMAGAL ID	$N_{cores}$	$M_J^{th}$ $M_\odot$	$\lambda_J^{th}$ au	$l_{min}^b$ au	$l_{max}^b$ au	$l_{mean}^b$ au	$l_{median}^b$ au	$l_{25\%}^b$ au	$l_{75\%}^b$ au	$l_{mode}^c$ au	$l_{-68\%}^c$ au	$l_{+68\%}^c$ au
AG010.2283–0.2064	9	0.5	5251	744	23 835	6519	3723	2443	13 387	7503	2773	21 050
AG010.3226–0.1599	11	1.3	8512	2964	16 956	5861	4336	3283	7756	8776	6160	13 596
AG010.6185–0.0314	6	1.0	13 997	1982	21 710	9671	2419	2007	20 236	2708	1162	26 768
AG010.6691–0.2196	5	0.8	10 478	6226	37 678	15 394	10 598	7072	37 678	12 291	6788	24 793
AG010.8851+0.1225	4	2.9	18 062	3518	5452	4219	3687	3518	5452	–	–	–
AG011.0829–0.5311	6	0.6	7802	2991	23 130	12 189	10 585	7676	16 563	13 387	5816	26 683
AG011.3614+0.8017	5	1.0	11 826	9575	66 471	26 125	17 459	10 997	66 471	24 358	12 821	52 485
AG011.9039–0.1403	11	2.0	12 787	4624	25 002	12 247	13 000	5659	16 270	21 127	6760	32 537
AG011.9183–0.6122	13	1.7	10 855	1589	30 588	8677	6800	4456	11 055	9616	4054	22 133
AG011.9370–0.6160	22	2.1	11 578	1872	41 345	9146	6556	3477	9418	8530	2779	17 002
AG012.2087–0.1017	23	0.8	5044	1012	7181	3037	2962	2135	3813	5615	3780	8545
AG012.4040–0.4681	6	1.2	12 773	14 294	41 567	27 021	28 396	21 264	29 583	38 511	24 560	55 522
AG012.4977–0.2233	6	0.8	10 009	15 213	34 785	26 925	29 814	20 368	34 443	51 966	33 141	67 659
AG012.6794–0.1830	8	0.8	6641	1898	22 572	8803	8355	3595	11 824	19 404	6398	43 866
AG012.7207–0.2175	7	0.9	8644	4273	10 152	7353	8588	4495	10 017	15 317	7458	20 427
AG012.7360–0.1031	11	1.3	15 477	3035	20 060	11 353	12 359	8505	13 506	15 213	10 302	21 802
AG012.8535–0.2265	11	1.5	12 384	2709	31 920	13 985	10 561	4667	23 744	28 203	5135	57 232
AG012.8887+0.4890	15	2.0	9109	1221	20 909	7090	4771	2083	9136	5304	2375	28 271
AG012.9008–0.2404	10	0.7	6696	1905	9649	4610	4311	3565	5014	9111	6670	12 296
AG012.9048–0.0306	14	1.7	12 699	651	52 167	11 452	4036	3047	7341	4533	2104	10 310
AG012.9084–0.2604	23	2.3	15 929	1408	16 930	5937	6124	3398	7657	7850	5006	11 802

**Notes.** <sup>(a)</sup> Clump-averaged quantities computed from properties reported in [Molinari et al. \(2025\)](#) <sup>(b)</sup> Minimum,  $l_{min}$ ; maximum,  $l_{max}$ ; mean,  $l_{mean}$ ; median,  $l_{median}$ ; first quartile,  $l_{25\%}$ ; and third quartile,  $l_{75\%}$ , of the distribution of projected separations. <sup>(c)</sup> The modal,  $l_{mode}$ , and the values at  $\pm 68\%$  peak intensity,  $l_{\pm 68\%}$ , are determined from empirical density distributions computed with the Gaussian kernel density estimate (KDE) method ([Silverman 1986](#)). The kernel bandwidth is derived from the data applying the Scott's rule ([Scott 1992](#)). The complete version of this table is available at the CDS.



**Fig. 9.** Relation between the median of the core separations in each clump and the mode determined with the KDE method. The colors indicate fields with different number of cores, ranging from 4 to 50 in logarithmic scale. The dashed line indicates the 1:1 relation, while the dotted lines indicate the range of variation of  $\pm 25\%$ .

separations starting from a single peaked distribution for the intrinsic 3D separations. One possibility is that these patterns are due to the limited degree of fragmentation observed, which could not be sufficient to properly sample the true distribution. Alternatively, they may also be the signature of different fragmentation conditions in groups of ALMAGAL clumps, with diverse characteristic separation lengths. We observed cases

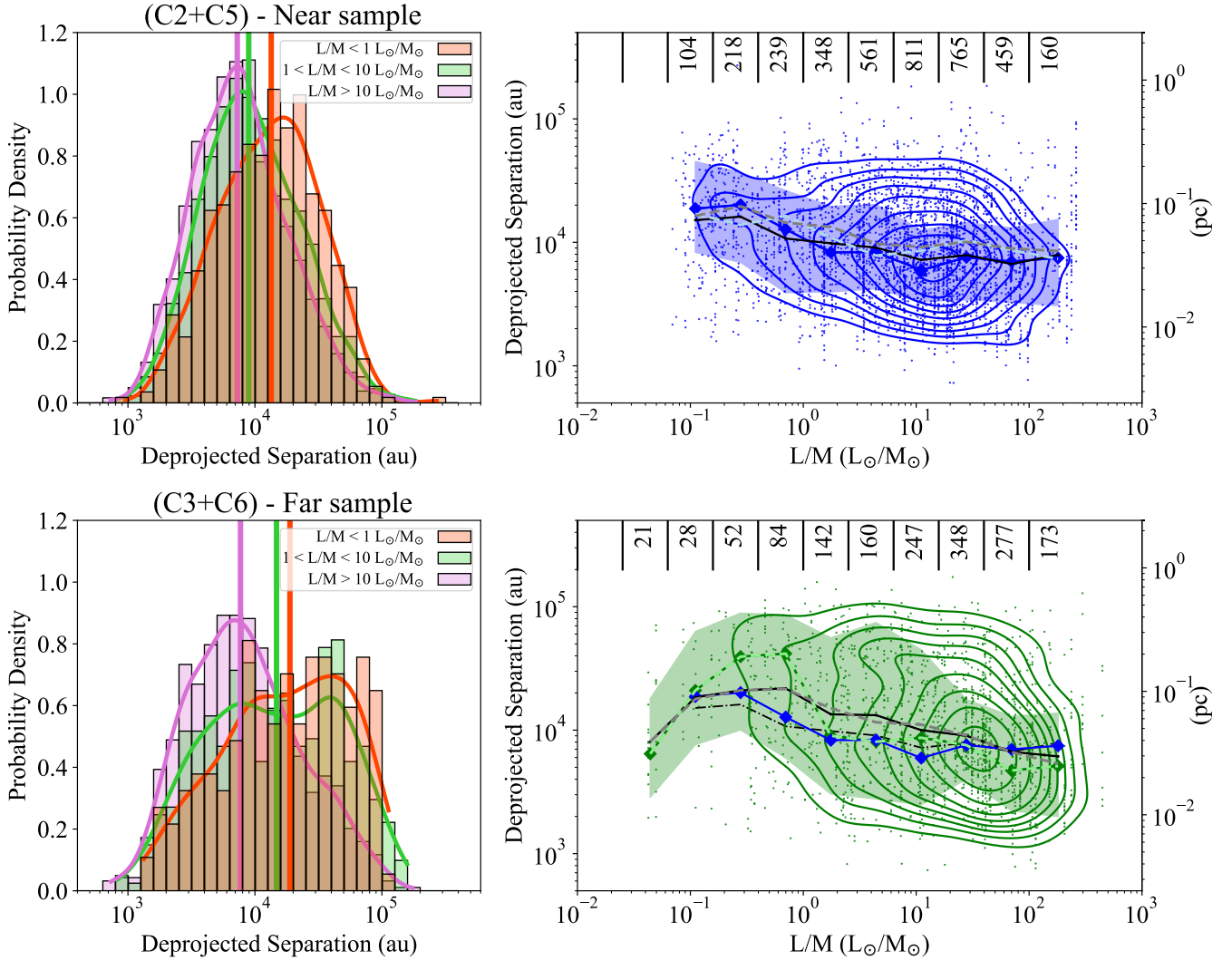
where a fraction of the core population is located farther away than the main group of more tightly packed cores. Indications that multiple characteristic lengths may be present in the fragmentation of massive clumps were already found in the literature ([Zhang et al. 2021](#)).

### 3.3. Relation between the core separations and the clump properties

The core positions directly probe the fragmentation of the clump as they indicate where the dense material collapsed. The distribution of patterns and core separations reflects the diversity of initial conditions or evolutionary paths that are observed in the ALMAGAL systems. This wide observed diversity has also been highlighted by [Elia et al. \(2026\)](#), who related the properties of the core population to the ones of the hosting clumps finding a significant scatter in all their relations, interpreted as a result of different fragmentation conditions ([Elia et al. 2026](#)). Nonetheless, they also identified statistical trends between some fragmentation metrics. For example, they found that the number of cores,  $N_{cores}$ , increases with  $\Sigma_{cl}$  and the evolutionary indicators  $L/M$  ratio and  $T_{bol}$ . Similarly, here we discuss how the separations between cores relate to the clump-averaged properties, using the values derived from the *Herschel* Hi-GAL data ([Molinari et al. 2025](#)) (see Sect. 2.1).

#### 3.3.1. Correlation with the clump evolution

The formation of cores is a consequence of the collapse of the dense material, which proceeds over timescales that are a few times the free-fall time of the gas. The degree of fragmentation



**Fig. 10.** *Left panels:* density distribution of the deprojected core separations measured in the all clumps with  $L/M \leq 1$  (orange),  $1 \leq L/M \leq 10$  (green), and  $L/M \geq 10 L_{\odot}/M_{\odot}$  (purple) with the continuous distribution computed with the KDE method overlapping with the colored continuous line. Clumps of the near and the far are shown in the top and bottom panel, respectively. The colored vertical thick lines show the median of these distributions. *Right panels:* Deprojected core separations as function of the clump  $L/M$  for the systems in the near (*top panel*) and the far (*bottom panel*) subsample, respectively. The thick black lines indicate the median of the distributions of the core separations computed in bins of  $L/M$  of width 0.3 dex. The thick colored lines, in blue and in green for the near and far subsample, respectively, indicate the peak of these distributions, while the colored area traces their spread defined as the width at 68% of peak intensity. The dashed lines show the variation of the mode (in light blue and green) and of the median computed over subsamples having  $3.7 \leq d_{cl} \leq 4.7$  kpc and  $4.7 \leq d_{cl} \leq 6.5$  kpc. The numbers at the top give the size of the subsamples from which the distribution in each bin is computed. The colored contours trace, from the outside inward, fractions of the sample lying outside that rise from 15% to 95%, with a step of 10% for the near (in blue) and far (in green) subsample, respectively. The median and the peak derived for the near sample are also indicated in the bottom panel for comparison purposes.

changes with time, as more clump material may become gravitationally unstable, depending on the local conditions, and cores evolve into stars. The local conditions vary over the entire clump evolution because of the internal dynamics of the clump or the feedback of the newly formed stars. Changes in the separations between cores during the clump evolution, if present, are useful to trace the variations of these conditions, determining how fragmentation proceeds.

We show in the left panels of Fig. 10 the normalized density distributions of the deprojected separations measured in clumps divided into three main classes according to the value of the ratio  $L/M^3$ , which is usually adopted as a diagnostic for the evolutionary state of the clump (Molinari et al. 2016b) (see

<sup>3</sup> Two clumps, AG014.9963–0.6733 and AG305.2021+0.2073, both of the near sample do not have estimates for luminosity and mass, as

Sect. 2.1). The classes were chosen to separate the clumps in the youngest phases of their evolution, which we considered as the systems with a ratio  $L/M < 1 L_{\odot}/M_{\odot}$ , from those that are highly evolved and host multiple protostars, which we selected here as those with a ratio  $L/M > 10 L_{\odot}/M_{\odot}$ . These values match the intervals identified by Molinari et al. (2016a) for different phases of clump evolution, and allow us to also compare our results with those of the surveys ASHES (Sanhueza et al. 2019), CORE (Beuther et al. 2018), and ASSEMBLE (Xu et al. 2024), each of which observed systems in a specific evolutionary phase. The two classes of young and evolved systems are composed of 129 and 209 clumps, respectively, leaving 176 clumps in the intermediate one (see Table 4). Figure 10 shows that the

they are saturated in *Herschel* data. Those systems have 4 and 15 cores respectively, so a total of 17 separations are excluded from the analysis.

**Table 4.** Size of subsamples defined by different intervals of  $L/M$  ratio and split into the near and far groups.

$L/M$	$N_{clump}$	$N_{clump}$	
$(L_{\odot}/M_{\odot})$	(C2+C5)	(C3+C6)	Total
	near sample	far sample	
$L/M \leq 1$	89	40	129
$1 \leq L/M \leq 10$	111	65	176
$L/M \geq 10$	130	79	209

distribution of the young class, with  $L/M < 1 L_{\odot}/M_{\odot}$ , and the evolved one, with  $L/M > 10 L_{\odot}/M_{\odot}$  are substantially different. The peak and median decrease to smaller values in the population of evolved clumps, which is also characterized by a narrower distribution. The Kolmogorov–Smirnov and Mann–Whitney U tests highlight the differences of these distributions, since they both return a  $p$ -value  $\ll 10^{-7}$ , rejecting the null hypothesis that these two datasets derive from the same underlying population. Specifically, in young systems with  $L/M < 1 L_{\odot}/M_{\odot}$ , the near and far distributions have median values equal to  $\sim 12\,000$  and  $\sim 18\,000$  au ( $\sim 0.06$ – $0.085$  pc), respectively. However, these values describe only partially the distributions presented in Fig. 10, which have complex shapes. The distribution peaks estimated with the KDE method correspond to  $\sim 14\,000$  and  $\sim 34\,000$  au, respectively. These values overlap with the separations equal to  $\sim 14\,000$ – $35\,000$  au measured in the 39 protoclusters observed by the ASHES survey (Sanhueza et al. 2019; Morii et al. 2024). We note that while the ASHES survey maps a larger angular area than ALMAGAL’s single pointings for near clumps, the two surveys cover a similar physical area in the more distant far sample, for which the comparison is less biased by the observed area. The distribution of the latter sample is broad and negatively skewed, with a peak that matches the upper values reported by Morii et al. (2024). An additional comparison is provided by the 12 young starless clumps observed by Svoboda et al. (2019) where the median of the measured projected separations is  $0.083$  pc, corresponding to a deprojected value of  $\sim 22\,000$  au, which is marginally larger than the median values reported above, but still compatible with the ones measured in the youngest systems of our ALMAGAL sample.

In the evolved clumps with  $L/M > 10 L_{\odot}/M_{\odot}$ , the separations decrease to median values of  $\sim 7\,300$ – $7\,700$  au ( $\sim 0.035$ – $0.038$  pc). These values match the results reported by the CORE (Beuther et al. 2018) and ASSEMBLE (Xu et al. 2024) surveys, which targeted the high-mass and high-luminosity clumps characterized by  $10 \leq L/M \leq 1000 L_{\odot}/M_{\odot}$  and that are in the later phases of their evolution. The median of the separations measured in the 20 clumps of the CORE survey is equal to  $\sim 6\,000$  au, with an interquartile interval varying from  $\sim 4\,000$  au to  $\sim 15\,000$  au, once their measurements are corrected for the projection effect (see Sect. 3.2.3). Even if this value is marginally smaller than our measurements, it is considered compatible given the widths of the distributions presented in Fig. 10. The typical separation measured in the 11 high-mass clumps of the ASSEMBLE survey is equal  $\sim 0.035$  pc ( $\sim 7\,200$  au) which is in agreement with our results.

A considerable fraction of the ALMAGAL clumps have a  $L/M$  ratio between the two values discussed above. In these systems, we find a different distribution of the separations for the clumps of the near and far sample. In the near sample, the

distribution is similar to that of the evolved systems, the opposite holds for the far sample, where it resembles that of the young clumps. We measured median values between  $\sim 9\,000$  and  $\sim 15\,000$  au, with the spread influenced by the wide area probed in the far sample. However, the KS test rejects the null hypothesis that each of the two sets of distributions derives from a unique underlying distribution with  $p \lesssim 10^{-7}$  and  $p \lesssim 0.05$  for the near and far sample, respectively. We mainly attribute the different shapes of the two distributions to the different degrees of fragmentation of their clump population. The near sample is dominated by highly fragmented clumps with  $N_{cores} > 8$ , which strongly impact the shape of the distribution since their separations contribute for more than  $\sim 80\%$  of the total. Conversely, the far sample consists mainly of poorly fragmented clumps ( $N_{cores} \leq 8$ ), whose separations are a relevant ( $\sim 65\%$ ) fraction of the distribution presented in Fig. 10. In summary, the different shapes are produced by the different relative contributions of poorly and highly fragmented systems combined with the evidence that separations are typically larger in clumps with  $N_{cores} \leq 8$  than in those with a higher degree of fragmentation. A minor, secondary effect may be caused by the larger physical area observed in the far group, which favors the measurements of wider separations.

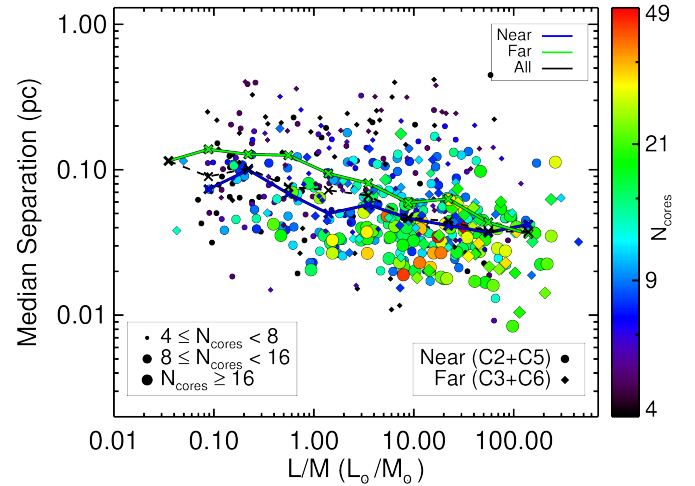
Only a few clumps with  $1 < L/M < 10 L_{\odot}/M_{\odot}$  were observed before ALMAGAL, a large fraction of them were obtained by the SQUALO (Traficante et al. 2023) and the DIHCA survey (Ishihara et al. 2024). A typical separation of  $\sim 7\,000$  au is reported from the 30 clumps of the DIHCA survey, although this includes systems with higher  $L/M$ , since their sample covers  $\sim 1$  to  $50 L_{\odot}/M_{\odot}$ . The average separation and, in general, the entire distribution closely match the results from the more evolved clump populations, similar to what is found for the near sample of the ALMAGAL survey. In contrast, the comparison with the SQUALO survey is severely limited since only 3 clumps (out of 13 surveyed in total) fall within this interval of  $L/M$  (Traficante et al. 2023). Traficante et al. (2023) reported a low level of fragmentation, measuring minimum separations between  $13\,000$  and  $29\,000$  au. Although these values should be considered as a lower limit because of the projection effect, these wide separations are nonetheless comparable with our results from the far sample systems.

In summary, our statistical analysis reveals an evolutionary trend in core separation. When ALMAGAL clumps are grouped by their  $L/M$  ratio, we find that the cores in more evolved systems are generally closer to one another than those in younger systems. Evidence of a relationship between core separations and the evolutionary diagnostic  $L/M$  was already proposed by Traficante et al. (2023), who compared the separations in the 13 young and moderately evolved targets of their SQUALO survey to those of the 20 high-mass and high-luminosity clumps of the CORE survey (Beuther et al. 2018). Similar indications were also found by including the 30 moderately evolved systems ( $\sim 1 < L/M < 50 L_{\odot}/M_{\odot}$ ) of the DIHCA survey (Ishihara et al. 2024) and comparing the 11 massive, luminous, and evolved clumps of the ASSEMBLE survey (Xu et al. 2024) with the 39 high-mass young clumps observed by the ASHES survey (Sanhueza et al. 2019; Morii et al. 2024).

The sample size and the uniform coverage across the  $L/M$  parameter space of the ALMAGAL survey provide robust statistical significance to this trend. It allows us to move beyond a coarse three-group comparison and analyze in detail the variation as a continuous function of  $L/M$ . The right panels of Fig. 10 show the entire ensemble of separations as a function of the clump  $L/M$  ratio. Although the separations  $l_i^{depr}$  have a

wide scatter for any given ratio  $L/M$ , we find a steady decreasing trend traced by the median of the distributions computed in equally spaced bins in logarithm of  $L/M$ . This behavior is not caused by the weak correlation found with distance. In fact, it is still present when selecting systems in the two distance intervals  $3.7 \leq d_{cl} \leq 4.7$  kpc and  $4.7 \leq d_{cl} \leq 6.5$  kpc, for which the statistical values of the separations do not show any dependence on distance (see Sect. 3.2.2), despite the minor quantitative differences emerging due to the smaller sample sizes. Apparently, the two samples near and far show different trends. In the near sample, we found that the median has a major decrease from  $\sim 15\,000$  to  $\sim 10\,000$  au occurring until  $L/M \approx 1 L_{\odot}/M_{\odot}$ , followed by a minor decrease to  $7\,000$ – $8\,000$  au for increasing values up  $L/M \approx 100 L_{\odot}/M_{\odot}$ . In the far sample, the median is typically larger in the early phases, with a value of  $\sim 20\,000$  au, and shows a steeper and more steady decrease starting at  $L/M \approx 1 L_{\odot}/M_{\odot}$ . These differences are due to the smaller sample size, lower level of fragmentation, and wider observed area in clumps of the far sample, as discussed above. The shift of the median to larger values is caused by the increase in the statistical weight of clumps with a limited degree of fragmentation whose cores are more distantly separated, as discussed above. It is interesting to compare the trend traced by the mode of the distributions. The right panel of Fig. 10 shows that in the two samples the modes are coincident for  $L/M \gtrsim 1 L_{\odot}/M_{\odot}$ . Instead, we observe a match between the values of the near sample and the shapes of the 2D density contours for  $L/M < 1 L_{\odot}/M_{\odot}$ . These shapes suggest a potential bimodality in the distribution, with one of the peaks that follows the trend found in the near sample. More generally, the distributions of the separations in the far sample are all broad for clumps with  $L/M \leq 10 L_{\odot}/M_{\odot}$ . Moreover, the profiles of the KDE distributions shown in the left panel of Fig. 10 present two local maxima, one approximately at  $\sim 40\,000$  au and another that shifts from  $\sim 10\,000$  down to  $\sim 7\,000$  au. The maximum at large separations is built mainly, but not exclusively, by systems with a low degree of fragmentation, whose cores are more widely distributed. In fact, about  $\sim 72\%$  of the separations greater than  $\sim 20\,000$  au of the far sample are measured in clumps with  $N_{cores} \leq 8$ . To verify that these maxima correspond to a statistically significant bimodality, we ran the Silverman (1981) test, following the implementation described in Beraldo e Silva et al. (2021, see their Appendix A), applied to the distributions in different intervals of  $L/M$ . In a few intervals of  $L/M$ , we were able to assess that the corresponding distribution of separations has more than one significant peak. However, we were unable to extend this to the entire  $L/M$  range due to the limited sample size in a few intervals of  $L/M$ . In general, we found that the null hypothesis that at least two peaks are present can be assessed only with a probability  $Prob > 0.8$ , although it becomes significant ( $Prob > 0.95$ ) over limited intervals of  $L/M$ .

Finally, we present in Fig. 11 the same analysis using clump-averaged separations, showing only the median separation as a function of the  $L/M$  ratio since the mode has a similar trend. Similarly to the case of the entire ensemble, the averaged quantities also show a gradual decrease in the separation between the cores with  $L/M$ , as we determined a Spearman correlation coefficient equal to  $\rho_s = -0.37$  [ $-0.38$ ] for the median [mode] with a  $p$ -value  $\ll 10^{-9}$ . These values indicate that a correlation is indeed present; however, the Pearson correlation coefficient of  $\rho_p = -0.26$  [ $-0.28$ ] suggests that it is not a single-linear relation over the entire interval of  $L/M$ . Figure 11 also indicates that the observed relationship is connected to the degree of fragmentation, which also statistically increases with the  $L/M$  ratio (see the



**Fig. 11.** Relation between the clump-averaged separations between the cores provided by the median and the  $L/M$  ratio, color-coded to indicate the number of cores detected in the clump. To support the visualization, different sizes are adopted to identify systems  $4 < N_{cores} < 8$  (low),  $8 \leq N_{cores} < 16$  (intermediate), and  $N_{cores} \geq 16$  (high). The different symbols refer to the clumps of the near (circle) and far (diamond) sample. The solid lines show the median of the distribution of the separations computed in bins of  $L/M$  of width 0.4 dex for the near (blue) and far (green) sample, respectively, while the black dashed line corresponds to the median of all the clumps with  $N_{cores} \geq 4$ .

discussion in Elia et al. 2026). As discussed for the case of the entire ensemble, the relationship changes for  $L/M \sim 1 L_{\odot}/M_{\odot}$ , above which it becomes rather flat. This value sets a boundary above which  $\sim 90\%$  of the fields are intermediate or clustered, according to the classification introduced in Sect. 3.1, or, alternatively, have a moderate to high degree of fragmentation, as they mostly host more than eight cores.

### 3.3.2. Relation with clump surface and volume density

The ALMAGAL clumps cover systems with surface densities,  $\Sigma_{cl}$ , varying from  $\sim 0.1$  to  $\sim 10$  g cm $^{-2}$ , which allow us to determine whether there is a connection between the places where the material collapsed and the averaged density of the clump. Density is a key quantity in the fragmentation theory and, in the case of Jeans fragmentation, sets the scales that ultimately shape the separation between cores. Figure 12 shows the distribution of all the separations as a function of the clump surface,  $\Sigma_{cl}$ , and volume number densities,  $n_{cl}$  for the near and far samples. The clump-averaged  $\Sigma_{cl}$  and  $n_{cl}$  are computed with the assumption of spherical geometry for the clump as

$$\Sigma_{cl} = \frac{M_{cl}}{\pi R_{cl}^2} \quad \text{and} \quad n_{cl} = \frac{3M_{cl}}{4\pi \mu m_H R_{cl}^3}, \quad (2)$$

where  $M_{cl}$  and  $R_{cl}$  are the clump mass and radius, respectively,  $m_H$  is the atomic mass and  $\mu$  is the mean molecular weight per free particle adopted to be equal to 2.37, which corresponds to a molecular gas with solar composition<sup>4</sup> (Kauffmann et al. 2008).

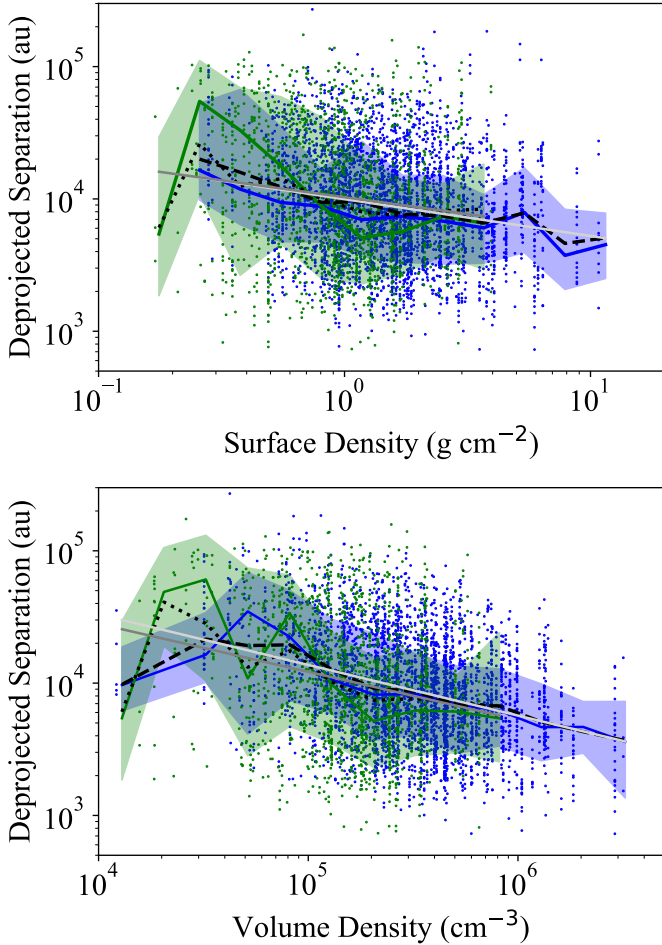
Quantitatively, the typical separation between the cores steadily decreases from  $\sim 25\,000$  au to  $\sim 8\,000$  au in clumps with an increasing  $\Sigma_{cl}$  from  $\sim 0.3$  g cm $^{-2}$  to  $\sim 5$  g cm $^{-2}$ , and a drop to

<sup>4</sup> This value of  $\mu$  is derived for a mixture composed only by  $H_2$  and He with mass ratio equal to the solar one, i.e.,  $X \approx 0.71$  and  $Y \approx 0.27$  (see Appendix A in Kauffmann et al. 2008).

**Table 5.** Results of the fits of a power-law relationship between separations and surface and volume density.

Sample	$A^a$		$B^a$		$C^b$		$D^b$	
	Median <sup>c</sup>	All data <sup>d</sup>	Median <sup>c</sup>	All data <sup>d</sup>	Median <sup>c</sup>	All data <sup>d</sup>	Median <sup>c</sup>	All data <sup>d</sup>
near	4.02±0.02	4.00±0.01	-0.34±0.04	-0.28±0.02	5.54±0.02	6.05±0.10	-0.29±0.06	-0.38±0.02
far	4.01±0.01	3.99±0.06	-0.17±0.15	-0.25±0.05	5.40±0.70	5.90±0.21	-0.27±0.14	-0.36±0.04

**Notes.** (a) Fitted relationship  $\log l^{depr} = A + B \log \Sigma_{cl}$ . (b) Fitted relationship  $\log l^{depr} = C + D \log n_{cl}$ . (c) Fit performed to the median values of separations calculated in bins of width 0.3 dex. (d) Fit performed using the entire set of separations.



**Fig. 12.** Deprojected core separations against the clump surface (*top panel*) and volume (*bottom panel*) density. The distributions of the core separations are computed in bins of  $\Sigma_{cl}$  and  $n_{cl}$  of width 0.15 and 0.2 dex, respectively. The peaks and spreads of these distributions, defined as the width at 68% peak intensity, are indicated by lines and area for the near (blue) and the far (green) samples, respectively. The medians of the distributions for these two sets are indicated as dashed and dotted black lines, respectively. The best fitting power-law relation computed over the entire set of separations and whose results are reported in Table 5 are indicated as light gray and dark gray lines for the near and far sample, respectively.

$\sim 6000$  au for  $\Sigma_{cl} \geq 5 \text{ g cm}^{-2}$ , where the relation appears to be flattening. A decreasing trend is also found when the separations are compared to the clump volume density  $n_{cl}$ , with a substantial decrease between  $\sim 10^4$  and  $\sim 10^6 \text{ cm}^{-3}$  (see the lower panel of Fig. 12). Both relations show a wide scatter of the measured separations, and weak linear relations appear only from the statistical estimators, i.e., the median and the mode. We checked

for possible power-law relations between  $l^{depr}$  versus  $\Sigma_{cl}$  and  $n_{cl}$  by fitting to the data with the functional forms  $l^{depr} \propto \Sigma_{cl}^\alpha$  and  $l^{depr} \propto n_{cl}^\beta$ . The derived exponents of these fits are reported in Table 5 and indicate that a weak correlation exists between the separations and these quantities. The empirical trends traced by the medians closely match the result of the fits derived using all the separations in each of the two samples and show a tight power-law behavior for  $n_{cl} \geq 10^5 \text{ cm}^{-3}$ .

### 3.4. Comparison with the Jeans fragmentation

In this section, we compare the observed distribution of core separations with the expectations of the Jeans fragmentation theory for a non-magnetized, homogeneous, isothermal and self-gravitating clump. Under these conditions, there is a critical mass (and a corresponding length) above which the medium is not able to remain stable and therefore collapses (Jeans 1902). These critical values depend only on the density,  $n$ , and temperature,  $T$  (see Appendix B). Although these conditions refer to an idealized case which may not be representative of real star-forming clumps, the analytical prescription given by Jeans theory offers a simplified model against which the observed fragmentation in the ALMAGAL clumps can be compared. Specifically, we are looking for whether thermal fragmentation is sufficient to statistically describe the separations between cores or whether additional contributions, such as turbulent support, are needed. Since core masses are affected by higher uncertainties than core separations, we focused our analysis on the comparison between the latter and the thermal Jeans lengths, while we report the results for the masses in Appendix C.

#### 3.4.1. The Jeans length and mass for ALMAGAL clumps

We computed the thermal Jeans length,  $\lambda_J^{th}$ , and mass,  $M_J^{th}$ , for the ALMAGAL clumps with Eq. (B.2) and Eq. (B.3), adopting as temperature the one of the dust,  $T_d$ , derived from *Herschel*/Hi-GAL data and reported in Table A.1 in Molinari et al. (2025). At the typical densities of ALMAGAL clumps ( $n > 10^4 \text{ cm}^{-3}$ ), gas and dust are expected to be thermally coupled (Goldsmith 2001). This has been observationally confirmed, with relatively small uncertainties discussed below, in a sample of 1086 Hi-GAL clumps (Merello et al. 2019). In these clumps, on average dust temperatures were found to be consistent with temperatures measured from  $\text{NH}_3$  emission (Merello et al. 2019), considered a reliable temperature tracer for the gas phase (Juvela et al. 2012; Guzmán et al. 2015). The values of  $\lambda_J^{th}$  and  $M_J^{th}$  for each ALMAGAL clump are reported in Table 3. The thermal Jeans length ranges from  $\sim 3500$  to  $\sim 60\,000$  au ( $\sim 0.017$ – $0.3$  pc), with a median value of  $\sim 14\,500$  au ( $\sim 0.07$  pc), and Jean masses from  $\sim 0.2$  to  $14 M_\odot$ , with a median equal to  $0.89 M_\odot$ .

The uncertainties of these quantities are determined by those on the temperature and density, which depend on both their direct measurements and on the assumptions adopted for their estimates. The fit to the *Herschel* data typically provides small uncertainties for  $T_d$ , quantified as  $\sim 2$  K (Elia et al. 2017; Svoboda et al. 2019). This value translates into a different contribution for the young and evolved systems of the ALMAGAL sample, characterized by a typical temperature of  $\approx 10$  and  $\approx 40$  K, respectively. The uncertainties in  $\lambda_j^{th}$  and  $M_j^{th}$  vary from 2% to 10% and 6% to 30%, with the highest values for younger systems. Although the dust and gas temperature are found to be in agreement on average, their relationship presents a measurable scatter whose amplitude quantifies how well we can assess the thermal coupling between these two phases. Typical discrepancies between  $T_g$  and  $T_d$  are of the order  $\sim 25\%$  (see Fig. 2 in Merello et al. (2019)), where the distribution of  $\log(T_g/T_d)$  has a width of  $\sim 0.2$  dex. These discrepancies introduce an additional random contribution of  $\sim 13\%$  and  $\sim 38\%$  to the uncertainties in  $\lambda_j^{th}$  and  $M_j^{th}$ , respectively. In addition to the random scatter,  $T_d$  and  $T_g$  show a systematic offset in different stages of the clump evolution, which may lead to an underestimation or overestimation of  $\lambda_j^{th}$  and  $M_j^{th}$ . Ammonia-derived gas temperatures are found to be systematically  $\sim 20\%$  higher than  $T_d$  in young systems such as prestellar clumps (Merello et al. 2019), where star formation activity is absent, while no significant offset is found in evolved systems. However, those evolved phases also have other temperature diagnostics as additional molecular tracers become available. The gas temperatures  $T_g$  estimated from these diagnostics are generally higher than  $T_d$  derived from the fitting of the spectral energy distribution of the *Herschel* data (Molinari et al. 2016b; Lin et al. 2022). Specifically, Molinari et al. (2016b) estimated gas temperatures from  $\text{CH}_3\text{C}_2\text{H}(2-1)$  emission finding that they are about  $\sim 40\%$  higher than dust ones. Similar discrepancies were also found by Lin et al. (2022), who instead analyzed the  $\text{CH}_3\text{CN}$  and  $\text{CH}_3\text{OH}$  emission lines in a few clumps with  $L/M \geq 10 L_\odot/M_\odot$ , although they also report extreme cases associated with HII regions, where estimated  $T_g$  are as high as  $\sim 130$  K which is a few times higher than  $T_d$ .

The discrepancy between gas and dust temperature may be explained by different reasons. Although fitting the dust continuum provides an accurate determination of  $T_d$ , this emission may include contributions from the foreground or background. In addition, it is not always clear whether the molecular line emission probes the entire clump or only a fraction of it (Mininni et al. 2025), in particular in more evolved systems where temperature gradients are certainly set by the hosted protostars. However, the systematic difference between  $T_d$  and  $T_g$  indicates that the clump-averaged  $\lambda_j^{th}$  and  $M_j^{th}$  may be different from our estimates in different evolutionary stages. Nonetheless, the expected average increase of these quantities does not exceed  $\sim 20\%$  and  $\sim 40\%$  for  $\lambda_j^{th}$  and  $M_j^{th}$ , respectively. At the same time, we consider an additional random contribution of  $\sim 10\%$  and  $\sim 30\%$  to the uncertainties in  $\lambda_j^{th}$  and  $M_j^{th}$  induced by the typical scatter of  $\sim 20\%$  between the two estimates  $T_g$  and  $T_d$ .

Unlike the case of temperature, the average volume density  $n_c$  is not directly measured, but is typically inferred from measurements of the clump mass and radius. Assuming an uncertainty of 50% in  $M_{cl}$ , 20% in the measured source size and 15% in the clump distance, the uncertainty in  $n_{cl}$  is about  $\sim 60\%$ , which results in  $\sim 30\%$  for  $\lambda_j^{th}$  and  $M_j^{th}$ . These estimates are obtained by assuming a uniform distribution of the material, which may not be realistic due to the density stratification present in clumps (Beuther et al. 2002; Williams et al. 2005; Palau et al. 2014; Beuther et al. 2024). In these systems, the density

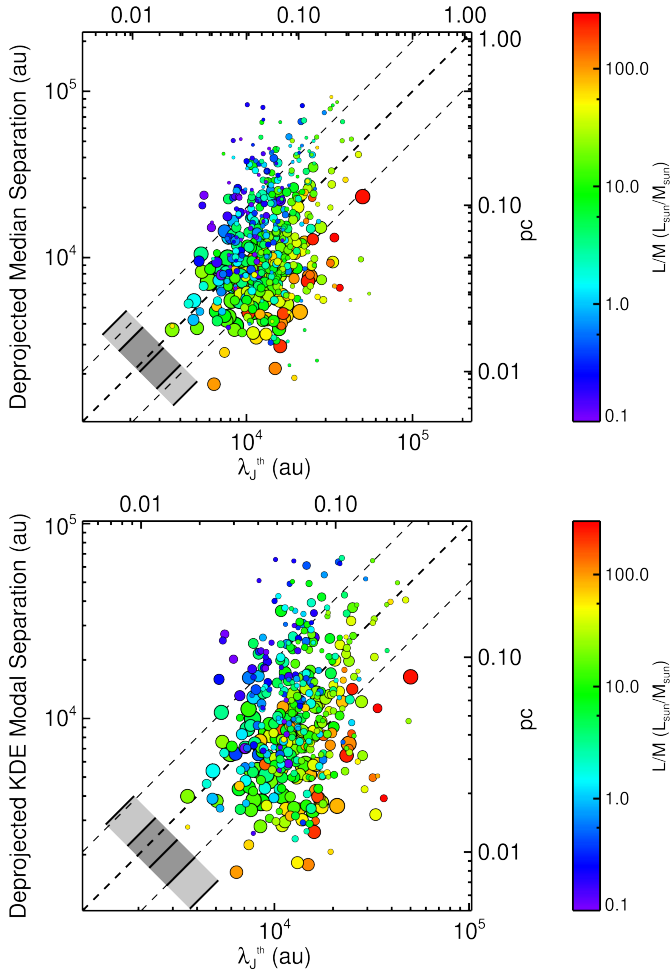
profile has a typical power-law shape, i.e.,  $\rho = \rho_0 \left(\frac{R}{R_0}\right)^{-p}$  with  $\rho_0$  and  $R_0$  the density at the reference radius  $R_0$ . The local density in the inner region of size  $R'$  is therefore enhanced by a factor  $\left(\frac{R'}{R_0}\right)^{-p}$ , which depends on the value of  $p$ . Observations of high-mass clumps indicate that  $p$  varies in the interval  $1 < p < 2$ , with an average value of  $p \approx 1.6$  (Beuther et al. 2002; Williams et al. 2005; Palau et al. 2014; Beuther et al. 2024, see also the discussion in Beuther et al. 2025 for a summary of the recent measurements). The observed variation appears to be connected to the clump evolutionary state. In fact, Lin et al. (2022) reported an increase from  $p \approx 1$  to  $\sim 1.5$  in their sample, and Gieser et al. (2022) found slightly shallower profiles with  $p \approx 1.6$  in the younger regions, whereas they also measured  $p = 2.0 \pm 0.2$  in the evolved massive clumps of the CORE surveys (Gieser et al. 2021; Beuther et al. 2024).

In principle, the density profile and the exponent  $p$  could be constrained for each clump by fitting the Hi-GAL observations. However, this would require a careful and detailed radiative transfer modeling to properly account for the temperature structure, which is beyond the scope of this work. Our aim here is to quantify the amplitude of the uncertainty that the assumed density structure introduces into our estimates for Jeans properties. In the case of ALMAGAL clumps, fragmentation is effectively taking place in a region defined by the cluster radius,  $R_{cluster}^{ph}$  (see Sect. 3.1 and Table 2). Due to density stratification, the average density is higher in the central regions, which results in a decrease for  $\lambda_j^{th}$  and  $M_j^{th}$  by a factor  $C^{\frac{1}{2}}$ , where  $C = \left(\frac{R_{cluster}^{ph}}{R_c}\right)$ . The median value of  $C$  is  $\sim 0.46$ , with quartiles of  $\sim 0.36$  and  $\sim 0.59$ , which translate into an expected  $\lambda_j^{th} = 0.68_{-0.1}^{+0.09} \times \lambda_j^{th}$ , for the young systems with  $p \approx 1$ , and  $\lambda_j^{th} = 0.46_{-0.1}^{+0.13} \times \lambda_j^{th}$  for evolved systems where  $p \approx 2$ .

In general, our assumptions in  $T_g$  and  $n_c$  introduce systematic shifts in the average estimates of  $\lambda_j^{th}$  and  $M_j^{th}$ , that go in opposite directions. The amplitudes of these shifts are similar, and, on average, compensate; therefore, we did not apply any correction. However, since their contribution leads to an increase in the spread of the uncertainty, in our analysis we considered a typical random error on  $\lambda_j^{th}$  of at least  $\sim 50\%$ .

### 3.4.2. ALMAGAL core separations compared to Jeans length

Figure 13 shows the comparison of the two clump-averaged estimators for the core separation, the median  $l^{median}$  and the mode  $l^{mode}$  (see Sect. 3.2.2), with the estimated thermal Jeans length,  $\lambda_j^{th}$ . In these plots, we mark the level of fragmentation and  $L/M$  of the clump with the symbol size and the color, respectively. A large fraction of ALMAGAL clumps have core separations that are consistent with the thermal Jeans length  $\lambda_j^{th}$ . The mode [median] separations range from  $\sim 0.07$  [0.1] $\times$  to  $\sim 6.5$  [11.0] $\times$   $\lambda_j^{th}$ . The distributions of the ratio  $l^{median}/\lambda_j^{th}$  and  $l^{mode}/\lambda_j^{th}$  are centered on values close to 1 and cover the narrow intervals included in Fig. 13 with dark and light gray shaded areas representing the interquartile and the 10th to 90th percentile interval, respectively. Quantitatively, the median and the mode are centered around  $0.91\times$  and  $0.76\times \lambda_j^{th}$ , respectively. The first and third quartiles are equal to  $0.56\times - 1.60\times$  and  $0.5\times - 1.26\times \lambda_j^{th}$  for the median and the mode, respectively. The 10th-90th percentile ranges extend from  $0.29\times$  to  $2.10\times$  for the mode and  $0.36\times$  to  $2.53\times \lambda_j^{th}$  for the median. Both the clump-averaged median and mode reveal a similar behaviour with the thermal Jeans length.



**Fig. 13.** Comparison between the clump-averaged estimators for the core separations, specifically the median (*top panel*) and mode (*bottom panel*) of the distribution in each clump, and the computed Jeans length. The fragmentation level in the clump is indicated by the symbol sizes, with increasing size going from low ( $4 \leq N_{\text{cores}} < 8$ ) and intermediate ( $8 \leq N_{\text{cores}} < 16$ ), up to high ( $N_{\text{cores}} \geq 16$ ) core numbers. The dashed lines are drawn as a reference and correspond to the separation equal to  $\lambda_J^{\text{th}}$  (thick line) and  $2\times$  and  $0.5\times$   $\lambda_J^{\text{th}}$  (thin lines), respectively. The gray shaded boxes indicate the amplitude of the distributions of the ratio  $l_i^{\text{depr}}/\lambda_J^{\text{th}}$  (see text), and correspond to the interquartile range (dark gray) and the 10th to 90th percentile range (light gray), respectively.

Although on average the core separations are in agreement with the thermal Jeans length, we observe that the systems with different  $L/M$  are distributed in different regions of Fig. 13, indicating that the ratio  $R = l^{\text{depr}}/\lambda_J^{\text{th}}$  varies with  $L/M$  and, therefore, changes during the clump evolution. The trend is shown in Fig. 14, where we present a panoramic view of how the ratio  $R$  varies with  $L/M$  for all separations of the ensemble. We found a decreasing trend which is the result of two effects: the decrease in the core separations  $l^{\text{depr}}$  discussed in Sect. 3.3.1, and a steady increase in  $\lambda_J^{\text{th}}$  with  $L/M$ . We note that the increase in  $\lambda_J^{\text{th}}$  is mainly driven by the general warming of the clump material by the radiation of newly born stars. Although it also depends on density  $n_{\text{cl}}$ , this has a marginal impact when averaging over bins of  $L/M$  ratio. In fact, our sample includes clumps with density spanning over the entire range of values in each of those bins (see Fig. 1). More generally, the selected ALMAGAL clumps do not show an increase in their density with evolution (see Sect. 2.1).

Given the uncertainties in  $\lambda_J^{\text{th}}$ , our result is that the observed separations are consistent with the thermal Jeans length computed with the average properties derived from *Herschel* observations for all clumps with the  $L/M$  ratio between 1 and  $100 L_{\odot}/M_{\odot}$ . Figure 14 shows an average systematic decrease in  $R$  with  $L/M$  and that the core separations become significantly smaller than the thermal Jeans lengths in the evolved systems with  $L/M \gg 50 L_{\odot}/M_{\odot}$  ( $L/M \gg 100 L_{\odot}/M_{\odot}$  in the case of near sample). This trend for  $L/M \gtrsim 1 L_{\odot}/M_{\odot}$  is mainly driven by an increase in  $\lambda_J^{\text{th}}$  and not by a decrease in the core separations. The observation of sub-Jeans separations in these evolved stages does not rule out gravity-dominated thermal fragmentation. In these phases, protostellar feedback could create local density enhancements that exceed the mean clump density used for our  $\lambda_J^{\text{th}}$  calculation. Therefore, Jeans-like fragmentation of these locally denser regions would naturally produce cores with separations smaller than our clump averaged estimates.

It is interesting to look in detail at the distribution of  $R$  for young and poorly fragmented clumps with  $L/M < 0.5$ – $1 L_{\odot}/M_{\odot}$ . The distributions have a median  $R \sim 1.5$ , but the shape of the density contours suggests a bimodality, which we confirmed with the statistical tests described in Sect. 3.3.1. Although the two peaks for the far sample systems are connected to the bimodal distribution of  $l_i$ , they are also present in the near systems. One of the two peaks of the distributions is centered at  $R \sim 1$ , indicating a population of cores which are separated by a thermal Jeans length, while the second peak is shifted at higher values, which are centered at about  $R \sim 3$ .

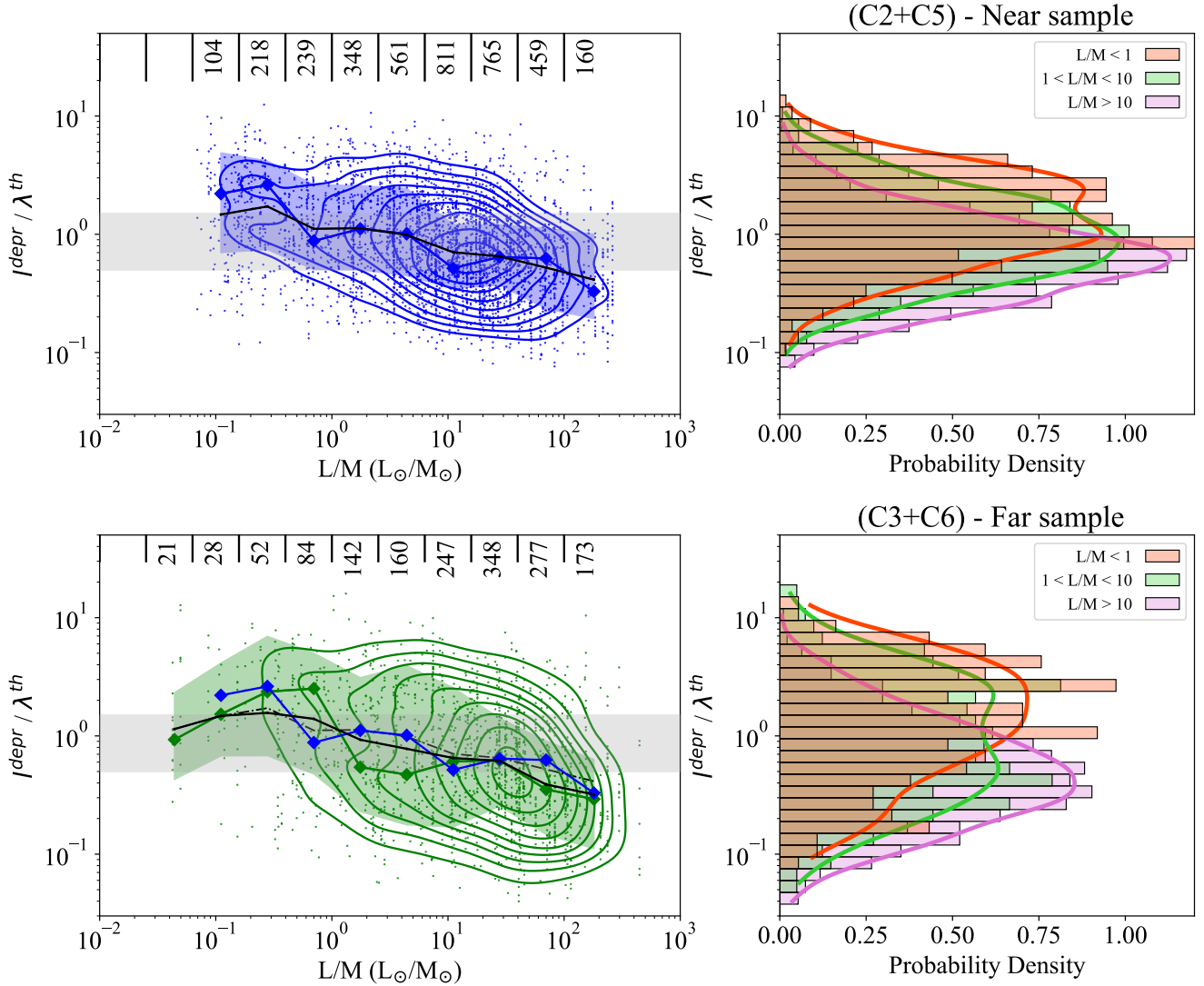
### 3.5. How are cores distributed in the fields?

In previous sections, we analyzed where cores are located in star-forming clumps and their relative separations to discuss how fragmentation takes place in potentially cluster-forming systems. How cores are distributed in the systems and the identification of potential different mass-dependent clustering of the cores offer additional information on the cluster formation process. To such aim, we analyzed two additional diagnostics, the  $Q$  parameter and the mass segregation ratio,  $\Lambda_{\text{MSR}}$ , which are typically used to characterize young proto-clusters (Allison et al. 2009; Cartwright & Whitworth 2004; Parker 2018; Dib & Basu 2018), and were also recently estimated for groups of cores (Dib & Henning 2019; Sanhueza et al. 2019; Sadaghiani et al. 2020; Avison et al. 2023; Morii et al. 2024; Xu et al. 2024).

#### 3.5.1. The $Q$ parameter

Statistical analysis provides several diagnostics capable to characterize the spatial distribution of young stellar groups. Cartwright & Whitworth (2004) introduced the  $Q$  parameter as a tool to quantify the degree of substructures in the distribution of the members of young stellar clusters and, if present, to identify cases where their radial density profiles are centrally concentrated (Cartwright & Whitworth 2004; Parker & Goodwin 2015; Jaffa et al. 2017). The  $Q$  parameter is defined as the ratio of the mean edge length of the cluster MST,  $\bar{m}$ , and the correlation length of the cluster,  $\bar{s}$ , both with normalization factors defined to remove any dependence on the number of members and the size of the cluster. In formula these are

$$\bar{m} = \sum_{i=1}^{N_{\text{c}}-1} \frac{l_i}{\sqrt{N_{\text{cores}} \pi R_{\text{cluster}}^2}}, \quad (3)$$



**Fig. 14.** *Left panel:* comparison of the ratio  $R$  of the deprojected separations to  $\lambda_j^{\text{th}}$  and the clump  $L/M$  ratio for the near (*top panel*) and far sample (*bottom panel*), respectively. The black dashed line shows the median of the distributions of  $R$  computed in bins of  $L/M$  equally spaced in logarithm of width 0.4 dex. The thick colored lines and the colored area trace the mode and the spread (defined as the width at 68% of the peak) of these distributions, adopting the blue and green colors for the near and far sample, respectively. The numbers at the top give the size of the subsamples adopted to compute the distribution in each bin. The gray shaded area shows the interval  $0.5 \leq R \leq 1.5$ , corresponding to an uncertainty on the  $\lambda_j^{\text{th}}$  of a factor of 2. The colored contours, from the outside inward, fractions of the sample lying outside that rise from 15% to 95% with a step of 10% for the near (in blue) and the far (in green) subsample, respectively. *Right panel:* normalized density histogram and estimated probability density distribution estimated with the KDE method (colored continuous lines) of the ratio  $R = l_i/\lambda_j^{\text{th}}$  measured in the three subsamples selected by different intervals of  $L/M$  tracing clumps in different evolutionary phases: young ( $L/M \leq 1 L_\odot/M_\odot$  in orange), intermediate ( $1 \leq L/M \leq 10 L_\odot/M_\odot$  in green), and evolved ( $L/M \geq 10 L_\odot/M_\odot$  in purple).

where  $l_i$  are the MST edges and  $R_{\text{cluster}}$  is the size of the cluster defined as in Sect. 3.1.1 and

$$\bar{s} = \frac{\bar{l}_i}{R_{\text{cluster}}}, \quad (4)$$

with  $\bar{l}_i$  defined as the average separation between the cluster members. The ratio  $Q = \bar{m}/\bar{s}$  allows us to distinguish between two specific spatial configurations, one where the members are radially distributed with a central concentration and the other where they have fractal type of subclustering. Cartwright & Whitworth (2004) showed that systems with members distributed in smooth and centrally concentrated patterns have  $Q > 0.8$ , with values that increase for higher concentrations represented by higher exponents  $p$  in the number density profile

$N(r) \propto r^{-p}$ . In contrast, systems whose members are organized in a fractal type of subclustering, described with a fractal dimension  $D$ , have  $Q < 0.8$ , and smaller values of  $Q$  correspond to systems with a higher level of subclustering (lower values of  $D$ ).

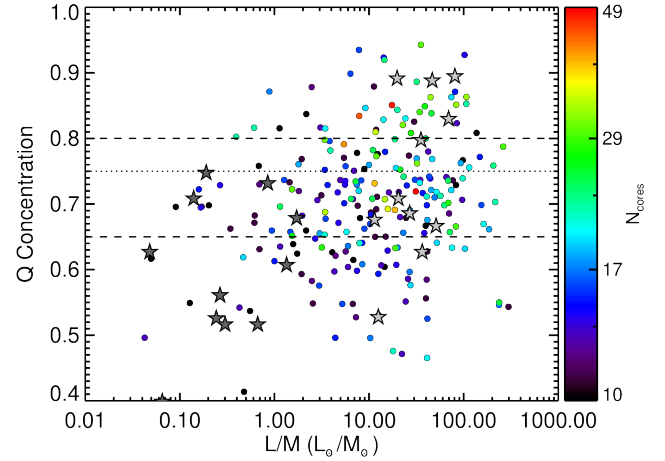
Although this diagnostic is typically used in the statistical analysis of young star clusters (Parker 2014; Gregorio-Hetem et al. 2015; Dib et al. 2018), it has only recently been applied to the distribution of star-forming cores, finding, for example, that  $Q$  correlates with star formation activity (Dib & Henning 2019). Specifically,  $Q$  increases with the density of star formation rate in low-mass star-forming clouds (Dib & Henning 2019). Moreover, the comparison between the core population observed by the ASHES (Sanhueza et al. 2019; Morii et al. 2024) and ASSEMBLE surveys (Xu et al. 2024) shows that different spatial arrangements are found depending on the evolutionary stage of

the clump. Most of the ASHES clumps ( $\sim 70\%$ ), that are young and with  $L/M \lesssim 2L_{\odot}/M_{\odot}$ , have  $Q \lesssim 0.8$ , which characterizes subclustered systems. In contrast, higher values of  $Q$  are typically measured in the more evolved clumps of the ASSEMBLE survey that have  $L/M \gtrsim 10L_{\odot}/M_{\odot}$  (Xu et al. 2024). The presence of a weak correlation between the  $Q$  parameter and the ratio  $L/M$  has been proposed to reflect a time evolution effect, where the signature of a gradual transition from a subclustered pattern to a centrally condensed configuration is a consequence of gravitational collapse (Xu et al. 2024). We tested the statistical robustness of the correlation  $Q$  versus  $L/M$  with the ALMAGAL sample, which largely extends the sample size of the ASHES and ASSEMBLE surveys from which it was initially derived.

The  $Q$  parameter measured in the ALMAGAL clumps with  $N_{\text{cores}} \geq 4$  shows values ranging from 0.4 to 0.95, with a median equal to 0.65, and half of the systems having  $0.6 < Q < 0.75$ . These values would suggest that the ALMAGAL sample is composed mainly of clumps where the cores are distributed in subclusters, if the classification criteria introduced by Cartwright & Whitworth (2004) were applied. However, recently Avison et al. (2023) questioned the diagnostic usefulness of the  $Q$  parameter when applied to a small number of sources ( $N_{\text{cores}} < 10$ ). They simulated clusters composed of small numbers of members ( $N_{\text{cores}} = 5$  and 10) to characterize the observations of the TEMPO survey, where they found between 2 and 15 fragments (Avison et al. 2023). Most of the ALMAGAL clumps with these conditions ( $N_{\text{cores}} < 10$ ) would be classified as highly subclustered as  $\sim 80\%$  of them have  $Q \leq 0.6$ . To quantitatively assess whether the  $Q$  parameter can be used as a useful diagnostic in the ALMAGAL fields that have a wider interval of the number of cores than TEMPO survey ( $N_{\text{cores}}$  from 2 to 49), we extended the analysis of Avison et al. (2023). This analysis is presented in Appendix D and is also applicable to the ASHES and ASSEMBLE fields, which host 8 to 37 and 13 to 37 fragments, respectively (Morii et al. 2024; Xu et al. 2024).

Our result is that the  $Q$  parameter retains a diagnostic power in probabilistic terms only for systems with  $N_{\text{cores}} \geq 10$ . We defined two intervals of  $Q$  where it is most likely that the cores follow a centrally concentrated ( $Q \gtrsim 0.8$  for high power-law exponent  $p$ ) or subclustered distribution ( $Q \lesssim 0.65$ , lower fractal dimension), with a value of  $Q = 0.75$  that approximately separates these two regimes. These results revise the initial impression on the ALMAGAL systems and indicate that both spatial distribution patterns are present in the 226 ALMAGAL systems with  $N_{\text{cores}} \geq 10$ . In fact, adopting the tentative limit of  $Q = 0.75$ , there are 155 fields ( $\sim 68\%$ ) with  $Q < 0.75$  and 71 fields ( $\sim 32\%$ ) with  $Q \geq 0.75$ , and similar proportions are found if higher cuts on  $N_{\text{cores}}$  are applied. However, our simulations show that the statistical distributions of  $Q$  for the two classes are not strongly separated, but they instead overlap. This led us to assess the classification with sufficient reliability only for the 60 fields where  $Q < 0.65$  and for the 38 where  $Q > 0.8$ . Although we reduce the sample size with this choice, we were still able to confirm that both patterns are observed in massive dense clumps.

We show in Fig. 15 the values of the  $Q$  parameter as a function of the clump  $L/M$  ratio for the 226 ALMAGAL fields with  $N_{\text{cores}} \geq 10$ , and we included measurements from the ASHES pilot (Sanhueza et al. 2019) and the ASSEMBLE survey (Xu et al. 2024). The 39 fields with  $L/M \leq 2L_{\odot}/M_{\odot}$  expand the sample provided by the ASHES survey for the early evolutionary phases (Sanhueza et al. 2019; Morii et al. 2024), and indicate that these early phases have cores that are mainly distributed in fractal-type subclusters, with 13 fields showing  $Q < 0.65$ . Fields with a clear signature of a concentrated distribution are



**Fig. 15.** Relation between the  $Q$  parameter and the  $L/M$  ratio of ALMAGAL clumps which host at least ten cores. The two dashed lines values  $Q = 0.65$  and  $Q = 0.8$  defining the intervals for which it is most likely that the hosted cores follow a fractal type of subclustering ( $Q \leq 0.65$ ) and radial distribution with a central concentration ( $Q \geq 0.8$ ), as indicated by the Monte Carlo simulations in Appendix D. The dotted line points to the  $Q = 0.75$ , which separates the two regimes with a lower statistical significance. Estimates from the ASHES (dark gray; Sanhueza et al. 2019) and ASSEMBLE surveys (light gray; Xu et al. 2024) are included, and are shown as star markers.

rare in these young clumps, and only four of them can be robustly classified as centrally concentrated ( $Q > 0.8$ ). Although the occurrence of systems with a more pronounced centrally concentrated distributions (i.e.,  $Q > 0.8$ ) is certainly higher in the later stages of their evolution, mild or high subclustered distributions of cores are still present at later stages, as  $Q < 0.65$  is found in 27 systems out of 124 with  $L/M > 10L_{\odot}/M_{\odot}$ .

We do not find a significant relationship between  $Q$  and  $L/M$ , on the contrary, we observe a large scatter of  $Q$  for any interval of  $L/M$ , which is not connected to the level of fragmentation or to other clump-averaged properties, for example the clump surface density. It is unclear what the reasons are behind this large scatter, which may not be only caused by statistics but also by additional factors not considered in our analysis, as, for example, the surrounding local environment. A weak dependence of  $Q$  with the  $L/M$  ratio may be identified, but the Pearson and Spearman correlation coefficients equal to  $\rho_p = 0.18$  and  $\rho_s = 0.185$  with  $p$ -values of 0.0067 and 0.0053 are significantly lower than the values reported by Xu et al. (2024):  $\rho_p = 0.60$  and  $\rho_s = 0.57$ . This indicates that if a transition between spatial distribution patterns exists in massive clumps, it may require additional factors than simply the internal clump evolution.

### 3.5.2. Mass segregation in the ALMAGAL fields

A further clue on the formation and early evolution process can be extracted by also taking into account the mass of the cores. Observations of evolved star clusters reveal that they are often mass segregated, with massive objects that are more concentrated towards the cluster centers than the low-mass ones (Hillenbrand & Hartmann 1998). This observational signature is often interpreted as a consequence of the dynamical evolution caused by close encounters between pairs of cluster members that lead the system to a global state of energy equipartition. As a result, the less massive cluster components rise to higher orbits around the gravitational center, whereas the more massive cluster components move more slowly, falling into lower orbits.

**Table 6.** Diagnostics of spatial statistics in ALMAGAL clumps.

ALMAGAL ID	N	$Q$ -par <sup>a</sup>	$\bar{m}$ <sup>a</sup>	$\bar{s}$ <sup>a</sup>	$N_{MSR}^{crit\ b}$	$\Lambda_{MST}^{max\ b}$	$N_{MST}^{max\ b}$	MSI <sup>c</sup>	MS Flag <sup>d</sup>	Prob1 <sup>e</sup>	Prob2 <sup>e</sup>
AG010.2283–0.2064	9	0.78	0.38	0.49	–	–	–	1.154			
AG010.3226–0.1599	11	0.71	0.42	0.59	–	–	–	1.011			
AG011.9039–0.1403	11	0.65	0.46	0.70	5	3.50	2	1.854	Rob MS	1.000	
AG011.9183–0.6122	13	0.72	0.36	0.51	–	–	–	1.223			
AG011.9370–0.6160	22	0.85	0.39	0.46	3	2.71	3	1.368	Pos MS		
AG012.2087–0.1017	23	0.80	0.50	0.63	–	–	–	1.574			
AG012.6794–0.1830	8	0.61	0.46	0.76	2	0.57	2	0.799	Pos Inv MS		
AG012.7360–0.1031	11	0.67	0.47	0.70	–	–	–	1.183			
AG012.8535–0.2265	11	0.70	0.57	0.82	–	–	–	1.170			
AG012.8887+0.4890	15	0.70	0.38	0.54	4	6.37	2	2.209	Rob MS	1.000	
AG012.9008–0.2404	10	0.62	0.60	0.95	2	2.58	2	1.218	Rob MS	0.964	
AG012.9048–0.0306	14	0.60	0.36	0.61	2	0.50	2	0.862	Rob Inv MS		0.992
AG012.9084–0.2604	23	0.89	0.68	0.77	4	2.91	4	1.484	Rob MS	0.948	
AG012.9156–0.3341	8	0.75	0.49	0.66	4	3.02	4	2.060	Rob MS	0.930	

**Notes.** <sup>(a)</sup> Mean edge length of cluster MST,  $\bar{m}$ ; the cluster correlation length,  $\bar{s}$ ; and the  $Q$  parameter defined as  $Q = \bar{m}/\bar{s}$ , following [Cartwright & Whitworth \(2004\)](#). <sup>(b)</sup> Quantities of the mass segregation ratio (MSR) profile: maximum number of cores for which MSR profile is different from unity,  $N_{MSR}^{crit}$ ; the MSR maximum value,  $\Lambda_{MSR}^{max}$ ; and its corresponding number of cores,  $N_{MST}^{max}$ . These quantities are measured only in the 110 systems where the mass segregation (MS) is assessed. <sup>(c)</sup> Mass segregation integral, MSI, defined by [Xu et al. \(2024\)](#). <sup>(d)</sup> Flags reporting whether a signature of direct (Pos MS) or inverse (Pos Inv MS) mass segregation is found and if it is robust (additional flag Rob) against the Monte Carlo simulations described in the text. <sup>(e)</sup> Probability of presence of direct (Prob1) or inverse (Prob2) mass segregation evaluated with Monte Carlo simulations. The complete version of this table is available at the CDS.

However, young clusters may have been born with a primordial MS set by the formation process itself. In fact, primordial MS with respect to the center of the gravitational potential is a natural outcome of the competitive accretion model for high-mass stars ([Bonnell & Bate 2006](#)).

Evidence of MS has already been found in observations of young cluster systems in both high-mass and intermediate- to low-mass environments ([Chen et al. 2007](#); [Gennaro et al. 2011](#); [Parker 2018](#); [Plunkett et al. 2018](#); [Dib et al. 2018](#); [Dib & Henning 2019](#); [Sadaghiani et al. 2020](#); [Nony et al. 2021](#)). This indicates that MS is not produced merely by long-term dynamical evolution, but it can be already established in the early phases. However, it remains unclear whether it is primordial, i.e., set in the initial formation stages, or the result of a rapid, early dynamical evolution of the system of cores. For instance, weak or no MS signature is found in the 39 young clumps of the ASHES survey ([Sanhueza et al. 2019](#); [Morii et al. 2023](#)). In contrast, supporting the idea that it develops over time, clear segregation is observed in the more evolved systems, with 8 of 11 clumps in the ASSEMBLE survey presenting clear signatures of MS ([Xu et al. 2024](#)).

We searched for MS signature in ALMAGAL systems using the mass segregation ratio  $\Lambda_{MSR}$  ([Allison et al. 2009](#); [Parker 2018](#)), which compares the total length of the MST of the  $N_{MST}$  most massive cores with the typical one of the entire protocluster. The value of  $\Lambda_{MSR}$  for  $N_{MST}$  cores is given by

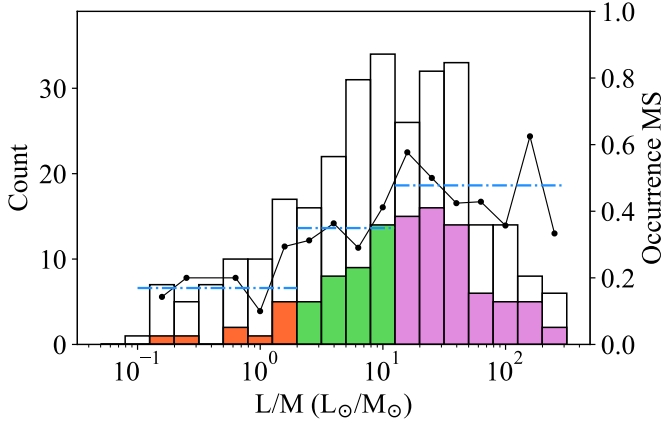
$$\Lambda_{MSR}(N_{MST}) = \frac{\langle l_{random} \rangle}{l_M} \pm \frac{\sigma_{l,random}}{l_M}, \quad (5)$$

where  $\langle l_{random} \rangle$  is the average of the total length of the MST derived by randomly selecting a total of  $N_{MST}$  cores from the cluster, while  $l_M$  is the total length of the MST of the  $N_{MST}$  most massive cores. We computed the average  $\langle l_{random} \rangle$  and the dispersion  $\sigma_{l,random}$  by randomly selecting the cores 1000 times as in [Morii et al. \(2023\)](#) and [Xu et al. \(2024\)](#). If MS is present, then the  $N_{MST}$  most massive cores are more concentrated, and the total

length of their MST would be shorter than the one built from randomly selected cluster members, which will give  $\Lambda_{MSR} \geq 1$ . In contrast,  $\Lambda_{MSR}$  would be  $\approx 1$  if the cores had a similar spatial distribution regardless of their mass.

We adopted the core masses estimated from [Coletta et al. \(2025\)](#) and computed  $\Lambda_{MSR}(N_{MST})$  in the 296 ALMAGAL fields hosting at least eight cores to limit the uncertainties due to their small number. This threshold corresponds to the minimum number of detections in the ASHES field for which the ratio  $\Lambda_{MSR}$  was also estimated ([Morii et al. 2024](#)). If we do not take into account any uncertainty in the core mass, we find that a signature of mass segregation is revealed in 171 fields, which have a MSR profile that differs from unity for more than the random uncertainty  $\sigma_{l,random}/l_M$  (it exceeds unity in 131 fields). We adopted a MonteCarlo approach to take into account the uncertainty on the core masses. For each system, we generated 500 unique realizations where every core mass was drawn from a Gaussian distribution centered on its value,  $M_{core}$ , with a standard deviation of 20%. We computed the  $\Lambda_{MSR}^j(N_{MST})$  with the relative error for each of these 500 synthetic realizations, fully mapping the range of possible outcomes given the assumed mass uncertainties. The fraction of realizations where  $\Lambda_{MSR}^j(N_{MST})$  exceeds unity over the dispersion  $\sigma_{l,random}$  defines the probability that mass segregation is present in the system. We classified a system as robustly mass segregated if this fraction is higher than 70%, a condition found in 110 fields ( $\sim 37\%$  of the total). This considerable fraction suggests that cores are already segregated in mass prior to dispersal of the gas present in the clump, and raises the question of whether segregation is set at the initial formation stage or is a consequence of the early evolution, which requires a comparative analysis between systems at different evolutionary stages.

Given their large numbers, we do not show all MSR profiles. Instead, we characterized each profile with metrics indicating average value, peak, and extent, reporting those quantities in Table 6. We included the mass segregation integral (MSI),  $I_{\Lambda}^{MSR}$



**Fig. 16.** Distribution of the ALMAGAL clumps hosting at least 10 cores in different bins of the ratio  $L/M$ . The colored histograms show the systems where a signature of MS is found, with different colors referring to the different intervals of  $L/M$  ratio, i.e., orange  $L/M \leq 2 L_{\odot}/M_{\odot}$ , green  $2 < L/M \leq 10 L_{\odot}/M_{\odot}$  and purple  $L/M > 10 L_{\odot}/M_{\odot}$ . The solid black line, which refers to the right y-axis scale, indicates the relative fraction of systems with MS with respect to the observed ones in bins of  $L/M$ , while the dot-dashed line shows the average occurrence fraction of MS systems in the three intervals of  $L/M$ .

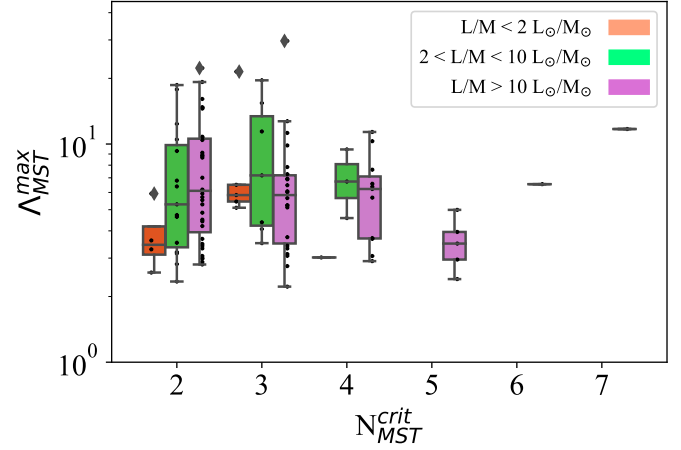
introduced by Xu et al. (2024) to quantify the fractional area of the MRS profile that exceeds unity. This quantity allows the comparison of the MSR profiles of different systems and is defined as

$$I_{\Lambda}^{MSR} = \frac{\sum_{i=2}^{N_{MST}^{max}} \Lambda_{MSR,i}}{N_{MST}^{max} - 1}, \quad (6)$$

where  $\Lambda_{MSR,i}$  is the value of the profile  $\Lambda_{MSR}(N_i)$  for the most massive cores  $N_i$ . Moreover, we determined the maximum of the profile,  $\Lambda_{MSR}^{max}(N_{MST}^{max})$ , and the largest number of cores,  $N_{MST}^{crit}$ , where it still remains above unity. These two quantities extend the information provided by the MSI, as they indicate how much more concentrated the most massive cores are than any other randomly selected group and the largest number of cores where the spatial distribution is found different.

Figure 16 shows the distribution of clumps with  $N_{cores} \geq 8$  as a function of the ratio  $L/M$  for systems with and without MS signature, indicating how frequently MS is found during the entire evolutionary sequence of massive clumps. We divided the sample of 296 clumps into three different groups based on the  $L/M$  ratio:  $L/M < 2 L_{\odot}/M_{\odot}$ ,  $2 < L/M \leq 10 L_{\odot}/M_{\odot}$ , and  $L/M > 10 L_{\odot}/M_{\odot}$ . The first and latter intervals are also sampled by the targets of the ASHES and ASSEMBLE surveys, allowing us to extend their results with larger sample sizes. These groups consist of 59, 83, and 154 clumps and have an average number of cores equal to 12, 15 and 17, respectively. The 110 systems where mass segregated is robustly confirmed divide in 10, 25, and 75 systems in each of these three groups. All of those systems have a different MSI distribution, and we rejected the null hypothesis that they derive from the same underlying one with the Kolmogorov–Smirnov and Mann-Whitney U tests with confidence levels set by  $p_{KS} < 0.029$  in the two cases with  $L/M < 2 L_{\odot}/M_{\odot}$  and  $2 < L/M < 10 L_{\odot}/M_{\odot}$ , and with a more robust assessment of  $p_{KS}$ ,  $p_U \lesssim 10^{-5}$  in the more numerous group with  $L/M > 10 L_{\odot}/M_{\odot}$ .

The occurrence of MS in the young clumps of the ALMAGAL sample ( $\sim 17\%$  with 10 out of 59) is similar (and compatible



**Fig. 17.** Boxplots showing the distribution of maximum value of the mass segregation ratio,  $\Lambda_{MST}^{max}$ , divided for the different values  $N_{MST}^{crit}$ , that identify the interval where the MSR profile is statistically higher than unity, in the 110 ALMAGAL clumps where signatures of MS are found. Those clumps are divided into three groups according to the  $L/M$  ratio composed of 59, 83, and 154 systems. The rectangular area of the boxplots shows the interquartile intervals, while the whiskers represent the 10th to 90th percentile intervals; the diamond symbols indicate the cases outside this interval.

within the uncertainty of the counting statistic) to the one measured in the ASHES sample (13% with 5 out of 39). These systems have a  $\Lambda_{MST}^{max}$  between  $\sim 2.5$  and  $\sim 7$  with an average value of  $\sim 5$ , indicating that when MS is present, even in these early phases, it usually has a moderate to high impact on the distribution of the cores. One peculiar system shows a very high value of  $\Lambda_{MST}^{max} \sim 20$ , but visual inspection of the system indicates that it is closely associated with multiple mid-infrared bright sources and, thus potentially affected by the strong feedback of nearby massive young stellar objects. Since there is only one case, we consider it an outlier. The fraction of clumps with MS steadily increases in more advanced stages of evolution. It first rises to 30% (25 out of 83) and then reaches  $\sim 49\%$  (75 out of 154) for  $2 < L/M < 10 L_{\odot}/M_{\odot}$  and  $L/M > 10 L_{\odot}/M_{\odot}$ , respectively. An indication of this effect was already suggested by Xu et al. (2024) who found that 8 out of 11 clumps with  $L/M > 10 L_{\odot}/M_{\odot}$  are mass segregated, but the large sample size provided by the ALMAGAL large program allows us to robustly and clearly assess this trend, also measuring the increase for intermediate evolutionary stages.

Furthermore, we also found that its signature affects a larger number of cores and becomes stronger in more evolved systems. Figure 17 shows the distributions of the maximum value of the mass segregation ratio,  $\Lambda_{MST}^{max}$ , and of  $N_{MST}^{crit}$  in the three aforementioned samples. The intensity of segregation is always greater than  $\sim 2.5$ , with most systems having  $\Lambda_{MST} > 3$ , which is considered a high level of MS (Pouteau et al. 2023). The median value of  $\Lambda_{MST}^{max}$  is  $\sim 5$  in less evolved clumps, and increases to  $\sim 6.4$  and  $\sim 6.2$  for the intermediate and more evolved subsamples, respectively. This suggests that the cores with higher mass become more concentrated as the clump evolves.

Segregation generally affects only two or three most massive cores of the clusters, as  $\sim 80\%$  of the system where segregation is present have  $N_{MST}^{crit} \leq 3$ , with 54 and 34 systems where  $N_{MST}^{crit} = 2$  and 3, respectively. While the evolution leads to an average increase in the number of cores (Elia et al. 2026), at the same time we found that the number of cores showing a tighter spatial distribution also increases and may involve up to

the five most massive cores in the system (only two systems show  $N_{MST}^{crit} \geq 6$ ).

In conclusion, although MS may be present in the earliest stages, it is about three times more common in evolved clumps ( $L/M > 10 L_{\odot}/M_{\odot}$ ). Segregation is more pronounced in systems in more evolved phases, when cores with masses of a few  $M_{\odot}$  in  $\sim 1000$ – $2000$  au are commonly found in the system (see Fig. 18 in Coletta et al. 2025). The gravitational concentration of the entire clump or the gas accretion toward its center has been suggested to be responsible for the observed increase in MS (Xu et al. 2024). We do not find any connection between the signature of MS and the clump surface density  $\Sigma_{cl}$  or the cluster radius  $R_c$ , as it would be expected if gravitational collapse was present. However, it is important to note that *Herschel* derived surface densities might not be sufficient to highlight the increase in density of the clump central region because of their poor resolution. Further analysis of the dynamics and the gas distribution in clumps with high signatures of MS is necessary using molecular line tracers at high angular resolution to verify whether clump material is efficiently transferred toward central regions and the most massive cores.

#### 4. Discussion

ALMAGAL observations reveal that fragmentation in high-mass clumps is widely diversified, with star-forming cores spatially arranged across a broad spectrum of configurations, from sparse, mildly concentrated, up to highly clustered. This observed heterogeneity, which mirrors the wide scatter found for core masses and fragmentation degrees discussed by Coletta et al. (2025) and Elia et al. (2026), does not preclude a deeper analysis. Rather, systematic analysis of spatial distributions and core-to-core separations provides a powerful diagnostic tool for the physical mechanisms that govern where and how fragmentation takes place. Our data show that most of the small-scale ( $\sim 1400$  au) cores are confined within the inner clump region ( $\leq 0.35$  pc). Additional cores forming at larger distances are rare or, if present, are likely low-mass objects undetectable in ALMAGAL fields, which would be compatible with their formation in low-density environment. This result might favor an in situ formation scenario for clusters (Dib et al. 2010). However, it does not contradict the more dynamic conveyor belt-type models (Krumholz & McKee 2020; Vázquez-Semadeni et al. 2019) that involve anisotropic rather than monolithic contraction, with material being fed into the clump along large-scale flows. The Global Hierarchical Collapse (GHC) model, an example of these dynamic models, predicts gravity-driven accretion along extended filaments to supply material to cores and hubs (Vázquez-Semadeni et al. 2019, 2025). GHC simulations show that low-mass cores may form along these peripheral filaments (Gómez & Vázquez-Semadeni 2014; Vázquez-Semadeni et al. 2017), a scenario consistent with observed cores found along filaments connected to high-density clumps (Myers 2009; Kirk et al. 2013; Peretto et al. 2014; Hacar et al. 2017; Chen et al. 2019; Álvarez-Gutiérrez et al. 2024). The ALMAGAL cores found on the outskirts of the clumps may be formed in similar flows whose underlying filaments remain undetected because of the limited sensitivity and coverage. In this context, the presence of aligned cores in several ALMAGAL systems can be an additional footprint of the presence of these flows. In fact, numerical simulations show that the cluster structure retains indications of the native distribution and organization in structures of the star-forming material (Vázquez-Semadeni et al. 2017).

Two significant results are found from the analysis of the separations between cores in the ALMAGAL high-mass clumps.

The separation is on average of the order, or smaller, of the pure thermal Jeans length over almost the entire early evolution of the clump sampled by the ALMAGAL survey. This evidence confirms and extends the results of previous studies based on a limited number of clumps (Palau et al. 2015, 2018; Beuther et al. 2018; Liu et al. 2020; Zhang et al. 2021; Morii et al. 2024; Ishihara et al. 2024). Separations close to the thermal Jeans length are predicted in the case that the fragmentation process is driven mainly by gravitational contraction (Palau et al. 2014; Vázquez-Semadeni et al. 2025). Nonetheless, we found indications of systems in which the separations between the cores are significantly larger than the thermal Jeans length,  $\lambda_J^{th}$ , by an average factor of three. One possible explanation is that there is a contribution due to turbulence, which provides additional support and increases the Jeans length (Mac Low & Klessen 2004). The turbulent Jeans length can be computed by replacing in Eq. (B.2) the sound speed  $c_s$  with the gas velocity dispersion  $\sigma_v$  given by  $\sigma_v = \sqrt{c_s^2 + \sigma_{nth}^2}$ , where  $\sigma_{nth}^2$  are the non-thermal motions present in the gas. The typical properties of the ALMAGAL clumps in the earliest phases have a median value of  $c_s \approx 0.2$  km s $^{-1}$ ; therefore, to explain the observed ratio  $R \sim 3$  it would be sufficient to have a moderate non-thermal velocity dispersion  $\sigma_{nt} \approx 0.6$  km s $^{-1}$ , which corresponds to Mach number  $\mathcal{M} \approx 5$ , evaluated as  $\mathcal{M} = \sqrt{3}\sigma_{nth}/c_s$ . To compare it with the average conditions present in the entire hosting molecular cloud, we matched the position of the ALMAGAL clumps with the clouds identified in the SEDIGISM survey using the  $^{13}\text{CO}(2-1)$  emission (Duarte-Cabral et al. 2021). We found that the typical  $\sigma_v$  of the matched clouds is  $\sim 1.8$  km s $^{-1}$ , with a distribution with interquartile interval from  $\sim 1.3$  to  $\sim 2.6$  km s $^{-1}$ . These values indicate a remarkable contribution of non-thermal velocity dispersion at the cloud scale, which would produce significantly larger separations between cores than the observed ones.

Another possibility is that different fragmentation regimes, each with its own characteristic length, are present in a fraction of ALMAGAL clumps, which would also explain the asymmetries found in the distribution of separations  $P(l_i)$ . These regimes may be due to different contributions of the magnetic field, local turbulence, and density distribution (Fontani et al. 2018), be the result of the assembly of several subclusters (Morii et al. 2023), or the fragmentation in the filaments connected to the accretion flow expected in the GHC model (Vázquez-Semadeni et al. 2019). A more robust interpretation of our observations requires a one-to-one comparison with dedicated simulations such as these developed within the Rosetta Stone project which model the formation and temporal evolution of massive clumps (Lebreuilly et al. 2025), produce post-processed maps at multiple wavelengths (Tung et al. 2025), and create mock observations that closely resemble ALMA data with high fidelity (Nucara et al. 2025).

In addition to the comparison with the Jeans fragmentation theory, an important result emerging from the ALMAGAL clusters of cores is the progressive statistical decrease exhibited by the separations with the  $L/M$  ratio, which can be interpreted as evidence that cores are more tightly packed in more evolved systems compared to the younger ones. This evidence has been advocated in support of the GHC model for high-mass and cluster formation as it naturally derives from the global contraction of the systems (Traficante et al. 2023; Xu et al. 2024; Vázquez-Semadeni et al. 2025). Our analysis shows that typical separations decrease by about a factor of three, with a major decrease from  $\sim 20\,000$  down to  $\sim 10\,000$  au taking place at  $L/M \approx 1 L_{\odot}/M_{\odot}$ . Above this value, our data show only marginal

variations, and the observed trend does not clearly reflect a decrease in core separations due to the ongoing global collapse of the clump. If a dynamical shrinking of the cluster is occurring, it is not the dominant effect traced by the median of the population of separation measured in ALMAGAL fields. The value  $L/M \approx 1 L_{\odot}/M_{\odot}$  is reached when the luminosity of already formed protostars becomes relevant for the entire clump (Molinari et al. 2008), and corresponds to an increase in the degree of fragmentation observed in these systems. The relation of the degree of fragmentation with evolution is discussed by Elia et al. (2026) who reported a steady increase on average in the number of cores over the entire interval  $0.1 \leq L/M \leq 100 L_{\odot}/M_{\odot}$ . The increase in fragmentation is also presented in Fig. 11, which also shows the average variation of the core separations. We note that the observed decrease in these separations follows the transition from clumps characterized mainly by a limited number of cores ( $N_{cores} \leq 8$ ) to more crowded systems.

We attempted to draw an overall scenario from ALMAGAL observations. Our results indicate that fragmentation occurs in multiple substructures of the clump, which are generally widely separated by  $\sim 15\,000$ – $30\,000$  au ( $\sim 0.07$ – $0.14$  pc), measured in systems where fragmentation is at the onset. These substructures host a limited number of already collapsed cores, but also fragments that did not yet have enough time to grow sufficiently to reach a high contrast over the parent structure to be detectable, as proposed by Louvet et al. (2024) to justify the observed small fragmentation present in young systems. As the clump evolves, additional cores collapse and become detectable in the ALMAGAL fields with a sufficiently high significance. These newly collapsed cores form in the midst of the original population, but a minor fraction of these objects also form in the outskirts of the clump, even though less material is available. Those effects produce a decrease in the observed average separation while the additional cores form by collapsing material.

Nonetheless, it is notable that systems with cores distributed following a smooth and radially concentrated profile start to be observed at the same  $L/M$  value where the typical core separations decrease. In addition to the increase in fragmentation level, the emergence of those systems may be the signature of a gradual merging and dispersion of the initial, fractal distribution in subclusters, that characterizes the young ALMAGAL clumps. This outcome is consistent with predictions from dynamical-type models where conveyor belt-like processes are active (Longmore et al. 2014; Krumholz & McKee 2020; Vázquez-Semadeni et al. 2025). According to these predictions, time evolution causes an increase in the degree of concentration of cores as the system shrinks and dynamically relaxes, as proposed by Xu et al. (2024). However, while our results indicate that smooth, centrally concentrated distributions occur more frequently at later stages, we do not observe a tight correlation with evolution that would robustly support this scenario over the entire sample. Instead, we still observe that subclustering signatures are present in a considerable fraction of systems with high values of the  $L/M$  ratio. As this pattern is often observed also in young embedded stellar clusters (Kuhn et al. 2014), this suggests the presence of mechanisms that work against the merging of these substructures or that are able to slow the process down for long timescales. The scatter of  $Q$  values and, more generally, the observed heterogeneity of the fragmentation in the ALMAGAL sample questions whether a unique path of evolution, like the one described above, can be traced for all the massive dense clumps observed. Additional factors that are not taken into account in this analysis may influence the evolution of the spatial distribution of cores and lead to the scatter of  $Q$  values for each given interval of the  $L/M$

ratio. Low values of the  $Q$  parameter are also measured when the main system increases its radial concentration, but additional substructures form in the outskirts of the field and become statistically significant. Moreover, it has been found that the intensity of the internal mechanical feedback affects the final structure and distribution of stars in clusters. Numerical simulations including protostellar jets have typically higher values of  $Q$  over time than those without any feedback or with only the radiative one (Verliat et al. 2022). In addition to the evolutionary effect, the variations in measured  $Q$  may reflect differences in the initial conditions with which the region had initially assembled and fragmented, as pointed out by Dib & Henning (2019) from the analysis of nearby low-mass star-forming regions. As the ALMAGAL fields cover different regions in the Galaxy, the clumps may be formed with different physical conditions imprinted by the large scale environment. In summary, our results are compatible with a dynamical evolutionary scenario, which defines the evolutionary path of a subset of massive clumps, and we cannot exclude the possibility that dynamical effects such as gravitational collapse, system shrinking, and dynamical relaxation are active in massive clumps.

## 5. Conclusion

We analyzed the spatial distribution of the 5728 dense cores extracted by Coletta et al. (2025) in the 514 ALMAGAL clumps with at least four detections. We provided a quantitative and statistically robust characterization of the fragmentation process in massive and potentially cluster-forming clumps, determining where the cores are located, the typical length of their reciprocal separations, and how they are distributed. Our measurements compose a comprehensive set of observational constraints for the fragmentation process against which to compare the predictions of different theories of cluster formation. We focused our analysis on how the properties of the statistical distribution change during the evolution of the systems before gas dispersal.

Our results are as follows:

- The ALMAGAL clumps show a wide variety of fragmentation patterns, with cores distributed over regions that extend from  $\leq 0.1$  up to  $\sim 0.8$  pc, which closely match the inner size of the hosting clump as determined from *Herschel* observations. Although cores are also observed in the outskirts of the clump, they are either rare or are low mass (and thus often undetected), as expected from formation in a low-density environment. We introduced a tentative classification based on the components and elongation of the convex hull polygon defined from the core positions, finding that several configuration patterns are present in the sample. We classified them into aligned, sparse, circular, elongated, and subclustered arrangements, with the most common pattern being moderately elongated, with aspect ratio  $\sim 2$ – $3$ , although we note that highly elongated systems are also present;
- The projected separations between the cores measured with the MST method show a broad distribution with a median value of  $\sim 6600$  au and  $\sim 7400$  au for clumps that are located at heliocentric distances  $2 < d_{cl} < 4.5$  kpc and  $4.5 < d_{cl} < 7.5$  kpc, respectively. Smaller median separations are measured in systems with  $d_{cl} \leq 3.7$  kpc, suggesting that there are different characteristic separations in the inner regions of the clumps, which are expected to be denser. Evidence that a single characteristic length for a clump may not completely describe the observed fragmentation is also found via a comparison of the medians and modes of the distribution of the

core separations, which differ from each other in 121 systems (~24% of the analyzed sample). These differences are due to a local subclustering or to the presence of a more widely spaced population of cores in the outskirts of the system that followed a different fragmentation pattern;

- We identified a statistically significant shift in the distribution of core separations with the  $L/M$  ratio, which can be interpreted as a variation during clump evolution. The observed shift varies significantly in the early phases ( $L/M \lesssim 2L_{\odot}/M_{\odot}$ ) when it drops from  $\sim 20\,000$  au to  $\sim 10\,000$  au for  $2 \leq L/M \leq 10L_{\odot}/M_{\odot}$ , and displays smaller changes for more evolved phases dropping to  $\sim 7000$  au for systems with  $L/M \geq 10L_{\odot}/M_{\odot}$ . Although the observed trend can be statistically described by a steady decrease in these two intervals, the larger variation appears to be produced by an increase in the degree of fragmentation in more evolved systems, as more cores are detected;
- We found a good agreement between the core separations and the thermal Jeans lengths  $\lambda_J^{th}$  computed from the clump-averaged physical parameters, with a median ranging between  $\sim 0.5$  and  $\sim 2$  times  $\lambda_J^{th}$ . This agreement remains when the separations are divided according to the clump  $L/M$ . Nevertheless, we identified a small population of cores whose separations are centered at  $\sim 3 \times \lambda_J^{th}$ , which are still smaller than the prediction of turbulent fragmentation of the hosting molecular cloud. These cores are mostly hosted in clumps with low levels of fragmentation and in early phases of evolution, with  $L/M \lesssim 0.5-1L_{\odot}/M_{\odot}$ . However, our results are compatible with thermal Jeans fragmentation being active in massive clumps. Although the masses are affected by large uncertainties and systematics, we note that the typical mass of cores is compatible with the thermal Jeans mass only in young and pristine clumps, where we observe that the core mass distribution is centered at  $M_{core} \sim 0.6M_J^{th}$ ;
- We did not find robust evidence of a progressive transition of the structure of the spatial distribution of the cores with evolution. The analysis using the  $Q$  parameter indicates that the cores are often distributed in a fractal type of subclustering in the early stages characterized by  $L/M \leq 1L_{\odot}/M_{\odot}$ . Although we found several clusters that are most likely distributed with radially concentrated profiles in the evolved clumps with  $L/M \geq 10L_{\odot}/M_{\odot}$ , subclustering is still often detected in those phases. The existence of these systems suggests that, in addition to the clump average properties, there are other factors that play a role in shaping the structure of the cluster over the evolution;
- Signatures of mass segregation are found in 110 fields out of the 296 ALMAGAL clumps hosting more than eight cores. The occurrence of mass segregation is lower ( $\sim 17\%$ ) in young systems, but steadily increases with the evolutionary stage to  $\sim 30\%$  in intermediate-evolved clumps with  $2 \leq L/M < 10L_{\odot}/M_{\odot}$  and up to  $\sim 48\%$  in the more evolved systems in the ALMAGAL sample with  $L/M \geq 10L_{\odot}/M_{\odot}$ .

## Data availability

Tables 2, 3, and 6 are available at the CDS via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/707/A221>.

*Acknowledgements.* We are very grateful to the anonymous referee for constructive comments that helped clarify a number of points in the paper. The Teams at INAF-IAPS, Heidelberg, Paris-Saclay and University of Bologna acknowledge financial support from the European Research Council via the

ERC Synergy Grant “ECOGAL” (project ID 855130). R.S.K. also acknowledges financial support from the Heidelberg Cluster of Excellence (EXC 2181 – 390900948) “STRUCTURES”, funded by the German Excellence Strategy, and from the German Ministry for Economic Affairs and Climate Action in project “MAINN” (funding ID 50002206), and is also grateful to the Harvard Radcliffe Institute for Advanced Studies and Harvard-Smithsonian Center for Astrophysics for their hospitality and support during his sabbatical. A.S.-M. acknowledges support from the RyC2021-032892-I grant funded by MCIN/AEI/10.13039/501100011033 and by the European Union “Next Generation EU”/PRTR, as well as the program Unidad de Excelencia María de Maeztu CEZ2020-001058-M, and support from PID2023-144675NB-I00 (MCI-AEI-FEDER, UE). Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). R.K. acknowledges financial support via the Heisenberg Research Grant funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant no. KU 2849/9, project no. 445783058. P.S. was partially supported by a Grant-in-Aid for Scientific Research (KAKENHI Number JP22H01271 and JP23H01221) of JSPS. P.S. was supported by Yoshinori Ohsumi Fund (Yoshinori Ohsumi Award for Fundamental Research). S.W. has been supported by the Collaborative Research Centre 1601 (SFB 1601) funded by the Deutsche Forschungsgemeinschaft (DFG). L.B. gratefully acknowledges support from the ANID BASAL project FB210003. Y.-W. T. is supported by NSTC grants 112-2112-M-001-066- and NSTC 114-2112-M-001-042. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2019.1.00195.L. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

## References

- Adamo, A., Zeidler, P., Kruijssen, J. M. D., et al. 2020, *Space Sci. Rev.*, **216**, 69
- Adams, F. C. 2010, *ARA&A*, **48**, 47
- Allison, R. J., Goodwin, S. P., Parker, R. J., et al. 2009, *ApJ*, **700**, L99
- Álvarez-Gutiérrez, R. H., Stutz, A. M., Sandoval-Garrido, N., et al. 2024, *A&A*, **689**, A74
- Anderson, M., Peretto, N., Ragan, S. E., et al. 2021, *MNRAS*, **508**, 2964
- André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, in *Protostars and Planets VI*, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 27
- Avison, A., Fuller, G. A., Frimpong, N. A., et al. 2023, *MNRAS*, **526**, 2278
- Baldeschi, A., Elia, D., Molinari, S., et al. 2017, *MNRAS*, **466**, 3682
- Ballesteros-Paredes, J., Vázquez-Semadeni, E., Palau, A., & Klessen, R. S. 2018, *MNRAS*, **479**, 2112
- Beraldo e Silva, L., Debattista, V. P., Nidever, D., Amarante, J. A. S., & Garver, B. 2021, *MNRAS*, **502**, 260
- Beuther, H., Schilke, P., Menten, K. M., et al. 2002, *ApJ*, **566**, 945
- Beuther, H., Mottram, J. C., Ahmadi, A., et al. 2018, *A&A*, **617**, A100
- Beuther, H., Gieser, C., Suri, S., et al. 2021, *A&A*, **649**, A113
- Beuther, H., Gieser, C., Soler, J. D., et al. 2024, *A&A*, **682**, A81
- Beuther, H., Kuiper, R., & Tafalla, M. 2025, *ARA&A*, **63**, 1
- Bonnell, I. A., & Bate, M. R. 2006, *MNRAS*, **370**, 488
- Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 2001, *MNRAS*, **323**, 785
- Bonnell, I. A., Bate, M. R., & Vine, S. G. 2003, *MNRAS*, **343**, 413
- Bressert, E., Bastian, N., Gutermuth, R., et al. 2010, *MNRAS*, **409**, L54
- Cartwright, A., & Whitworth, A. P. 2004, *MNRAS*, **348**, 589
- Cartwright, A., Whitworth, A. P., & Nutter, D. 2006, *MNRAS*, **369**, 1411
- Chen, L., de Grijs, R., & Zhao, J. L. 2007, *AJ*, **134**, 1368
- Chen, H.-R. V., Zhang, Q., Wright, M. C. H., et al. 2019, *ApJ*, **875**, 24
- Chen, Y., Li, H., & Vogelsberger, M. 2021, *MNRAS*, **502**, 6157
- Coletta, A., Molinari, S., Schisano, E., et al. 2025, *A&A*, **696**, A151
- Csengeri, T., Bontemps, S., Wyrowski, F., et al. 2017, *A&A*, **600**, L10
- Dale, J. E., & Bonnell, I. 2011, *MNRAS*, **414**, 321
- Dib, S. 2023, *ApJ*, **959**, 88
- Dib, S., & Basu, S. 2018, *A&A*, **614**, A43
- Dib, S., & Henning, T. 2019, *A&A*, **629**, A135
- Dib, S., Shadmehri, M., Padoan, P., et al. 2010, *MNRAS*, **405**, 401
- Dib, S., Schmeja, S., & Parker, R. J. 2018, *MNRAS*, **473**, 849
- Dib, S., Bontemps, S., Schneider, N., et al. 2020, *A&A*, **642**, A177
- Dobbs, C. L., Bending, T. J. R., Pettitt, A. R., & Bate, M. R. 2022, *MNRAS*, **509**, 954
- Duarte-Cabral, A., Colombo, D., Urquhart, J. S., et al. 2021, *The SEDIGISM survey: molecular clouds in the inner Galaxy*
- Elia, D., Molinari, S., Schisano, E., et al. 2017, *MNRAS*, **471**, 100

- Elia, D., Merello, M., Molinari, S., et al. 2021, *MNRAS*, 504, 2742
- Elia, D., Coletta, A., Molinari, S., et al. 2026, *A&A*, 705, A100
- Elmegreen, B. G. 2010, in *IAU Symposium*, 266, Star Clusters: Basic Galactic Building Blocks Throughout Time and Space, eds. R. de Grijs, & J. R. D. Lépine, 3
- Elmegreen, B. G., & Falgarone, E. 1996, *ApJ*, 471, 816
- Fontani, F., Commerçon, B., Giannetti, A., et al. 2018, *A&A*, 615, A94
- Gennaro, M., Brandner, W., Stolte, A., & Henning, T. 2011, *MNRAS*, 412, 2469
- Gieser, C., Beuther, H., Semenov, D., et al. 2021, *A&A*, 648, A66
- Gieser, C., Beuther, H., Semenov, D., et al. 2022, *A&A*, 657, A3
- Girichidis, P., Federrath, C., Banerjee, R., & Klessen, R. S. 2011, *MNRAS*, 413, 2741
- Girichidis, P., Offner, S. S. R., Kritsuk, A. G., et al. 2020, *Space Sci. Rev.*, 216, 68
- Goldsmith, P. F. 2001, *ApJ*, 557, 736
- Gómez, G. C., & Vázquez-Semadeni, E. 2014, *ApJ*, 791, 124
- Goodwin, S. P., & Whitworth, A. P. 2004, *A&A*, 413, 929
- Gregorio-Hetem, J., Hetem, A., Santos-Silva, T., & Fernandes, B. 2015, *MNRAS*, 448, 2504
- Grudić, M. Y., Guszejnov, D., Hopkins, P. F., et al. 2018, *MNRAS*, 481, 688
- Grudić, M. Y., Guszejnov, D., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021, *MNRAS*, 506, 2199
- Guszejnov, D., Grudić, M. Y., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021, *MNRAS*, 502, 3646
- Gutermuth, R. A., Megeath, S. T., Myers, P. C., et al. 2009, *ApJS*, 184, 18
- Guzmán, A. E., Sanhueza, P., Contreras, Y., et al. 2015, *ApJ*, 815, 130
- Hacar, A., Alves, J., Tafalla, M., & Goicoechea, J. R. 2017, *A&A*, 602, L2
- Hacar, A., Tafalla, M., Forbrich, J., et al. 2018, *A&A*, 610, A77
- Hacar, A., Clark, S. E., Heitsch, F., et al. 2023, in *Astronomical Society of the Pacific Conference Series*, 534, Protostars and Planets VII, eds. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 153
- Hillenbrand, L. A., & Hartmann, L. W. 1998, *ApJ*, 492, 540
- Hoffman, R., & Jain, A. K. 1983, *Pattern Recognit. Lett.*, 1, 175
- Ishihara, K., Sanhueza, P., Nakamura, F., et al. 2024, *ApJ*, 974, 95
- Ishihara, K., Nakamura, F., Sanhueza, P., & Saito, M. 2025, *A&A*, 695, L25
- Jaffa, S. E., Whitworth, A. P., & Lomax, O. 2017, *MNRAS*, 466, 1082
- Jeans, J. H. 1902, *Philos. Trans. Roy. Soc. Lond. Ser. A*, 199, 1
- Jia, B.-S., Zhang, G.-Y., Men'shchikov, A., et al. 2025, *Res. Astron. Astrophys.*, 25, 085018
- Juvela, M., Harju, J., Ysard, N., & Lunttila, T. 2012, *A&A*, 538, A133
- Kainulainen, J., Stutz, A. M., Stanke, T., et al. 2017, *A&A*, 600, A141
- Kauffmann, J., & Pillai, T. 2010, *ApJ*, 723, L7
- Kauffmann, J., Bertoldi, F., Bourke, T. L., Evans, II, N. J., & Lee, C. W. 2008, *A&A*, 487, 993
- Kirk, H., Myers, P. C., Bourke, T. L., et al. 2013, *ApJ*, 766, 115
- Könyves, V., André, P., Men'shchikov, A., et al. 2015, *A&A*, 584, A91
- Krause, M. G. H., Offner, S. S. R., Charbonnel, C., et al. 2020, *Space Sci. Rev.*, 216, 64
- Kruijssen, J. M. D. 2012, *MNRAS*, 426, 3008
- Krumholz, M. R. 2006, *ApJ*, 641, L45
- Krumholz, M. R., & McKee, C. F. 2008, *Nature*, 451, 1082
- Krumholz, M. R., & McKee, C. F. 2020, *MNRAS*, 494, 624
- Krumholz, M. R., Bate, M. R., Arce, H. G., et al. 2014, in *Protostars and Planets VI*, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 243
- Kuhn, M. A., Feigelson, E. D., Getman, K. V., et al. 2014, *ApJ*, 787, 107
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, 41, 57
- Lebreuilly, U., Traficante, A., Nucara, A., et al. 2025, *A&A*, 701, A217
- Li, G.-X. 2024, *MNRAS*, 528, L52
- Li, H. B., Goodman, A., Sridharan, T. K., et al. 2014, in *Protostars and Planets VI*, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 101
- Li, H., Vogelsberger, M., Marinacci, F., & Gnedin, O. Y. 2019, *MNRAS*, 487, 364
- Lin, Y., Wyrowski, F., Liu, H. B., et al. 2022, *A&A*, 658, A128
- Liu, T., Evans, N. J., Kim, K.-T., et al. 2020, *MNRAS*, 496, 2790
- Liu, H.-L., Tej, A., Liu, T., et al. 2022, *MNRAS*, 511, 4480
- Liu, X., Liu, T., Zhu, L., et al. 2024, *Res. Astron. Astrophys.*, 24, 025009
- Longmore, S. N., Kruijssen, J. M. D., Bastian, N., et al. 2014, in *Protostars and Planets VI*, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 291
- Louvet, F., Sanhueza, P., Stutz, A., et al. 2024, *A&A*, 690, A33
- Lumsden, S. L., Hoare, M. G., Urquhart, J. S., et al. 2013, *ApJS*, 208, 11
- Mac Low, M.-M., & Klessen, R. S. 2004, *Rev. Mod. Phys.*, 76, 125
- Maschberger, T., Clarke, C. J., Bonnell, I. A., & Kroupa, P. 2010, *MNRAS*, 404, 1061
- McKee, C. F., & Tan, J. C. 2003, *ApJ*, 585, 850
- Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2016, *AJ*, 151, 5
- Megeath, S. T., Gutermuth, R. A., & Kounkel, M. A. 2022, *PASP*, 134, 042001
- Merello, M., Molinari, S., Rygl, K. L. J., et al. 2019, *MNRAS*, 483, 5355
- Mininni, C., Molinari, S., Soler, J. D., et al. 2025, *A&A*, 699, A34
- Molinari, S., Pezzuto, S., Cesaroni, R., et al. 2008, *A&A*, 481, 345
- Molinari, S., Swinyard, B., Bally, J., et al. 2010, *A&A*, 518, L100
- Molinari, S., Schisano, E., Faustini, F., et al. 2011, *A&A*, 530, A133
- Molinari, S., Merello, M., Elia, D., et al. 2016a, *ApJ*, 826, L8
- Molinari, S., Schisano, E., Elia, D., et al. 2016b, *A&A*, 591, A149
- Molinari, S., Schilke, P., Battersby, C., et al. 2025, *A&A*, 696, A149
- Morii, K., Sanhueza, P., Nakamura, F., et al. 2021, *ApJ*, 923, 147
- Morii, K., Sanhueza, P., Nakamura, F., et al. 2023, *ApJ*, 950, 148
- Morii, K., Sanhueza, P., Zhang, Q., et al. 2024, *ApJ*, 966, 171
- Moscadelli, L., Beuther, H., Ahmadi, A., et al. 2021, *A&A*, 647, A114
- Motte, F., Bontemps, S., & Louvet, F. 2018, *ARA&A*, 56, 41
- Motte, F., Bontemps, S., Csengeri, T., et al. 2022, *A&A*, 662, A8
- Myers, P. C. 2009, *ApJ*, 700, 1609
- Nony, T., Robitaille, J. F., Motte, F., et al. 2021, *A&A*, 645, A94
- Nucara, A., Traficante, A., Lebreuilly, U., et al. 2025, *A&A*, 701, A219
- Padoan, P., Pan, L., Juvela, M., Haugbølle, T., & Nordlund, Å. 2020, *ApJ*, 900, 82
- Palau, A., Estalella, R., Girart, J. M., et al. 2014, *ApJ*, 785, 42
- Palau, A., Ballesteros-Paredes, J., Vázquez-Semadeni, E., et al. 2015, *MNRAS*, 453, 3785
- Palau, A., Zapata, L. A., Román-Zúñiga, C. G., et al. 2018, *ApJ*, 855, 24
- Palau, A., Zhang, Q., Girart, J. M., et al. 2021, *ApJ*, 912, 159
- Parker, R. J. 2014, *MNRAS*, 445, 4037
- Parker, R. J. 2018, *MNRAS*, 476, 617
- Parker, R. J., & Alves de Oliveira, C. 2017, *MNRAS*, 468, 4340
- Parker, R. J., & Goodwin, S. P. 2015, *MNRAS*, 449, 3381
- Peretto, N., Fuller, G. A., André, P., et al. 2014, *A&A*, 561, A83
- Plunkett, A. L., Fernández-López, M., Arce, H. G., et al. 2018, *A&A*, 615, A9
- Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, *ARA&A*, 48, 431
- Pouteau, Y., Motte, F., Nony, T., et al. 2023, *A&A*, 674, A76
- Prim, R. C. 1957, *Bell Syst. Tech. J.*, 36, 1389
- Rebolledo, D., Guzmán, A. E., Contreras, Y., et al. 2020, *ApJ*, 891, 113
- Rigby, A. J., Peretto, N., Anderson, M., et al. 2024, *MNRAS*, 528, 1172
- Ripley, B. D., & Rassin, J.-P. 1977, *J. Appl. Probab.*, 14, 483
- Sadaghiani, M., Sánchez-Monge, Á., Schilke, P., et al. 2020, *A&A*, 635, A2
- Sánchez, N., & Alfaro, E. J. 2009, *ApJ*, 696, 2086
- Sánchez-Monge, Á., Brogan, C. L., Hunter, T. R., et al. 2025, *A&A*, 696, A150
- Sanhueza, P., Jackson, J. M., Zhang, Q., et al. 2017, *ApJ*, 841, 97
- Sanhueza, P., Contreras, Y., Wu, B., et al. 2019, *ApJ*, 886, 102
- Sanhueza, P., Liu, J., Morii, K., et al. 2025, *ApJ*, 980, 87
- Scalo, J. M. 1985, in *Protostars and Planets II*, eds. D. C. Black, & M. S. Matthews, 201
- Schisano, E., Rygl, K. L. J., Molinari, S., et al. 2014, *ApJ*, 791, 27
- Schisano, E., Molinari, S., Elia, D., et al. 2020, *MNRAS*, 492, 5420
- Schmeja, S., & Klessen, R. S. 2006, *A&A*, 449, 151
- Schmeja, S., Kumar, M. S. N., & Ferreira, B. 2008, *MNRAS*, 389, 1209
- Schneider, N., Csengeri, T., Hennemann, M., et al. 2012, *A&A*, 540, L11
- Schneider, N., Ossenkopf-Okada, V., Clarke, S., et al. 2022, *A&A*, 666, A165
- Scott, D. W. 1992, *Multivariate Density Estimation* (Wiley)
- Shima, K., Tasker, E. J., & Habe, A. 2017, *MNRAS*, 467, 512
- Silverman, B. W. 1981, *J. Roy. Stat. Soc.*, 43, 97
- Silverman, B. W. 1986, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall)
- Smith, R. J., Treß, R. G., Sormani, M. C., et al. 2020, *MNRAS*, 492, 1594
- Svoboda, B. E., Shirley, Y. L., Traficante, A., et al. 2019, *ApJ*, 886, 36
- Tang, Y.-W., Koch, P. M., Peretto, N., et al. 2019, *ApJ*, 878, 10
- Traficante, A., Jones, B. M., Avison, A., et al. 2023, *MNRAS*, 520, 2306
- Tung, N.-D., Traficante, A., Lebreuilly, U., et al. 2025, *A&A*, 701, A218
- Vázquez-Semadeni, E., González-Samaniego, A., & Colín, P. 2017, *MNRAS*, 467, 1313
- Vázquez-Semadeni, E., Palau, A., Ballesteros-Paredes, J., Gómez, G. C., & Zamora-Avilés, M. 2019, *MNRAS*, 490, 3061
- Vázquez-Semadeni, E., Palau, A., Gómez, G. C., et al. 2025, *MNRAS* [arXiv:2408.10406]
- Verliat, A., Hennebelle, P., González, M., Lee, Y.-N., & Geen, S. 2022, *A&A*, 663, A6
- Williams, J. P., Blitz, L., & McKee, C. F. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, & S. S. Russell, 97
- Williams, S. J., Fuller, G. A., & Sridharan, T. K. 2005, *A&A*, 434, 257
- Xu, F., Wang, K., Liu, T., et al. 2024, *ApJS*, 270, 9
- Zavala-Molina, R., Ballesteros-Paredes, J., Gazol, A., & Palau, A. 2023, *MNRAS*, 524, 4614
- Zhang, Q., Qiu, K., Girart, J. M., et al. 2014, *ApJ*, 792, 116
- Zhang, Q., Wang, K., Lu, X., & Jiménez-Serra, I. 2015, *ApJ*, 804, 141
- Zhang, S., Zavagno, A., López-Sepulcre, A., et al. 2021, *A&A*, 646, A25

Zhou, J.-W., Liu, T., Evans, N. J., et al. 2022, *MNRAS*, 514, 6038  
 Zhou, J. W., Dib, S., Juvela, M., et al. 2024a, *A&A*, 686, A146  
 Zhou, J.-w., Kroupa, P., & Dib, S. 2024b, *A&A*, 688, L19  
 Zhou, J.-w., Dib, S., & Kroupa, P. 2025, *MNRAS*, 537, 845  
 Zinnecker, H., & Yorke, H. W. 2007, *ARA&A*, 45, 481

- 
- <sup>1</sup> INAF – IAPS, via Fosso del Cavaliere, 100, 00133 Roma, Italy; Istituto Nazionale di Astrofisica (INAF)-Istituto di Astrofisica e Planetologia Spaziali, Via Fosso del Cavaliere 100, 00133 Roma, Italy
- <sup>2</sup> Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185, Roma, Italy
- <sup>3</sup> Physikalisches Institut der Universität zu Köln, Zùlpicher Str. 77, 50937 Köln, Germany
- <sup>4</sup> Institut de Ciències de l’Espai (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, 08193, Bellaterra (Barcelona), Spain
- <sup>5</sup> Institut d’Estudis Espacials de Catalunya (IEEC), 08860, Castelldefels (Barcelona), Spain
- <sup>6</sup> Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
- <sup>7</sup> Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Straße 2, D-69120 Heidelberg, Germany
- <sup>8</sup> Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
- <sup>9</sup> Center for Astrophysics, Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
- <sup>10</sup> Radcliffe Institute for Advanced Studies at Harvard University, 10 Garden Street, Cambridge, MA 02138, USA
- <sup>11</sup> Dipartimento di Fisica, Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy
- <sup>12</sup> SRON Netherlands Institute for Space Research, Landleven 12, 9747, AD Groninger, The Netherlands
- <sup>13</sup> Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands
- <sup>14</sup> Université Paris-Saclay, Université Paris-Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
- <sup>15</sup> Istituto Nazionale di Astrofisica (INAF), Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, Firenze, Italy
- <sup>16</sup> INAF-Istituto di Radioastronomia & Italian ALMA Regional center, Via P. Gobetti 101, 40129 Bologna, Italy
- <sup>17</sup> Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

- <sup>18</sup> Institute of Astronomy and Astrophysics, Academia Sinica, 11F of ASMAAB, AS/NTU No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan
- <sup>19</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
- <sup>20</sup> Faculty of Physics, University of Duisburg-Essen, Lotharstraße 1, 47057 Duisburg, Germany
- <sup>21</sup> Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
- <sup>22</sup> SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield SK11 9FT, UK
- <sup>23</sup> Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile
- <sup>24</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, China
- <sup>25</sup> Dipartimento di Fisica e Astronomia, Alma Mater Studiorum - Università di Bologna Dipartimento di Fisica e Astronomia “Augusto Righi”, Via Gobetti 93/2, 40129, Bologna, Italy
- <sup>26</sup> Research Center for Astronomical computing, Zhejiang Laboratory, Hangzhou, China
- <sup>27</sup> Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
- <sup>28</sup> Dipartimento di Fisica e Astronomia, Università degli Studi di Firenze, Via G. Sansone 1, 50019 Sesto Fiorentino, Firenze, Italy
- <sup>29</sup> University of Connecticut, Department of Physics, 2152 Hillside Road, Unit 3046 Storrs, CT 06269, USA
- <sup>30</sup> National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
- <sup>31</sup> Max-Planck-Institute for Extraterrestrial Physics (MPE), Garching bei München, Germany
- <sup>32</sup> LUX, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 75014 Paris, France
- <sup>33</sup> East Asian Observatory, 660 N. A’ohoku, Hilo, Hawaii, HI 96720, USA
- <sup>34</sup> School of Engineering and Physical Sciences, The University of Lincoln, Brayford Way, Lincoln LN6 7TS, UK
- <sup>35</sup> UK Astronomy Technology center, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK
- <sup>36</sup> INAF – Astronomical Observatory of Capodimonte, Via Moirariello 16, 80131 Napoli, Italy
- <sup>37</sup> School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK
- <sup>38</sup> Center for Data and Simulation Science, University of Cologne, Germany
- <sup>39</sup> Universidad Autónoma de Chile, Av. Pedro de Valdivia 425, Providencia, Santiago de Chile, Chile

## Appendix A: Probability distribution of the projected separation

The measurable separations between cores are the projection on the plane of the sky of the real 3D distances between pair members. To derive these 3D distances, we can only rely on statistical analysis that, once applied to the distribution of projected separations, allows us to infer the 3D distance between cores. This approach is based on the estimation of the probability distribution density of the projected distances  $L_p$ ,  $p(L_p)$  of a pair of objects that are randomly placed in the 3D space. As a first step, we consider only the effect of the projection on these two objects with fixed separation  $L$ , without considering any distribution of their separation. Assuming two points in the 3D space with a fixed separation  $L$ , the projected separation observed along a generic direction is equal to  $L_p = L \sin i$ , where  $i$  is the inclination angle of the vector connecting the two points and the line of sight (see Fig. A.1). For a randomly oriented connecting vector between the two points, the probability of observing a separation  $L_p$  is equal to the area of the annulus  $dS = 2\pi L_p L di$  normalized to the surface of a sphere of radius  $L$ , in other words:

$$dP(L_p) = \frac{dS}{4\pi L^2} = \frac{L_p di}{L^2} = \frac{\sin i di}{2}. \quad (\text{A.1})$$

Differentiating  $L_p = L \sin i$  and doing some algebra yields the following:

$$di = \frac{dL_p}{L \cos i} = \frac{1}{L} \frac{dL_p}{\sqrt{1 - \sin^2 i}}. \quad (\text{A.2})$$

Replacing Eq. A.2 in Eq. A.1, we obtain the probability density distribution function  $p(L_p)$ ,

$$p(L_p) = \frac{dP}{dL_p} = \frac{L_p}{L \sqrt{L^2 - L_p^2}}, \quad (\text{A.3})$$

which is normalized to unity as  $\int_0^L p(L_p) dL_p = 1$ .

The expected value  $\langle L_p \rangle$  of the projected separation  $L_p$  is derived from the probability density distribution function  $p(L_p)$  as

$$\langle L_p \rangle = \int_0^L L_p p(L_p) dL_p = \frac{\pi}{4} L, \quad (\text{A.4})$$

providing the average factor that link the intrinsic 3D separation to the projected one in the case of random orientation for a pair of objects. A similar estimate has been provided by Ishihara et al. (2024), against which this computation provides a statistical distribution. More generally, in the case that the 3D distance between the two points is not fixed, but has a probability distribution function  $P'(L)$ , then the projected separations will be distributed as

$$p(L_p) = \int \frac{1}{L} \frac{P'(L) L_p dL}{\sqrt{L^2 - L_p^2}}, \quad (\text{A.5})$$

whose formulation has the advantage of describing how to statistically deproject the projected probability distribution function

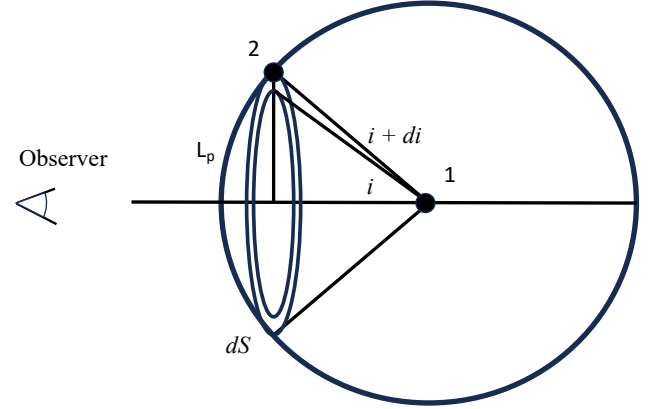


Fig. A.1. Scheme of projection of a pair of randomly oriented objects in 3D space

$P'(L)$  that can be derived from the observed separations. Equation A.5 can be inverted with numerical methods, or by assuming a shape for  $P'(L)$ . An important aspect of Eq. A.5 is that it is extremely unlikely for a unimodal 3D distribution  $P'(L)$  to produce a multimodal 2D projected distribution  $p(L_p)$ .

## Appendix B: Thermal Jeans analysis

The linear analysis of the stability of a non-magnetized, homogeneous, isothermal, and self-gravitating clump shows that there is a critical length above which density perturbations are always unstable against gravitational contraction (Jeans 1902; Mac Low & Klessen 2004). This critical length, and the mass contained in a sphere of this diameter, are usually referred to as the thermal Jeans length,  $\lambda_J^{th}$ , and mass,  $M_J^{th}$ , which, in the absence of turbulent support, depend on the density,  $\rho_{cl}$ , and the internal sound speed,  $c_s$  as

$$\lambda_J^{th} = \left( \frac{\pi c_s^2}{G \rho_{cl}} \right)^{\frac{1}{2}} \quad \text{and} \quad M_J^{th} = \frac{\pi^{\frac{5}{2}}}{6} \frac{c_s^3}{\sqrt{G^3 \rho_{cl}}}, \quad (\text{B.1})$$

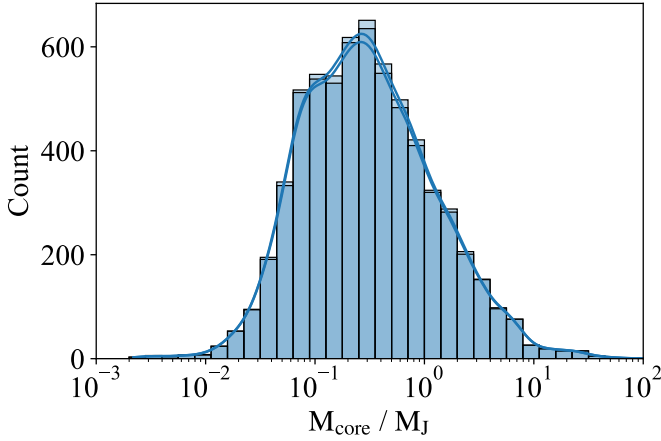
where  $G$  is the gravitational constant.

Expressing the sound speed in terms of the temperature of the clump and assuming a mean molecular weight  $\mu = 2.37$ , which corresponds to a molecular gas with solar composition (Kauffmann et al. 2008; see also our Sect. 3.3.2), these relations become

$$\lambda_J^{th} \approx 4300 \text{ au} \left( \frac{T}{10 \text{ K}} \right)^{\frac{1}{2}} \left( \frac{n}{10^6 \text{ cm}^{-3}} \right)^{-\frac{1}{2}} \quad \text{and} \quad (\text{B.2})$$

$$M_J^{th} \approx 0.28 M_{\odot} \left( \frac{T}{10 \text{ K}} \right)^{\frac{3}{2}} \left( \frac{n}{10^6 \text{ cm}^{-3}} \right)^{-\frac{1}{2}}, \quad (\text{B.3})$$

where  $n$  is the number (volume) density in the system.



**Fig. C.1.** Distribution of the ratio  $M_{\text{core}}/M_J^{\text{th}}$  observed in the ALMAGAL clumps. Light blue: Entire population of 6218 cores; dark blue: Population of cores in clumps with  $N_{\text{cores}} \geq 4$ .

### Appendix C: ALMAGAL core masses versus thermal Jeans mass

The estimated thermal Jeans masses in the ALMAGAL clumps range from  $\sim 0.3$  to  $\sim 14 M_{\odot}$ , a range of masses that is fully observable by the sensitivity of the survey (Molinari et al. 2025), and that is above the completeness limit of  $\sim 0.2 M_{\odot}$  of the ALMAGAL core catalog (Coletta et al. 2025).

Together, the distribution of the ratio  $M_{\text{core}}/M_J^{\text{th}}$  presents a wide spread over about two orders of magnitude, but is heavily skewed towards low values of the ratio. The median value is  $M_{\text{core}}/M_J^{\text{th}} \sim 0.3$  and the first and third quartiles are equal to 0.1 and 0.75, respectively, indicating that most of the ALMAGAL cores have masses considerably lower than  $M_J^{\text{th}}$ . Specifically, 4987 of 6218, equal to  $\sim 80\%$  of the entire sample, have  $M_{\text{core}} < M_J^{\text{th}}$ . Even taking into account the large uncertainties in the core mass and  $M_J^{\text{th}}$ , the result is that a large fraction of cores with sub-Jeans masses are present. Similar results were also found in other surveys, like, for example, the ASHES survey, where a large population of sub-Jeans mass cores has also been detected (Sanhueza et al. 2019; Morii et al. 2023).

Figure C.2 shows the density distribution of the ratio  $M_{\text{core}}/M_J^{\text{th}}$  as a function of  $\Sigma_{\text{cl}}$  and  $L/M$ . We do not observe any strong correlation between the clump surface density  $\Sigma_{\text{cl}}$  and the ratio  $M_{\text{core}}/M_J^{\text{th}}$ . The large majority ( $\sim 70\%$ ) of the detected cores are identified in clumps with  $0.6 \lesssim \Sigma_{\text{cl}} \lesssim 2.5 \text{ g cm}^{-2}$  whose distribution of  $M_{\text{core}}/M_J^{\text{th}}$  does not depend on the clump surface density  $\Sigma_{\text{cl}}$ . They all have a wide distribution that extends from 0.02 to  $\sim 10$ , with a typical peak around  $\sim 0.1 - 0.3$ .

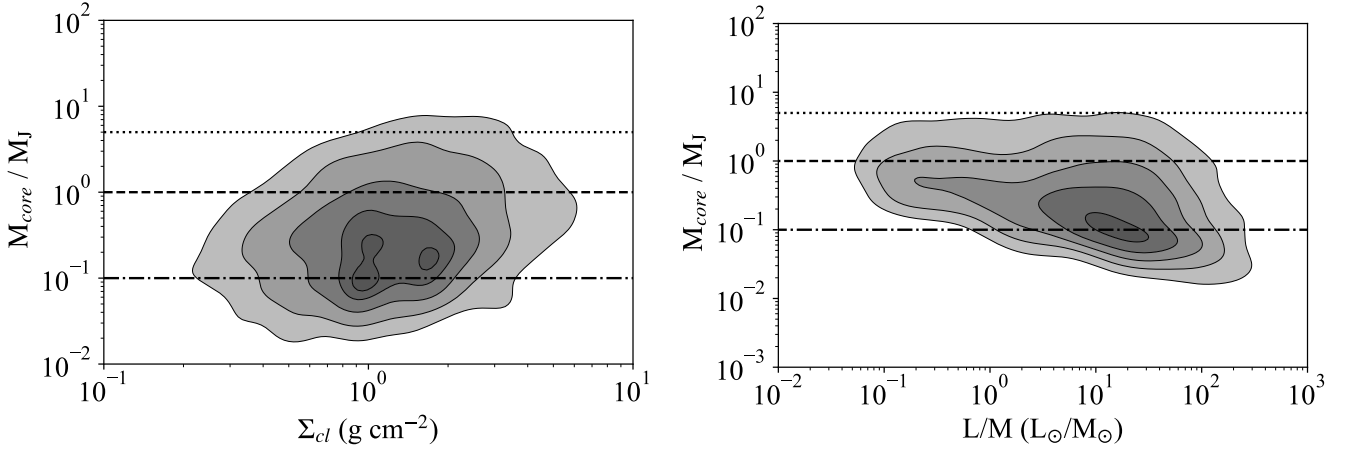
It is more interesting to analyze the distribution of the ratio  $M_{\text{core}}/M_J^{\text{th}}$  as a function of the  $L/M$  ratio. While the ratio  $M_{\text{core}}/M_J^{\text{th}}$  still shows a general wide spread, we find a broad general trend with the  $L/M$  ratio. It is important to note that for clumps in the early stages of evolution, we identify a better agreement of the core masses with the computed thermal Jeans mass. In fact, the distribution of  $M_{\text{core}}/M_J^{\text{th}}$  for cores in clumps with  $L/M \lesssim 1 L_{\odot}/M_{\odot}$  peaks at  $\sim 0.6$ , and shows a relatively narrow distribution with the first and third quartiles equal to 0.3 and 1.33. This agreement is not due to the sample size, as the distribution is derived from 1186 cores, equivalent to  $\sim 20\%$  of the entire sample. On the other hand, for the core separations, we observe that the distribution  $M_{\text{core}}/M_J^{\text{th}}$  shifts its peak towards

lower values and broadens with increasing values of  $L/M$ . The peak gradually shifts to  $M_{\text{core}}/M_J^{\text{th}} \sim 0.1$  for clumps with  $L/M \gtrsim 50 L_{\odot}/M_{\odot}$ . Moreover, we note that while the peak shifts, the distribution becomes progressively more right skewed, presenting a considerable tail at high values of the ratio, and increase its width. This observed behavior is due to the combination of two effects. First, we measured an increase in the estimated  $M_J^{\text{th}}$  with  $L/M$  in the sample of ALMAGAL clumps, which shifts the distribution of the ratio  $M_{\text{core}}/M_J^{\text{th}}$ . The increase in  $M_J^{\text{th}}$  is mainly driven by a general warm-up of the clump material in systems with higher  $L/M$ , due to the heating from the already formed protostars. The clump-averaged volume density,  $n$ , has a minor impact on the estimated  $M_J^{\text{th}}$ , since the observed ALMAGAL clumps have a distribution of  $n$  that spans the same interval independently of  $L/M$ . Secondly, Coletta et al. (2025) found that the core mass distribution extends to higher masses in the clumps with  $L/M$ , which explains the broadening of the distribution of the ratio in these systems. Significant differences between  $M_{\text{core}}$  and  $M_J^{\text{th}}$  are measured in clumps with  $L/M > 2 L_{\odot}/M_{\odot}$ , where the average ratio  $M_{\text{core}}/M_J^{\text{th}}$  is found to be  $\lesssim 0.3$ , indicating that the newly formed cores have masses smaller than the thermal Jeans mass. One possible explanation for the abundance of sub-Jeans masses may reside in the increase in the number of cores with evolution. The formation of additional fragments increases the strength of the tidal forces present in the cluster. Those forces are suggested to also promote gravitational instabilities, with the effect of reducing the critical mass required for local collapse (Zavala-Molina et al. 2023)

### Appendix D: Simulations of the $Q$ parameter in regimes with low number of fragments

The  $Q$  parameter introduced by Cartwright & Whitworth (2004) is usually adopted to characterize the spatial distributions of the components of clusters, since its value distinguishes between systems where the elements are centrally concentrated and those where they are arranged in subclusters, which are well represented by a fractal distribution (Cartwright & Whitworth 2004; Parker 2018; Dib & Henning 2019). Cartwright & Whitworth (2004) used synthetic clusters to identify the intervals of the values of the  $Q$  parameter that characterize these different spatial distributions. They found that  $Q \approx 0.8$  separates between the two aforementioned distributions: centrally concentrated clusters have  $Q > 0.8$ , while subclustered systems have  $Q < 0.8$ . The threshold value of  $Q \approx 0.8$  identifies systems whose elements are uniformly distributed.

However, as correctly pointed out by Avison et al. (2023), this classification was defined for star clusters composed of a large number of components, from several tens to hundreds of elements. On the contrary, the typical cluster of cores observed with ALMA is generally composed of a smaller number of sources per field, up to a few tens in the case of the ALMAGAL survey. Poorly crowded clusters with 2 to 15 detected sources per field were found in the TEMPO survey by Avison et al. (2023), who were the first to investigate the reliability of the  $Q$  parameter applied to the case of small sample sizes. Their result was that  $Q$  parameter can lead to inconclusive results when  $N_{\text{cores}} < 10$ , as it typically returns values below the threshold value of 0.8 independently of the intrinsic spatial distribution of the cluster elements. Although their results suggest that the  $Q$  parameter has a limited applicability to typical ALMA observations, they analyzed only two sets of synthetic clusters with  $N_{\text{core}} = 5$  and 10, as these values represent the degree of fragmentation observed



**Fig. C.2.** Density distribution plot of the core mass normalized to the thermal Jeans mass,  $M_J$ , presented as function of the clump  $L/M$  (right panel) and surface density,  $\Sigma_c$  (left panel). The contours identify the different areas for which, going from the external to the internal, a fraction of 10%, 25%, 50%, 75%, 95% of the entire distribution lies outside. The horizontal lines are reference values of the ratios: unity (dashed), 0.1 (dot-dashed), and 5 (dotted).

in the TEMPO survey. Here, we extend their work to the case of ALMAGAL clumps, where Coletta et al. (2025) detected between 2 and 49 sources, investigating the reliability of the  $Q$  parameter in these cases.

We followed the prescription of Cartwright & Whitworth (2004) to simulate synthetic clusters with centrally condensed and subclustered spatial distributions for their components. We simulated 3D clusters following the former type of distribution as spherical systems composed of  $N$  objects radially distributed with a number density  $N(r) \propto r^{-p}$ . These are simply generated by creating  $N$  triples of polar coordinates  $(r, \theta, \phi)$  with the following formulae:

$$r = \left( \frac{(3-p)R}{3} \right)^{\frac{1}{3-p}}, \quad (\text{D.1})$$

$$\theta = \arccos((2\Theta) - 1), \quad (\text{D.2})$$

$$\phi = 2\pi\Phi. \quad (\text{D.3})$$

Here  $R$ ,  $\Theta$ , and  $\Phi$  are numbers extracted randomly with a uniform distribution between 0 and 1.

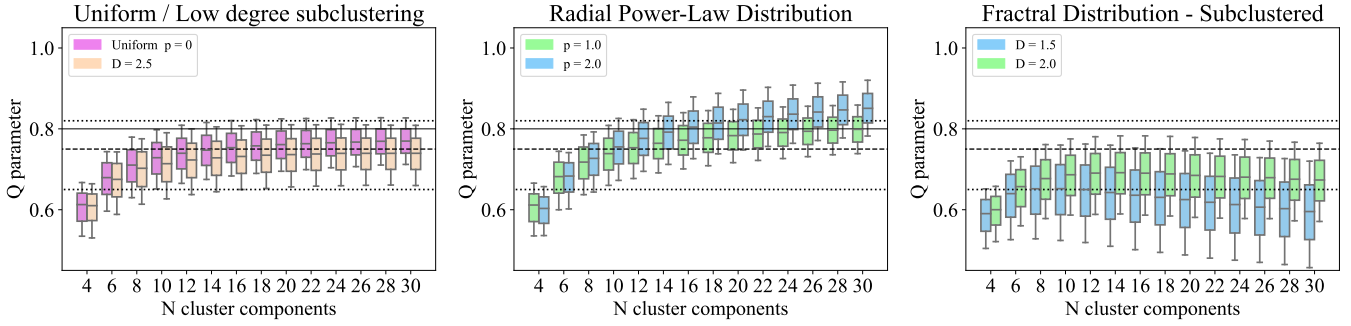
Simulations of subclustered systems are generated by creating fractal distributions with the box method described in Goodwin & Whitworth (2004) and also adopted by Cartwright & Whitworth (2004) and Cartwright et al. (2006). In short, these patterns are created by initially dividing the region of the cluster into eight cubes with each side equal to  $R_c/2$ , placing a first generation of parent seeds in the center of each of these cubes. Each cube is further split into 8 subcubes, whose centers are populated by a first generation of children, which may become fertile. Only fertile elements are kept in each cycle, and, in the next one, those children become parents, spawning a new generation. The iterative process creates the fractal distribution for the cluster that contains a larger or smaller number of substructures depending on the probability function adopted for the fertility. This probability is set by the fractal dimension  $D$  as  $2^{(D-3)}$ , allowing fewer children to mature for lower values of the fractal dimension  $D$ , and thus generating a system with more substructures. Moreover,

to avoid extremely regular structures, we introduce a random displacement of amplitude  $0.01 \times R_c$  to the position of the spawning children. Cycles are iterated until a number of cluster components larger than the target  $N$  are generated, then we randomly select  $N$  of these components to define a single synthetic cluster realization.

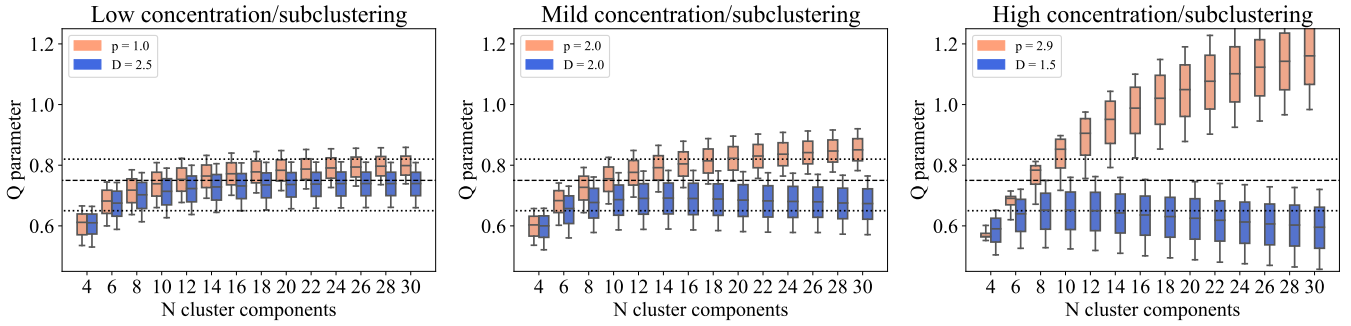
We ran seven sets of simulations: three for each of two types of distribution sampling with  $p = 1, 2$ , and  $2.9$  (low, medium, and high concentrations) and with  $D = 1.5, 2.0$ , and  $2.5$ , (high, mild, and low level of subclustering), respectively, and one where the components are uniformly distributed in the cluster, which is simply generated adopting  $p = 0$  (an approximately similar result is obtained with  $D = 3$ ). Each set is composed of a group of synthetic clusters computed varying  $N_{cores}$  from 4 to 30 with a step of 2. Our simulations are designed over a range of  $N_{cores}$  smaller than that of the ALMAGAL survey, since higher number of fragments rarely occurs in our observations. We computed  $10^5$  synthetic realizations for each value of  $N_{cores}$ , projecting the 3D spatial distribution in two dimensions and then deriving the distribution of the  $Q$  parameter. The large number of realizations is chosen to carefully determine the probability distribution function of the  $Q$  parameter. To speed up the calculation of the  $10^5$  realizations, we initially simulated  $10^4$  systems with three times the intended number of sources  $N_{cores}$ , and then, for each system, we selected  $N_{cores}$  elements at random ten times.

The simulation campaign is composed of 1.1M synthetic clusters, and the results are shown in Fig. D.1 where we report how the probability distribution functions (PDFs) of the  $Q$  parameter, represented as boxplots at the 10, 25, 75, and 90th percentiles, vary as a function of the number of cluster elements  $N$  for different types of spatial distribution. For each measured  $Q$  parameter, these PDFs could be theoretically adopted to assess the likelihood of a system being a concentrated cluster or a subclustered structure. However, we prefer to use these results to define the ranges of  $Q$  that are most likely associated with one or the other type of spatial distribution.

The plots in Fig. D.1 show that in the case of very small populations ( $N_{core} < 8$ ) the measured PDFs of different spatial distributions strongly overlap, confirming the result of Avison et al. (2023) that the  $Q$  parameter is highly unreliable in these systems. The situation improves for  $N_{core} \geq 10$ , above which



**Fig. D.1.** Boxplots representing the distributions of the  $Q$  parameter derived from the simulated clusters as a function of the number of cluster components. Each box indicates the median, the first, and the third quartile, while the whiskers extend from the 10th to the 90th percentile of the distribution of measured  $Q$ . The left panel compares the results of the simulations with uniform distribution, obtained with  $p = 0$ , and with the lowest level of subclustering in the simulation campaign, corresponding to a fractal dimension  $D = 2.5$ . The central panel presents the results for clusters with a radial distribution of their components, for the two cases with exponents  $p = 1$  and  $p = 2$ . The right panel presents the results for clusters with a fractal distribution of their components with two different levels of subclustering, determined by a fractal dimension  $D = 2.0$  (mild subclustering) and  $D = 1.5$  (high subclustering). The reference values for  $Q$  discussed in the text are reported as horizontal lines, corresponding to  $Q = 0.65$  and  $Q = 0.83$  (dotted lines),  $0.75$  (dashed line), and the threshold  $Q = 0.8$  introduced by Cartwright & Whitworth (2004, thin line).



**Fig. D.2.** Same as Fig. D.1, but showing the direct comparison of the distributions of the  $Q$  parameter computed from the sets of synthetic clusters with the components distributed with a radial power law and with a fractal distribution. The comparison is shown for sets with low, mild, and high concentration degrees and level of subclustering.

the PDFs of the two types of spatial distributions start to progressively shift one with respect to the other, although they still retain a certain degree of overlap. For this reason, we focus the discussion on the results of cases where  $N_{core} \geq 10$ .

These plots show that the criteria introduced by Cartwright & Whitworth (2004) cannot be directly applied to the case of small sample sizes. In particular, the threshold limit  $Q \approx 0.8$ , which should also identify the case with uniform distribution, does not reflect the results obtained from our simulation campaign. Our simulations of the uniform distribution indicate that the median of the probability distribution functions slightly increases from  $\sim 0.73$  to  $0.76$  when the number of cluster components increases. More generally, the PDFs in this limit case are confined in an interval equal to  $0.65 \leq Q \leq 0.82$ . Values close to  $\sim 0.8$  are only found in the upper tails of the distribution, for values greater than the third quartile, while the average of the medians is equal to  $Q = 0.75$ . The simulation with uniform distribution is a limit case for the two types of spatial distribution analyzed; therefore, its statistical properties can be adopted as a reference to discuss the results of the other simulations. If we look at the median of the PDFs, we notice that the clusters with centrally concentrated distribution have  $Q \geq 0.75$ , while those with mild and high levels of subclustering have  $Q < 0.75$ .

However, adopting a defined value as a threshold limit does not take into account that PDFs have widths that extend over wide intervals of  $Q$ , and often overlap for different simulated spatial distributions. Figure D.2 shows the PDFs derived from

the simulation campaign, comparing the sets of simulations for increasing degrees of concentration or subclustering. The two different spatial distributions may both produce systems with  $Q$  values between  $0.65 \leq Q \leq 0.8$ , but outside this interval it can be determined with a reasonably high confidence whether the cluster components are distributed as a radial power-law or as subclusters. Simulations with subclustering have PDFs whose 90th percentile is always lower than  $0.8$ , and systems with  $Q > 0.8$  are unlikely to have a fractal distribution. Similarly, since the 10th percentile of the probability distribution functions of the clusters with radial distribution is always higher than  $0.65$ , we expect that when  $Q < 0.65$  is measured, these systems are most likely to present a level of subclustering.