Astronomy & Astrophysics

ALMAGAL

I. The ALMA evolutionary study of high-mass protocluster formation in the Galaxy: Presentation of the survey and early results

S. Molinari[1,*], P. Schilke[2], C. Battersby[3], P. T. P. Ho[4], Á. Sánchez-Monge[5,6], A. Traficante[1], B. Jones[2], M. T. Beltrán[19], H. Beuther[7], G. A. Fuller[9,2], Q. Zhang[13], R. S. Klessen[10,11,13,14], S. Walch[2,12], Y.-W. Tang[4], M. Benedettini[1], D. Elia[1], A. Coletta[1,47], C. Mininni[1], E. Schisano[1], A. Avison[8], C. Y. Law[19], A. Nucara[1,16], J. D. Soler[1], G. Stroud[9], J. Wallace[3], M. R. A. Wells[7], A. Ahmadi[17], C. L. Brogan[15], T. R. Hunter[15], S.-Y. Liu[4], S. Pezzuto[1], Y.-N. Su[4], B. Zimmermann[2], T. Zhang[2], F. Wyrowski[18], F. De Angelis[1], S. Liu[1], S. D. Clarke[4], F. Fontani[19,20,21], P. D. Klaassen[22], P. Koch[4], K. G. Johnston[23], U. Lebreuilly[25], T. Liu[26], S. L. Lumsden[24], T. Moeller[2], L. Moscadelli[19], R. Kuiper[28], D. Lis[30], N. Peretto[32], S. Pfalzner[29], A. J. Rigby[24], P. Sanhueza[31,33], K. L. J. Rygl[27], F. van der Tak[35,36], H. Zinnecker[37], F. Amaral[2], J. Bally[38], L. Bronfman[44], R. Cesaroni[19], K. Goh[2], M. G. Hoare[24], P. Hatchfield[30], P. Hennebelle[25], T. Henning[7], K.-T. Kim[40,41], W.-J. Kim[2], L. Maud[42], M. Merello[44,45], F. Nakamura[33], R. Plume[43], S.-L. Qin[46], B. Svoboda[34], L. Testi[39], V. S. Veena[2,18], and D. Walker[9]

(Affiliations can be found after the references)

ABSTRACT

Context. A large fraction of stars form in clusters containing high-mass stars, which subsequently influences the local and galaxy-wide environment.

Aims. Fundamental questions about the physics responsible for fragmenting molecular parsec-scale clumps into cores of a few thousand astronomical units (au) are still open, that only a statistically significant investigation with ALMA is able to address; for instance: the identification of the dominant agents that determine the core demographics, mass, and spatial distribution as a function of the physical properties of the hosting clumps, their evolutionary stage and the different Galactic environments in which they reside. The extent to which fragmentation is driven by clumps dynamics or mass transport in filaments also remains elusive.

Methods. With the ALMAGAL project, we observed the 1.38 mm continuum and lines toward more than 1000 dense clumps in our Galaxy, with $M \geq 500\,M_\odot$, $\Sigma \geq 0.1\,\mathrm{g\,cm^{-2}}$ and $d \leq 7.5$ kiloparsec (kpc). Two different combinations of ALMA Compact Array (ACA) and 12-m array setups were used to deliver a minimum resolution of ∼1000 au over the entire sample distance range. The sample covers all evolutionary stages from infrared dark clouds (IRDCs) to HII regions from the tip of the Galactic bar to the outskirts of the Galaxy. With a continuum sensitivity of 0.1 mJy, ALMAGAL enables a complete study of the clump-to-core fragmentation process down to $M \sim 0.3\,M_\odot$ across the Galaxy. The spectral setup includes several molecular lines to trace the multiscale physics and dynamics of gas, notably CH_3CN, H_2CO, SiO, CH_3OH, DCN, HC_3N, and SO, among others.

Results. We present an initial overview of the observations and the early science product and results produced in the ALMAGAL Consortium, with a first characterization of the morphological properties of the continuum emission detected above 5σ in our fields. We used "perimeter-versus-area" and convex hull-versus-area metrics to classify the different morphologies. We find that more extended and morphologically complex (significantly departing from circular or generally convex) shapes are found toward clumps that are relatively more evolved and have higher surface densities.

Conclusions. ALMAGAL is poised to serve as a game-changer for a number of specific issues in star formation: clump-to-core fragmentation processes, demographics of cores, core and clump gas chemistry and dynamics, infall and outflow dynamics, and disk detections. Many of these issues will be covered in the first generation of papers that closely follow on the present publication.

Key words. stars: formation – stars: protostars – ISM: clouds – evolution – HII regions – submillimeter: ISM

1. Introduction

Most stars form in clusters (Lada & Lada 2003; Portegies Zwart et al. 2010; Adams 2010; Adamo et al. 2020) and about 50% of them, including the Sun itself (and the fraction probably was even higher in the early Universe), form in very rich clusters of at least 1000 stars, containing at least one 10 $M_\odot$ star. High-mass

stars ($M \geq 8\,M_\odot$) influence their immediate environment through gravitational (ejection of stars from the cluster, disk truncation, etc.), mechanical (winds, outflows), radiative interaction (e.g., radiative heating, photo-ionization, and radiation pressure), and, eventually, through their supernovae explosions. Hence, the formation process of low-mass stars and their associated planets in such clusters will vastly differ from star formation in isolation or small clusters. However, the latter scope is where most low-mass star formation studies have taken place so far, mostly

* Corresponding author; sergio.molinari@inaf.it

because such objects are close by and easier to study. However, they are not representative of the most significant fraction of star formation.

High-mass stars dominate the energy input into the ISM through feedback processes, not only locally, but also on galactic scales (see, for example, Veilleux et al. 2005; Bolatto et al. 2013). They also enrich the ISM with heavy elements, which, in turn, influences subsequent events of star formation (Klessen & Glover 2016). Yet, their formation processes differ significantly from low-mass stars: while the Kelvin-Helmholtz timescale of low-mass stars is significantly longer than the time required to assemble them, for any reasonable accretion rate, it is shorter for high-mass stars. Hence, high-mass stars will continue accreting up to and even after reaching the main sequence (e.g., Maeder & Meynet 2000; Keto & Wood 2006; Zinnecker & Yorke 2007). The combination of being deeply embedded in dusty cores through most of their evolution, their scarcity and correspondingly high average distance, together with evolution in clusters with many sources close together, make these observations challenging.

Cluster formation involves cores (linear scales ≤0.1 pc) forming inside molecular clumps (linear scales of 0.5–1.0 pc), as indeed observed at high spatial resolution at mm wavelengths (Zhang et al. 2015; Beuther et al. 2018; Svoboda et al. 2019; Sanhueza et al. 2019; Traficante et al. 2023; Avison et al. 2023). A full understanding of the fragmentation process and its role in allowing high-mass stars to collect material up to their final mass is still elusive. Furthermore, thermal Jeans (1902) mass fragmentation leads to low-mass objects (Rees 1976; Larson 1985) that are less massive than the observed masses of high-mass stars (M≥8 M$_\odot$). Mechanisms to stop fragmentation and delay collapse, allowing the collection of enough gas in existing fragments to form massive stars, are a possible solution. Turbulent support (e.g., McKee & Tan 2003) or magnetic and radiation feedback (e.g., Krumholz et al. 2009; Commercon et al. 2011) have recently been invoked. In this model family (hereafter monolithic collapse), the final stellar mass of the emerging stars is pre-assembled in the cores, and we would expect to find high-mass monolithic prestellar cores. Yet, the search for high-mass protostellar clumps without any sign of star formation has only resulted in very few candidates (e.g., Tackenberg et al. 2012; Beuther et al. 2015; Nony et al. 2018; Motte et al. 2018); so, this is unlikely to be a dominant path to high-mass star formation.

In the more dynamical scenario of competitive accretion, cores compete (Bonnell et al. 2007) for gas from the cloud mass reservoir that is not initially local to the core itself (Klessen 2000; Bonnell & Bate 2006; Peters et al. 2011; Girichidis et al. 2012), whereas high-mass monolithic prestellar cores should not feasibly exist (Smith et al. 2009). Infall motions would be dominated by accretion from the cloud onto the core. Indeed, observations reveal large-scale infall motions in massive star-forming regions, showing that high-mass clumps are not isolated from the cloud mass reservoir and they are shown, instead, to globally accrete while star formation is internally ongoing (e.g., Wu & Evans 2003; Rygl et al. 2013; Klaassen et al. 2012; He et al. 2015; Traficante et al. 2017, 2020; Contreras et al. 2018) also via continuous mass flow along filaments to stars forming at filamentary hubs (Peretto et al. 2013; Chen et al. 2019; Wells et al. 2024). Theoretical studies have shown that infall motions are crucial both for initiating the formation of high-mass stars and in subsequent evolutionary stages for maintaining accretion flows to increase the stellar mass (e.g., Jijina & Adams 1996; Yorke & Sonnhalter 2002; Gong & Ostriker 2009; Peters et al. 2010; Vázquez-Semadeni et al. 2019).

To make fundamental progress in our understanding of the formation of high-mass stars and their surrounding clusters, we want to investigate two key issues in a statistically significant way: (1) what physical processes govern the fragmentation of cluster-forming clumps and how they evolve with time; and (2) how cores gain mass and how this process is influenced by internal feedback from the cores into the clump gas.

So far, ~300 dense and massive clumps have been observed at interferometric spatial resolution in total over fewer than a dozen different observing programs. In particular, early-stage, 70 µm-dark, or IRDC-like targets have received considerable attention. Svoboda et al. (2019) targeted 12 such massive clumps at a resolution of ~3000 au, revealing fragmentation with separations comparable to the thermal Jeans length, as previously proposed by Palau et al. (2015). Likewise, ASHES Sanhueza et al. (2019) observed a sample of similar size at slightly lower resolutions (~4500 au), confirming fragmentation in these early stages of evolution, Jeans-compatible fragmentation lengths, and also suggesting sub-clustering in the distribution of the fragments. Anderson et al. (2021) observed with ALMA six hub-filament systems in IRDCs at resolutions ~6000 au; combining them with data from 29 clumps previously observed by Csengeri et al. (2017) offered evidence of clump-fed accretion. A similarly sized sample of IRDCs was recently studied by Rigby et al. (2024), displaying gas kinematics that is also consistent with the latter scenario.

At the other end of the evolutionary path, Beuther et al. (2018) observed 20 relatively more evolved massive clumps with luminosities in excess of 10^4 L$_\odot$ using NOEMA with high resolutions (~1000 au), revealing various levels of fragmentation, again compatible with thermal Jeans length and typical fragment separations showing no dependence on core masses. More recently, ATOMS (Liu et al. 2020) targeted 146 evolved UC/HC-HII regions, while ASSEMBLE (Xu et al. 2024) observed 11 evolved massive clumps at ~2000 au resolution. Similarly, Ishihara et al. (2024) observed 30 hot cores at ~1000 au resolution confirming that thermal Jeans fragmentation may be at work. Other surveys have a wider coverage in the evolutionary stage of targeted clumps, with TEMPO (Avison et al. 2023) reporting (with resolutions ~3000–4000 au) Jeans-compatible fragmentation distance only in a fraction of the clumps; or, as in SQUALO (Traficante et al. 2023), suggesting that the mass and minimum distance between fragments, respectively, increase or decrease with evolution.

The significance and robustness of these results are hampered by the relatively low number of targets in each study, as well as the different selection criteria that prevent trustworthy conclusions from studies of composite samples that, in principle, could be assembled. In addition, the different frequency setups and different linear resolutions used make a quantitative comparison difficult. Large ALMA projects like ALMA-IMF (Motte et al. 2022) partially overcome these limitations by imaging 15 well-known large star formation complexes (e.g., the W43 Galactic starburst), harbouring regions of very different mass and evolution with extensive spectroscopic coverage, to address the issue of the emergence of the IMF.

Answering the fundamental questions outlined above requires a number of crucial observables to be measured in a statistically significant way and compared to predictions from numerical simulations over a wide variety of evolutionary stages and Galactic environments: i) the spatial distribution of dense cores as a function of mass within dense clumps; ii) their fragmentation properties, such as average distances and their relation to Jeans masses or filament fragmentation scales; iii) their
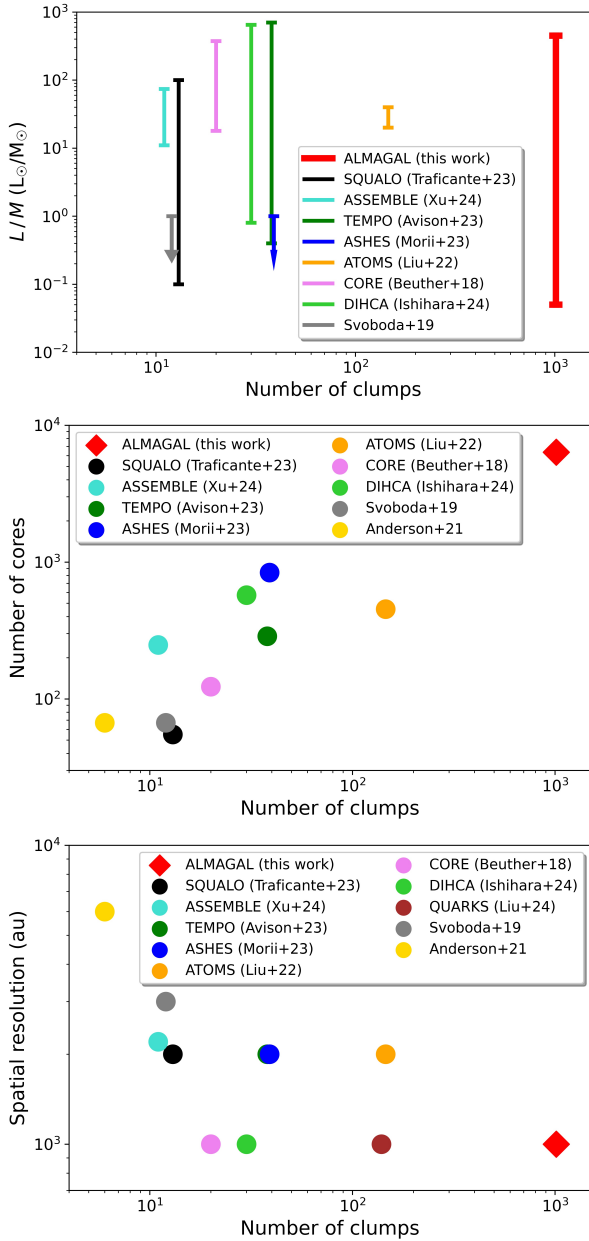
Fig. 1. ALMAGAL in the context of other interferometric surveys of star-forming regions in terms of number of target clumps vs. evolutionary stage (top), number of detected fragments (middle, from Coletta et al. 2025), and spatial resolution achieved (bottom). The panels show surveys for which the necessary information was immediately available in the papers and for which observations consist of single pointings or relatively small mosaics. For the latter reason, ALMA-IMF is not included in the figure as the program does relatively large maps of 15 star-forming complexes.

evolution both in morphology and in total number; iv) the temperature and density distribution of gas in the clumps; v) the velocity field of the clumps gas and the presence of global infall motions; and vi) the dynamical state of the compact fragments.

The ALMAGAL Large Program was specifically designed to deliver this key science by mapping more than 1000 intermediate and high-mass dense clumps with ALMA in band 6 (Sect. 3) at a spatial resolution of 1000 au. Figure 1 shows that ALMAGAL is a game-changer with respect to coupling the ability to map

the variance of physical, evolutionary, and environmental conditions of the targets (Fig. 2) with an unprecedented resolution and statistical significance.

The paper is structured as follows. In Sect. 2, we summarize the characteristics of the ALMAGAL observations, deferring a much more thorough discussion to the companion paper by Sanchez-Monge et al. (2025). In Sect. 3, we discuss in detail the properties of the observed clumps and the process of revision of source distances following the ALMAGAL observations and the homogeneization of the way in which physical parameters (e.g., masses and luminosities) are derived. In Sect. 4, we present a characterization of the continuum emission detected with ALMA in the context of clump-integrated properties (Sect. 4.1) and in terms of the morphology of the emission (Sect. 4.2). In Sect. 5, we outline the potential of the ALMAGAL science in a number of areas, anticipating the dedicated papers that are anticipated in the near future. Finally, in Sect. 6, we summarize our conclusions.

2. Observational setup

The ALMA Band 6 217 to 220 GHz frequency range is ideally suited to provide access to the J=2–1 lines of ^{13}CO and $C^{18}O$ (for column density and velocity structure), SiO (5–4) (to trace outflows), three H_2CO lines (temperature and density, infall and outflow tracers), CH_3OH (temperature estimates), the K-ladder of the CH_3CN (12–11) and $CH_3^{13}CN$ (12–11) lines (temperature, velocity structure), and several other species (DCN (3–2), OCS (19–18), HC_3N(24–23) etc.). The correlator was configured to produce two 1.875-GHz wide windows, providing ~ 1.4 km s^{-1} resolution, plus two higher resolution windows with a ~ 0.34 km s^{-1} resolution. To be able to address our science goals, it is imperative to recover signal from the minimum target 1000 au scale up to the clump-size scale. To achieve this goal in an optimal way, given the large distance range of our sources, all the targets have been observed in single-pointing, with the ACA plus two configurations of the 12-m array with two different antenna configuration combinations, depending on the source distances; the C-5/C-2/ACA setup was chosen for sources with d<4.7 kpc (hereafter the "near" subsample, amounting to 538 objects), while the C-6/C-3/ACA setup was adopted for sources with d>4.7 kpc (the "far" subsample, with 479 objects). In band 6, these combinations provide maximum angular resolutions of 0.13″ (C-6 for the "far" subsample) and 0.24″ (C-5 for the "near" subsample), along with a largest recoverable scale of about 30″. More details about the observations setup are given in the companion paper (Sanchez-Monge et al. 2025).

We will refer to as "7M" the data taken with the ACA, "TM2" the data taken in short baseline (hence lower resolution) configurations of the 12-m array (C-2 for the "near" sample, and C-3 for the "far" sample), and "TM1" the data taken in long baseline (hence higher resolution) configurations (C-5 for the "near" sample, and C-6 for the "far" sample). A fundamental step in data processing was the production of continuum images and spectral cubes by jointly deconvolving the combined data taken with ACA 7M array and the different 12-m array configurations. In a first set of products, we combined ACA and the TM2 into what is called here and in all ALMAGAL papers the 7M+TM2 products. In a subsequent step, the TM1 data were also combined to obtain the full resolution 7M+TM2+TM1 products.

We required a 0.1 mJy continuum sensitivity for all targets irrespective of clump properties, which should (in principle) allow for detections at the 3σ level of compact cores with a
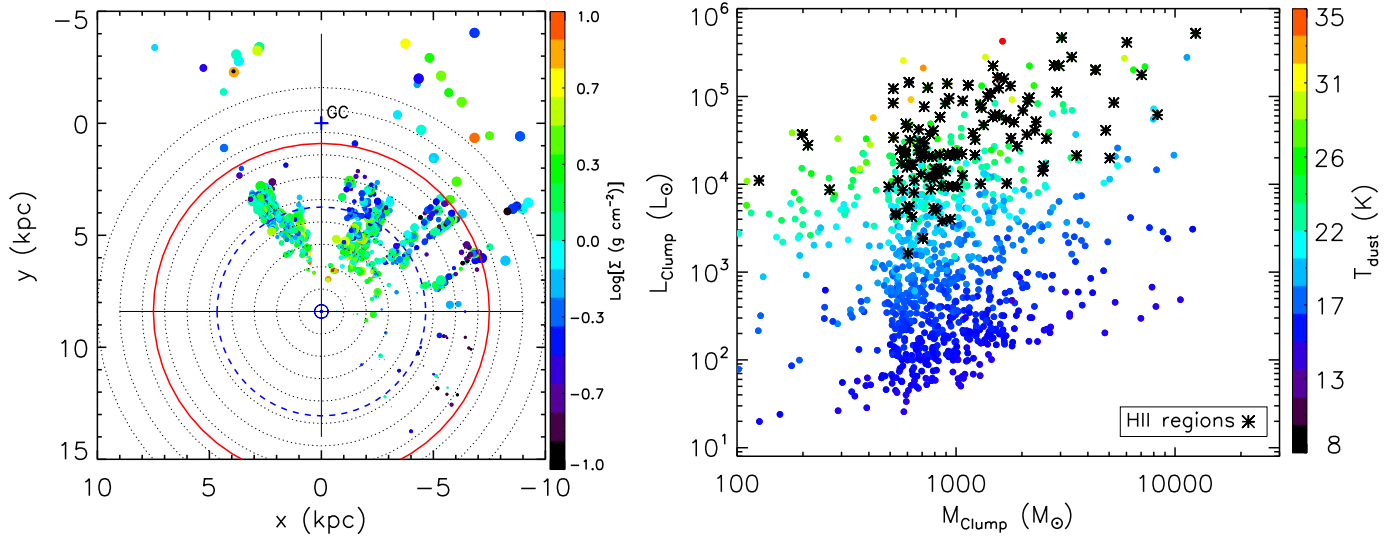
Fig. 2. Left panel: Galactic distribution of ALMAGAL target clumps are shown on the left, with a symbol size of $\propto \mathrm{Log}(M_c)$ (clump mass) and color coded by Σ_c (surface density), as determined from the Hi-GAL data (non-beam-deconvolved), reflecting the updated distances and physical parameters described in Sect. 3.2 and reported in Table 1, available at the CDS. Dotted circles represent 1 kpc distance interval centered on the Sun and the plus symbol marks the location of the Galactic Center. The blue dashed circle marks the distance threshold used to split the target sample among the two ALMA antenna configurations designed to provide a minimum 1000 au linear resolution. The red circle marks the 7.5 kpc distance originally used as an upper limit for source selection. Right panel: L_{bol}/M_{clump} plot for the 1017 selected clumps (color-coded by T_{dust}) is shown on the right. Asterisks indicate Hi-GAL sample sources associated with HII regions from the surveys of CORNISH (Purcell et al. 2013), and CORNISH-S (Irabor et al. 2023).

mass of $\geq 0.3\ M_\odot$ (assuming dust temperature of 20K and opacity from Ossenkopf & Henning 1994 with index $\beta = 1.75$) at a distance of 7.5 kpc; this is sufficient to reliably sample the peak of the core mass function in a star-forming region such as Aquila (Konyves et al. 2010), but at up to 7.5 kpc from the Sun. This estimate is based on pure sensitivity requirements, while the effective detection of compact objects in interferometric maps critically depends on local background conditions and residuals from the CLEANing procedures. More details are given in Sanchez-Monge et al. (2025); Coletta et al. (2025). The ALMAGAL Large Program (2019.1.00195.L) was approved in ALMA Cycle 7, for a total granted observing time of 117.7 hours for all the 12-m array configurations, and 88.4 hours for the ACA observations (Sanchez-Monge et al. 2025).

3. The ALMAGAL sample and observations

3.1. Initial selection and survey planning

The ALMAGAL sample initially consisted of 1017 targets with declination $\delta \leq 0°$, distributed from the near tip of the Galactic bar to the third quadrant, spanning a large range of clump masses, evolutionary stages and Galactocentric distances.

The larger fraction of the sample (915 objects) was selected from the complete catalog of dense clumps from Elia et al. (2017) based on the Hi-GAL 70–500 μm photometry (Molinari et al. 2016b) complemented with ancillary data at $\lambda \leq 24$ μm and $\lambda > 500$ μm and with a distance determined after cross-comparison with an extensive suite of CO surveys of the Galactic plane, extinction maps, and HI line profile analysis, following Mege et al. (2021). The following selection criteria were adopted: i) distance <7.5 kpc from the Sun to be able to resolve the target 1000 au spatial scale with the selected configurations setup (Sect. 2); ii) clump masses >500 $M_\odot$ in the inner and 250 $M_\odot$ in the outer Galaxy; and iii) surface densities $\Sigma \geq 0.1$ g cm^{-2} threshold that is critical for high-mass star

formation (Kauffmann & Pillai 2010; Krumholz et al. 2014; Tan et al. 2014; Traficante et al. 2020). These selection criteria, adopted at the time of the proposal, are no longer 100% fulfilled, following a revision of target physical parameters using the ALMAGAL data itself (see Sect. below). Overall, more than 90% of targets fulfill the initial selection requirements. Target sources are spread over 3–14 kpc in Galactocentric distance from the near tip of the Galactic bar to the outskirts of the Milky Way (Fig. 2, left panel). They span the full evolutionary path from the IRDC to the HII region stage (Fig. 2, right panel). They reside in very different environments, such as arm and inter-arm regions, and proximity to triggering agents, such as OB associations and expanding HII region bubbles.

Sources in the inner 20° around the Galactic Center were excluded because this range subtends the central regions of the Galaxy. These regions are heavily affected by non-circular motions that make distance estimates unreliable (e.g., Hunter et al. 2024). In addition, our main goal is to characterize the variance of the environmental conditions in the Galactic disc. In this sense, the central regions show peculiar conditions not representative of the disc and they are typically studied with dedicated programs such as ACES (Nonhebel et al. 2024).

We visually inspected the PACS 70–160 μm and the SPIRE 250–350–500 μm *Herschel* maps of all candidate targets to confirm the presence of compact emission at the position of the catalog source. The position of the 250 μm peak was adopted for the ALMA observations. From the evolutionary viewpoint, both the L/M and the shape of the SED for $\lambda \leq 70$ μm have been used as a broad evolutionary classification of the clumps (Molinari et al. 2008; Duarte-Cabral et al. 2013; Molinari et al. 2016a, 2019; Traficante et al. 2018; Merello et al. 2019; Wells et al. 2022). The values of clump L/M go from ∼0.05, which is typical of early-stage IRDC-like clumps, to ∼450, which is common to IR-bright clumps hosting actively forming protostellar objects often associated with HII regions (e.g., Cesaroni et al. 2015; Elia et al. 2021).

To extend the coverage of the sample toward relatively more evolved H II regions, we also included 98 objects drawn from the Red Midcourse Space Experiment Source (RMS) catalog of star-forming regions (Lumsden et al. 2013), using the same distance cut and with luminosity in excess of 3000 $L_\odot$. The RMS initial selection being based on the the mid-IR 8–21 μm wavelength range allows it to be more representative of relatively late stages; the sources selected do indeed display $1 \leq L/M \leq 200$. The distribution of the final ALMAGAL sources sample in the Galaxy and in the L versus M evolutionary plot is presented in Fig. 2, where the asterisk-symbols mark the location of the target clumps with H II counterparts (Purcell et al. 2013; Irabor et al. 2023) in the CORNISH and CORNISH-S 5 GHz radio continuum surveys and occupy mostly the region above $L/M = 10$. This confirms that the ALMAGAL sample spans the entire path from IRDC-hosted clump to the H II region phase.

During the analysis stage, after the observations were taken, it was realized that the template WR stellar system η Car, which has the target ID '653755' in the name of the data files in the ALMA Archive was included by error in the sample. We also overlooked that three fields (with target IDs '615590', 'G348.7342-01.0359B', and 'G323.7410-00.2552C' for reference in the ALMA Archive) had coordinates closer than 10″ to other targets in the sample and they are therefore considered to be duplications. Since they have larger noise levels compared to the fields they duplicate, we removed them from the ALMAGAL sample data considered in the rest of the paper and, thus, from further scientific analysis (amounting to 1013 targets). Figure 2 reports the Galactocentric distribution and the L-M plot for the sample targets, with distances and physical parameters updated based on the new ALMAGAL observations, as explained in the next subsection below.

3.2. Revision of ALMAGAL targets physical properties

A critical aspect in ALMAGAL science is relating the physical properties of ALMA detected cores, as well as the properties of the dense gas at all sampled spatial scales, to the physical properties of the parent dense clumps (Coletta et al. 2025; Mininni et al. 2025, Elia et al., in prep, Jones et al., in prep.). To this end, we decided to revise and update the target clump distances (also based on ALMAGAL itself) and physical properties, for three reasons:

- The methods by which the distance to the targets is estimated may have changed due to new evidence. For the fraction of the sample selected from the Hi-GAL survey, a new source catalog with new distance estimates and revised physical properties was released (Elia et al. 2021). For the RMS-based selected targets, their properties are constantly updated on the project's website[1].
- The v_{LSR} of the bulk gas traced by high-critical density lines in the ALMAGAL spectral cubes provide a much more reliable input for kinematic distance estimates, especially in cases where CO (even with its rarer isotopologs) shows multiple gas components along the line of sight (Sect. 3.2.1).
- A meaningful and reliable science analysis of the ALMAGAL data requires that clumps physical properties (in particular mass and luminosity) are estimated in a homogeneous way for the Hi-GAL and the RMS-based fractions of the sample (Sect. 3.2.2).

We report these updates in the following subsections.

---

[1] http://rms.leeds.ac.uk

3.2.1. Revision of targets distance

Assigning a heliocentric distance to a source detected in 2D thermal continuum images is a mandatory critical step for any subsequent physical parameter estimate. The typical backbone workflow of any distance determination recipe can be summarized as follows (Urquhart et al. 2018; Mege et al. 2021) in order of precision and priority: i) check source position against literature catalogs of maser or stellar parallax that offer the most reliable of all estimates; ii) correlate source position and velocity against literature catalogs of optical, radio, or absorption line H II regions, or in YSOs where the distance has already been determined via, for instance, optical spectroscopy; iii) correlate source position with submm/radio molecular line surveys, use a Galactic rotation curve to derive the distance and (in cases where this places the source in the inner Galaxy) to try to solve the kinematic distance ambiguity (KDA) using additional evidence such as an association with IRDCs, H I absorption features, and so on. A similar methodology was used to derive distances for the Hi-GAL and RMS catalogs, with only small differences.

To obtain a homogenized set of properties for the sample of target clumps and to take advantage of the ALMAGAL spectroscopy that offers the ideal database to improve existing kinematic distance estimate, in particular, for those objects (the majority), where only CO and its isotopologues were previously used in the literature, we revised the distances by measuring the v_{LSR} from several high critical density transitions and applied the final module of the Mege et al. (2021) algorithm (previously used for deriving distances for the Hi-GAL catalog) that converts v_{LSR} into kinematic distance. The details of this work will be reported in Benedettini et al. (in prep.). In summary, for each ALMAGAL source, we extracted subcubes from the ALMAGAL 7M datacubes over large velocity ranges (± 180km s^{-1}) centered on CH$_3$OH, H$_2$CO, DCN (3–2), SO (6$_5$–5$_4$), HC$_3$N(24–23), SiO (5–4), C^{18}O (2–1), and ^{13}CO (2–1) lines. For each line, the cubes were co-added over the emitting region, defined by selecting the spatial pixels where the moment zero map is above three times its noise level, to obtain integrated spectra that were fed to an automatic 3σ-clipping line detection and Gaussian fitting algorithm. The v_{LSR} of the targets were then derived averaging the fitted line centers of all the detected lines with a good fit among the transitions with critical density higher than 10^5 cm^{-3}; namely, CH$_3$OH, H$_2$CO, DCN (3–2), SO (6$_5$–5$_4$), HC$_3$N(24–23) and SiO (5–4). C^{18}O (2–1), and ^{13}CO (2–1) were used only for 207 targets where none of the other higher critical density lines were detected. We assigned an uncertainty to the derived v_{LSR} of 1.4 km s^{-1}, which is the spectral resolution of the data. The formal uncertainty obtained from averaging the v_{LSR} of the individual lines used is much lower than this. However, since the resolution element is sampled with only 2 channels and the linewidths are generally below ∼2 km s^{-1}, we decided to adopt the resolution as a conservative uncertainty. It is worth noting that in 257 ALMAGAL targets we found that more than one emission feature is present in the spectrum centered toward the C^{18}O (2–1) frequency, indicating that multiple components of gas are present in the same ALMA field of view (FOV). In these cases we selected the densest component choosing the velocity where line emissions were detected also in the transitions with higher critical densities ($>10^5$ cm^{-3}). We also have 12 targets with two components both revealed also in the high density tracers for which the most intense component was adopted. Figure 3a shows the extent of the variations found in the v_{LSR} derived from ALMAGAL data with respect the ones used at the time of the proposal and that were used to define the two subsamples
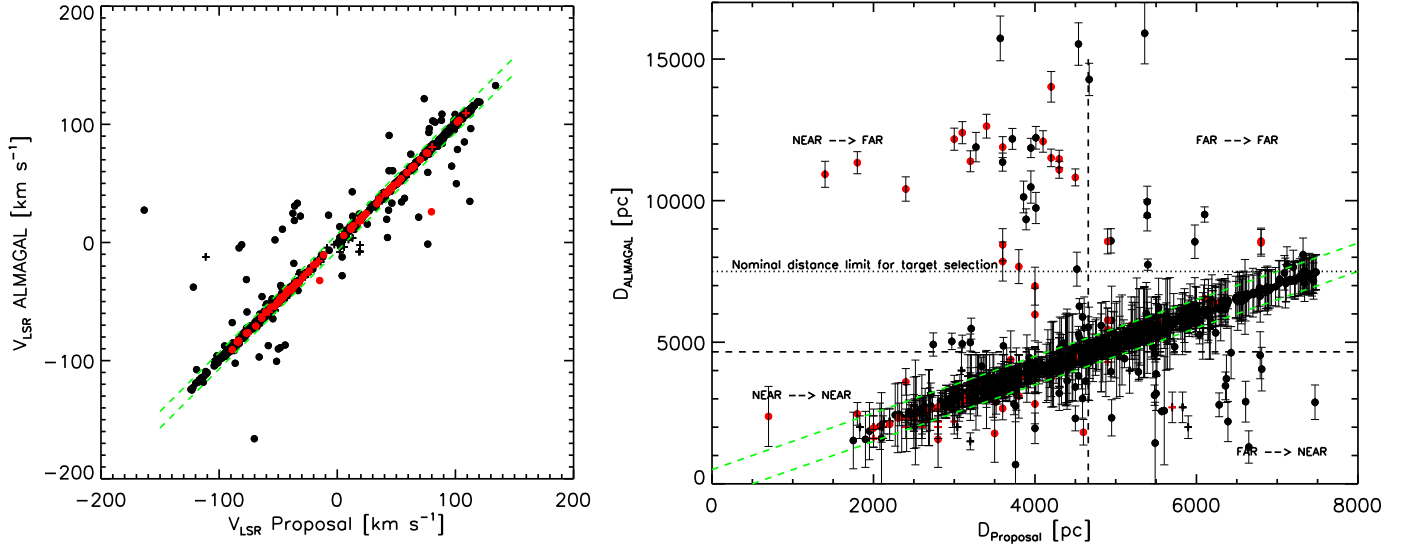
Fig. 3. Physical parameters derived from the ALMAGAL observations. General properties of the ALMAGAL targets. *Left*: plot of the ALMAGAL-determined v_{LSR} vs the values in the original catalogs at the time of the proposal. Black and red symbols identify sources extracted from the Hi-GAL and RMS survey catalogs. The formal uncertainty in the determination (1.4 km s^{-1} see text) is smaller than the size of the points, while green dashed lines mark the extent of a ±7 km s^{-1} v_{LSR} variation typically assigned to local non-circular motions (see text). *Right*: new distances determined from the ALMAGAL-derived v_{LSR} vs the original values (colors as in left panel). Green dashed lines mark a ±500 pc distance variation. The dashed lines mark the distances that were adopted to generate the "near" and "far" ALMAGAL subsamples, with text showing quadrants where sources would change subsample membership (see text for more details).

observed with different 12m-array configurations (see Sect. 2). In 77 sources, the new v_{LSR} values differ from the original ones by more than 7 km s^{-1} in absolute value. This is a threshold that we consider here, along with common literature (e.g., Mege et al. 2021 and references therein), a reliable upper limit for local gas motions above which v_{LSR} variations are likely to reflect different kinematic distances. We therefore chose to adopt this limit as an uncertainty in the determination of the v_{LSR} of a target. Propagating this uncertainty on distance greatly depends on the Galactic longitude of the source, as we discuss below. The differences (sometimes very large) that we found in v_{LSR} with respect to the initial values at target selection are due to an incorrect choice of the CO component that was associated with the sources in the Hi-GAL/RMS samples, among the multiple components found along the sources' line of sight. This resulted in a substantial revision of distances and physical properties for these 77 sources (see below).

The newly determined ALMAGAL-based v_{LSR} values were converted to distances using the final module of the Mege et al. (2021) algorithm, with one important exception. In cases where there was no way to help resolve the KDA for Hi-GAL sources in the inner Solar circle, Mege et al. (2021) adopted (by default) the far distance, while in similar conditions, RMS sources (Urquhart et al. 2018) were put at the near distance. To provide as a homogeneous as possible table of source properties, we chose for this study to adopt the same approach of the RMS catalog and place the Hi-GAL sources for which KDA could not be resolved (7 out of the 915 Hi-GAL sources in the ALMAGAL sample) at the *near* distance. In estimating the error associated with the distance we considered possible deviations from purely circular motions and proper motions as anticipated above, by applying an additional offset to the velocity of 7 km s^{-1} as in Mege et al. (2021); the corresponding offset recovered in the distance to each source will of course be dependent on the Galactic longitude of the object, and is adopted as the distance uncertainty as reported in Table 1, available at the CDS. Error propagation

transfers twice the distance relative uncertainty to source luminosity and masses. No uncertainty originating from distance affects distance-independent quantities such as L/M or clump surface densities.

In Fig. 3b, we compare the newly obtained distances with the original ones at the time of the proposal. We find that for 95 objects out of the total 1013 sources sample, the new distance differs by more than 0.5 kpc from the original one. For 24 objects the change in distance is such that the source would switch from the "near" to the "far" ALMAGAL subsample. Having been observed with the "near" C-5/C-2 ALMA antenna configuration, the target 1000 au linear scale was not reached for these objects. Conversely, for 39 objects, the change in distance is such that the source would switch from the "far" to the "near" ALMAGAL subsample; having been observed with the "far" C-6/C-3 ALMA antenna configuration, the linear scale achieved for these objects was smaller than the target 1000 au. Finally, for 44 objects, the new distance would put them beyond the 7.5 kpc distance limit adopted for the initial source selection; for these sources as well, the target resolved 1000 au linear scale was not reached.

3.2.2. A homogeneous methodology for SED fitting

Of the ALMAGAL targets extracted from the RMS sample, 59 were also present in the Hi-GAL clumps catalogs of Elia et al. (2017, 2021), where the requirement for inclusion was that a source should have a convex SED with at least 4 adjacent *Herschel* bands. For these sources (the "G" sources in Col. 3 of Table 1 with "no sed" flag in the last column) the set of physical parameters reported in Table 1 were adopted from the Hi-GAL catalogs. For the rest of the RMS sources in the sample the above criteria were failed and hence they required a custom analysis of their SED to estimate parameters with the same approach as in Hi-GAL (described in detail in the above cited Elia's papers). The different situations encountered can be grouped as follows, identified by a flag that is also reported in Table 1:

- **-sed_irr**: the RMS source had an entry with at least four adjacent bands in the Hi-GAL band-merged photometric catalog, but the *Herschel* SED was "irregular" in the sense that it showed changes in concavity directions. For the present work, the processing to derive the mass, L_{bol}, T_{dust}, and T_{bol}, was also applied to these sources.

- **-sed_rebuilt**: the RMS source had multiple "<4-bands" entries in the Hi-GAL band-merged photometric catalog. A typical occurrence would be a RMS source with two counterparts at 250 μm, where one of them would be linked to a 70–160 μm SED branch, and the other to another 350–500 μm SED branch. Both branches would feature less than four adjacent bands and would not pass the criteria to be included in the Hi-GAL physical parameters catalog. For these cases, in the present work the partial branches were reconnected (e.g., in the example above, forcing a single 250μm source to be extracted).

- **-sed_sat**: the RMS source is associated with saturated pixels in *Herschel* 250 or 350 μm bands. As such, the fluxes for the source could not be extracted in four adjacent bands. For the present work we forced the Hi-GAL processing allowing for "holes" in the SED.

- **-sed_noband**: the RMS source has a 3-band 70–160–250 μm counterpart in the Hi-GAL band-merged catalog, vanishing with at 350 and 500 μm. For the present work we forced the Hi-GAL processing with 3-bands *Herschel* SEDs.

- **-sed_noprops**: the RMS source has a counterpart only at 70 μm and, hence, no Hi-GAL-like processing can be applied to derive physical parameters.

Finally, in addition to the evolutionary information provided by the L/M ratio, we also report the classification according to the method from Urquhart et al. (2022) who define classes as quiescent, protostellar, YSO, and HII regions based on visual inspection at different wavelengths. Each region is visually inspected at 70, 24 and 8 μm. Roughly 65% of the ALMA-GAL sources were cross-matched with the ATLASGAL catalog and the classification adopted from there. The remaining 35% of sources were classified by hand using the same methodology.

Table 1, available at the CDS, reports the consolidated properties of the ALMAGAL targets as revised above. Uncertainties are explicitly given in the table for the distance to each source. Uncertainties on masses and luminosities are easily derived from those and are not explicitly reported. Columns explanation is available in Appendix A.

4. Properties of continuum emission

The ALMAGAL target sample spans large ranges in several physical and environmental parameters (Sect. 3) and this is reflected in the wide variety of properties and morphologies of the 1.38 mm continuum emission. While a complete gallery of 1.38 mm continuum images is available on the project's website[2], in Fig. 4, we illustrate 12 continuum emission maps that illustrate the whole spectrum of morphologies in the ALMAGAL sources. Fields AG013.4582-0.0377 (#46) and AG284.0055-0.8459 (#400) shows relatively extended emission that is mostly resolved at the 1000 au scale of the full-resolution images. We can then see fields with variable level of fragmentation, from single main cores as in AG013.7869-0.2359 (#48), to small clusters more (AG285.2633-0.0501, #417), or less compact as in AG024.5252-0.1390 (#186) and AG311.4675+0.3724

(#592), to larger clusters embedded in sometimes considerable extended emission with either distinctively filamentary, as in AG028.3456+0.0605 (#263), AG320.2463-0.2942 (#647), or more irregular (the remaining panels) shapes.

In this section, we provide a characterization of the continuum emission properties in our ALMA images in relationship to the integrated properties of the target clumps. The analysis is based on the 1.38 mm continuum emission detected over 5σ level over the root mean square (rms). This noise level is computed in each field using the residual map outside of the largest CLEAN mask used in the ALMAGAL pipeline processing (Sanchez-Monge et al. 2025) and it is stored in the keyword AGSTDREM of the maps FITS Header. The 5σ threshold was adopted to minimize the inclusion of noise residuals in the morphological analysis. For each of the 1013 independent fields we produce masks containing the pixels of the continuum images with flux above the 5σ level, which we call RoI (Region of Interest). Each cluster of connected pixels constitutes an individual RoI, so that each field may have many RoIs. For each RoI, we compute the area and the perimeter, and all RoIs smaller than the beam are discarded from further analysis. For each valid RoI, we computed the fluxes (peak, total, and median fluxes) by projecting the RoI mask onto the primary beam-corrected images (PBCOR). Figure 5 reports the distribution of the number of RoIs, and their total area, per imaged clump where emission was detected. It is dominated by fields with relatively low number of RoIs, showing that the threshold level is adequate to recover areas with reliable emission. The distribution of areas is instead relatively flat, implying that large areas of emission do not necessarily result from the total contribution of many RoIs in the field.

In addition, for each field, we also computed the perimeter and area of the convex hull for each of the 5σ emission contours, which is the smallest convex figure containing the contour (see Fig. 14 for an example). Any of these geometrical properties of the dust emission in the observed fields can then be related to other integrated field properties (e.g., clump mass, surface density, L/M as estimated from *Herschel* measurements), or can be compared among different fields in the sample.

4.1. ALMA 1.38 mm emission vs. Herschel clump properties

It is interesting to compare the ALMA 1.38 mm emission properties with global clump parameters mostly based on *Herschel* observations and listed in Table 1, particularly as parameters such as the far-IR (FIR) flux and surface density have been used as the selection criteria (Sect. 3).

One initial aspect we want to quantify is how the thermal dust emission at different spatial scales is conserved. Interferometers introduce a spatial filtering of the emission that has a spatial scale larger than the one sampled by the shorter antenna baselines. The selection of the ALMAGAL targets sample has been based on properties directly (flux) or indirectly (mass, luminosity, surface density) estimated from the *Herschel* thermal far infrared emission. The comparison of these properties with similar properties derived from the millimeter thermal dust emission from ALMA can provide interesting insight concerning the spatial distribution of the cold dust within the target clumps. For this comparison, we chose to use the 7M+TM2 continuum images with primary-beam-corrected fluxes (PBCOR).

Figure 6 shows the relationship between the *Herschel* 350 μm flux of the ALMAGAL target fields and the 350 μm-rescaled total flux obtained from the total 1.38 mm continuum
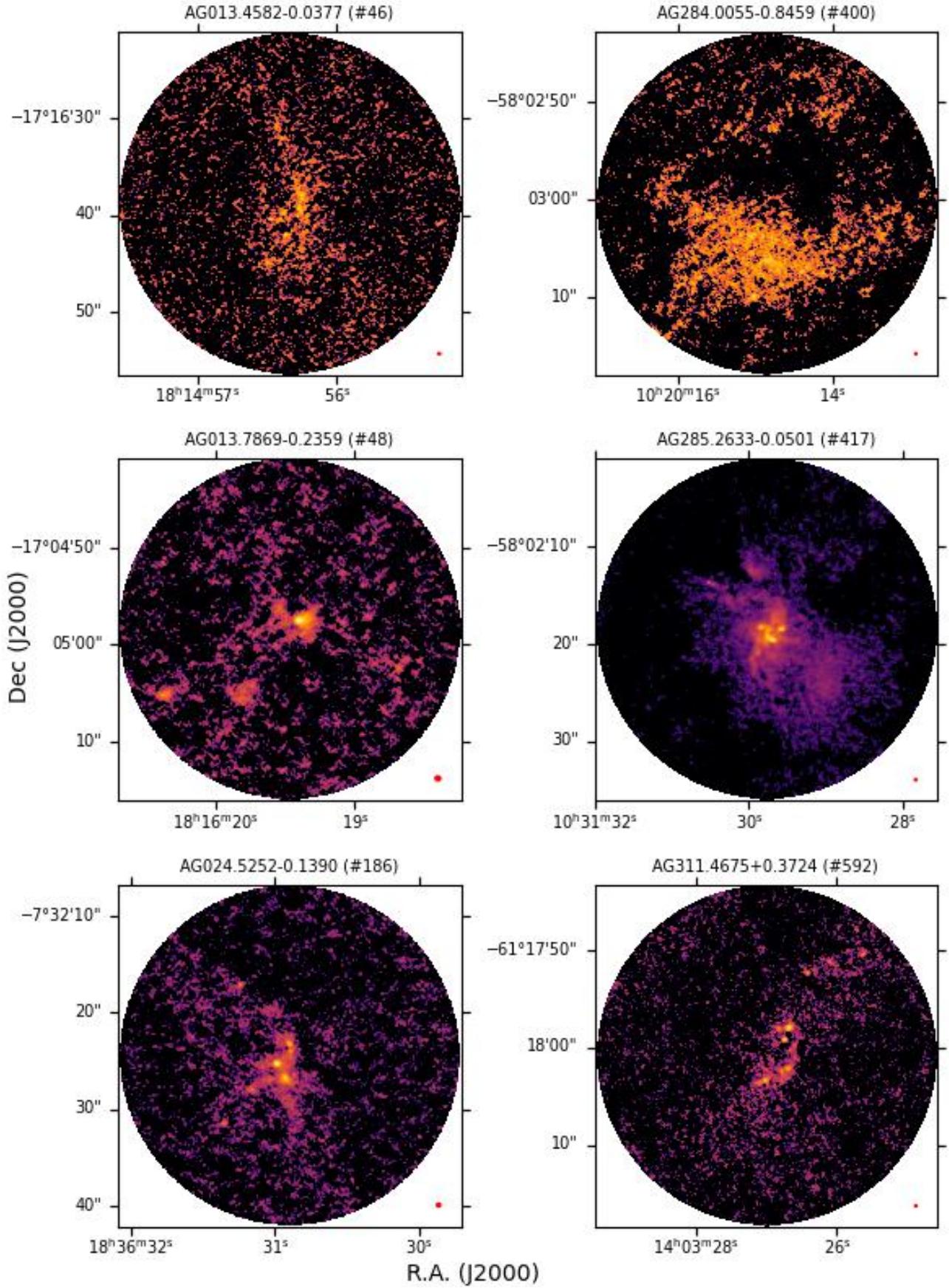
---

[2] www.almagal.org

Fig. 4. ALMA 1.38 mm continuum 7M+TM2+TM1 images of a selection of ALMAGAL fields showing the large variety of fragmentation levels and extended emission found. The target AG name and running number are from Cols. 2 and 1 of Table 1. The small red ellipse in the bottom-right corner represents the synthesized beam.
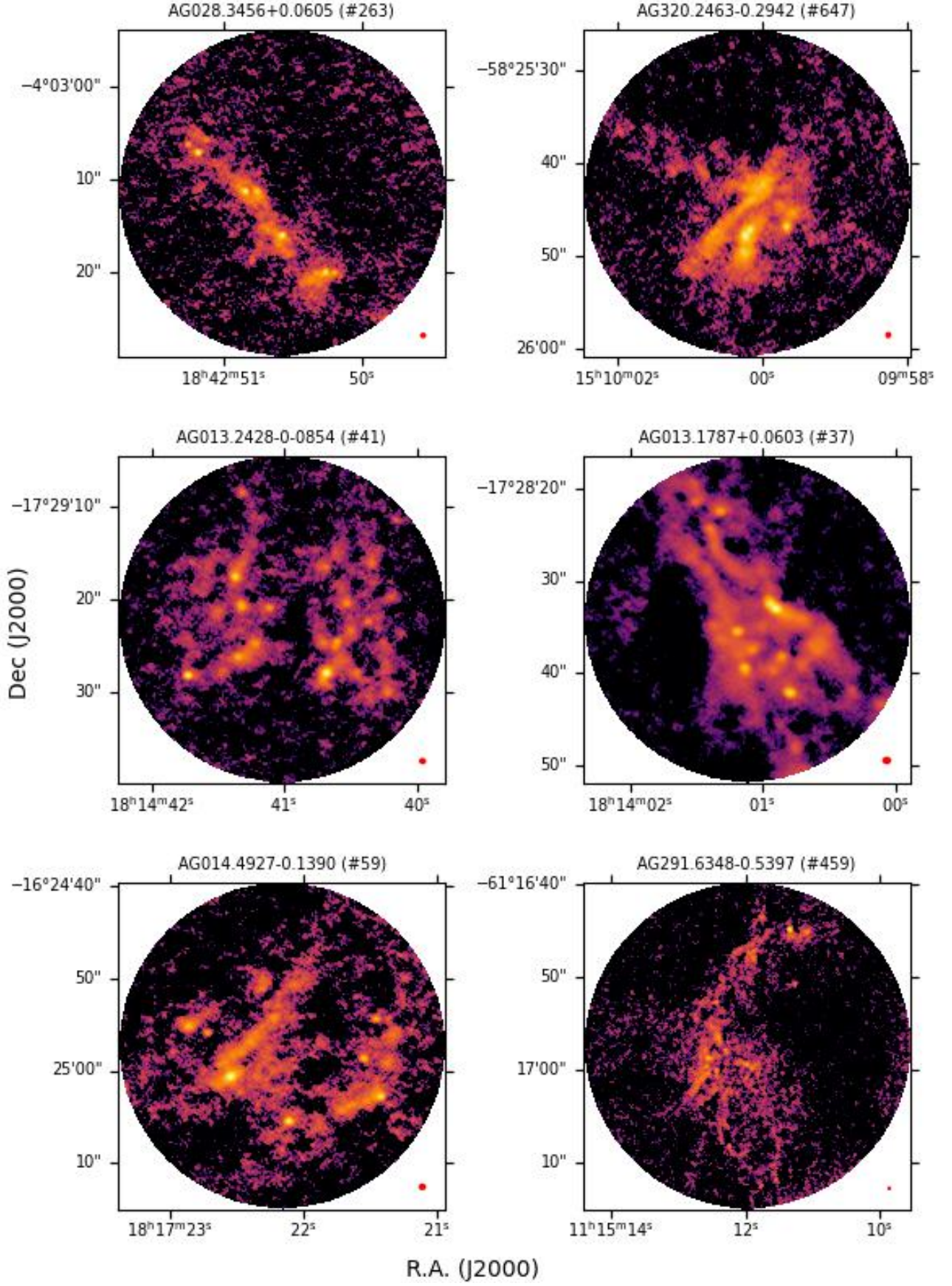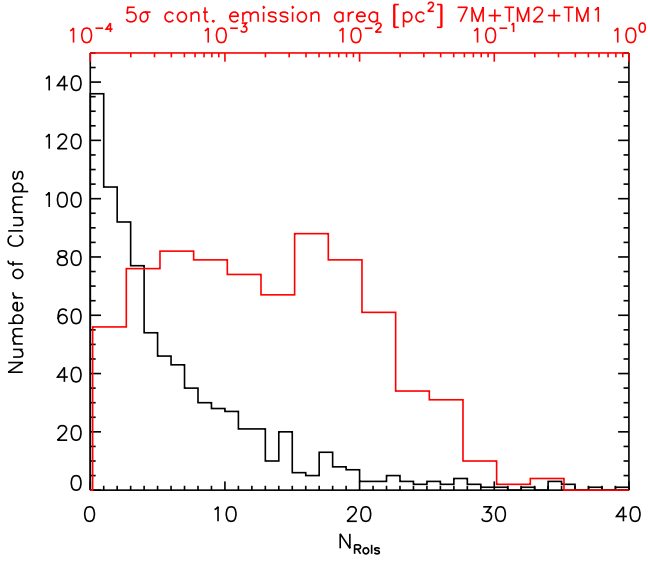
AG028.3456+0.0605 (#263)

AG320.2463-0.2942 (#647)

AG013.2428-0-0854 (#41)

AG013.1787+0.0603 (#37)

AG014.4927-0.1390 (#59)

AG291.6348-0.5397 (#459)

Dec (J2000)

R.A. (J2000)

Fig. 4. continued.

Fig. 5. Distribution of the number of RoIs (in black, bottom X axis) and of their area in pc^2 (in red, top X axis) revealed for the 1.38 mm continuum at 5σ per clump.
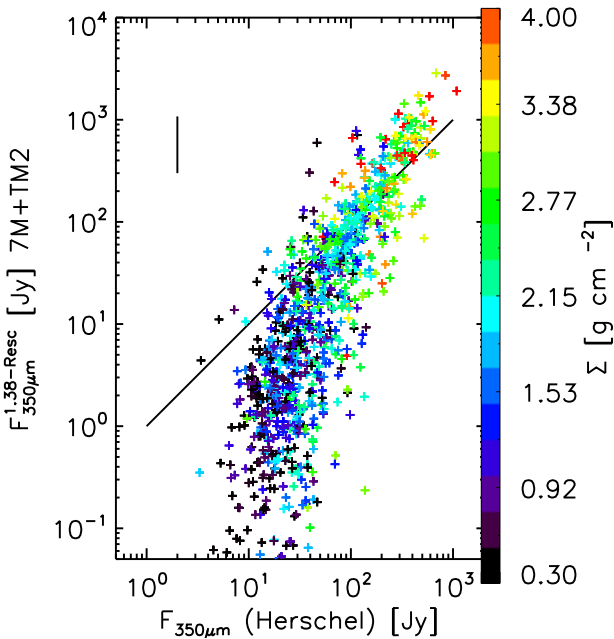
Fig. 6. Total flux within the 4σ contours of the 7M+TM2 PBCOR images rescaled to 350 μm, as a function of the 350 μm flux from *Herschel*. For the rescaling of the ALMA 1.38 mm flux to 350 μm we use optically thin assumptions with $\beta = 1.75$ and the dust temperature from the *Herschel* SED fitting. The black line is the identity line, and the color scale is proportional to the clump surface Σ, again from *Herschel*. The vertical black line shows the spread of values that the 350 μm-rescaled 1.38 mm flux can assume for different dust temperatures and opacities.

in the 4σ contour of the PBCOR images in the 7M+TM2 configuration, provided that the area of the contour is larger than the synthesized beam. In particular, we extrapolated the mm flux to 350μm using a modified black-body assuming different dust opacities and temperatures (see below). A 4σ was specifically adopted here (instead of the 5σ use in the rest of the analysis) to make sure we could recover as much as possible the emission area for a most reliable assessment of a missing flux problem.
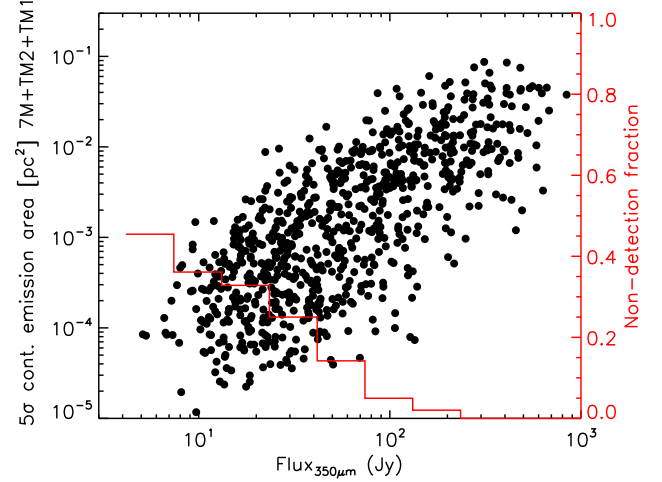
Fig. 7. ALMA 1.38-mm total emission area above 5σ in the 7M+TM2+TM1 images as a function of the Hi-GAL 350 μm flux (r_P=0.74). The red histogram reports the fractional distribution of the 350 μm flux for ALMAGAL sources with no 5σ detection (to be read on the red right y-axis).

Signal is detected toward 941 out of 1013 fields ($\sim$93%), and in 336 out of 941 fields ($\sim$36%), the discrepancy between the fluxes is less than 50%. For a fraction of the fields the rescaled 350 μm flux is above the *Herschel* measured flux, but that is compatible with the uncertainties due to the assumptions in terms of dust temperature and opacities that we are forced to make to rescale the 1.38 mm flux to 350 μm. Indeed, the vertical bar in the figure represents the extent of the overall spread of 350 μm-rescaled fluxes obtained varying the temperature between 20 K and 50 K, dust spectral index β between 1.5 and 2 (a range typical for diffuse ISM and dense clumps), and assuming opacities from Preibisch et al. (1993), Ossenkopf & Henning (1994) and Draine (2003).

The figure shows that the distribution is strongly skewed toward low 7M+TM2-rescaled fluxes. For about 64% of the fields the rescaled 7M+TM2 recovered flux is much less than 50% of the *Herschel* flux, and for these fields this is a clear indication of missing emission. The extent of the emission in the 350 μm *Herschel* images is generally larger than the ALMA FOV, so that the discrepancy could result simply from the different sizes of the areas sampled. However, the beam of *Herschel* at 350 μm is $\sim$80% of the ALMA FOV and therefore the far-IR emission could plausibly come from smaller areas within the ALMA FOV. In particular, for 90% of the targets the measured clump FWHM from *Herschel* is smaller than the ALMA FOV; the percentage rises to 95% if the beam-deconvolved *Herschel* size is adopted. We conclude that it was only in 10% of the targets that the discrepancy in recovered flux could be attributed to differences in the areas of the sky sampled by *Herschel* and ALMA.

Instead, the color scale of the points shows that these fields have the lowest surface density, as estimated from *Herschel*; hence, the dust distribution is expected to be less compact and more prone to spatial filtering by the interferometer. This indicates that for a significant fraction of targets we may be missing flux from larger angular scales, and the use of physical parameters such as the total clump mass that could be derived from integrated ALMA millimeter fluxes is not reliable for the majority of the fields.

We wanted to concentrate on the morphology of continuum emission for this work. Figure 7 shows the relationship between
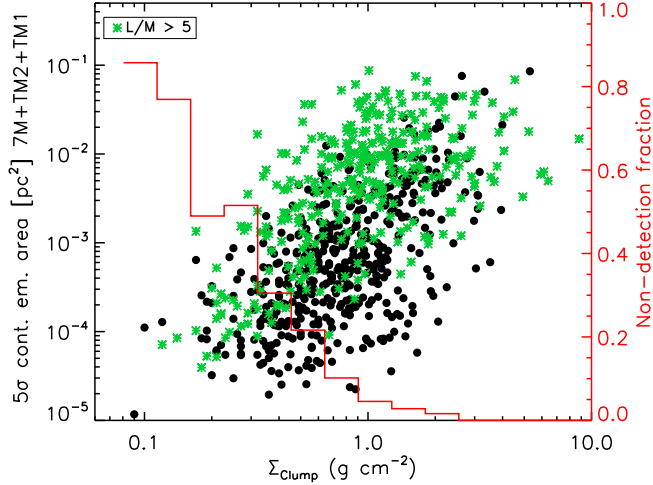
Fig. 8. ALMA 1.38 mm total emission area above 5σ in 7M+TM2+TM1 images as a function of the clump surface density from Table 1. The green symbols are for target clumps with $L/M \geq 5$. The red histogram reports the fractional distribution of the surface density for ALMAGAL sources with no 5σ detection (to be read on the red right y-axis).

Fig. 9. ALMA 1.38 mm total emission area above 5σ in 7M+TM2+TM1 images as a function of the clump L/M from Table 1. The green symbols are for target clumps with $\Sigma_{\mathrm{Clump}} \geq 0.8$ g cm^{-2}. The red histogram reports the fractional distribution of the surface density for ALMAGAL sources with no 5σ detection (to be read on the red right y-axis).

the 1.38 mm total emission area above 5σ level in the ALMA 7M+TM2+TM1 image as a function of the *Herschel* 350 µm flux. The plot outlines a very good correlation among the two parameters that is not influenced by the different distances of the targets as neither of the two parameters depends on it (Figs. B.1, B.2). The Pearson linear correlation coefficient is r_P=0.74.

The fraction of fields with no detected emission above 5σ in ALMA is concentrated toward the low end of the 350 µm flux distribution, but there is no indication of an exact far-IR flux threshold for the 1.38mm detection. The figure confirms that the 350 µm flux cut adopted for the selection of the ALMAGAL target clumps is well matched to the chosen ALMAGAL target sensitivity. A fraction of targets below 10% is not detected with ALMA also for relatively large 350 µm fluxes. The average properties of non-detected fields are discussed in Elia et al. (in prep.).

Figure 8 shows the relationship between the ALMA 1.38 mm 5σ area and the clump surface density estimated from Hi-GAL data. As in the previous figure, the red histogram shows the Σ_{Clump} distribution for fields without 5σ detection. Here again a correlation is clearly apparent, with r_P=0.57 suggesting indeed a good correlation, although with a large scatter in the y-axis distribution.

The problem of possible missing flux may play a role in determining the scatter of the points; however this should affect only the relatively lower surface density clumps (see Fig. 6), while the scatter we see in the Y direction in Fig. 8 is independent from the range of surface density. The points reported in green color identify the clumps with $L/M \geq 5$, showing that they predominantly populate the upper range of the points distribution. Therefore, it is the clump evolutionary stage that introduces most of scatter.

The relation between the Hi-GAL Σ_{clump} and the ALMA-GAL 1.38-mm 5σ area is interesting for two key reasons. First, it again reinforces the adopted choice of parameters to drive the ALMAGAL source selection. The 1.38 mm emitting area, under the optically thin assumption, is a proxy for the dust column density spatially resolved in the ALMAGAL maps; the surface density estimated from *Herschel* observations is also a measurement of column density that is, however, an integrated
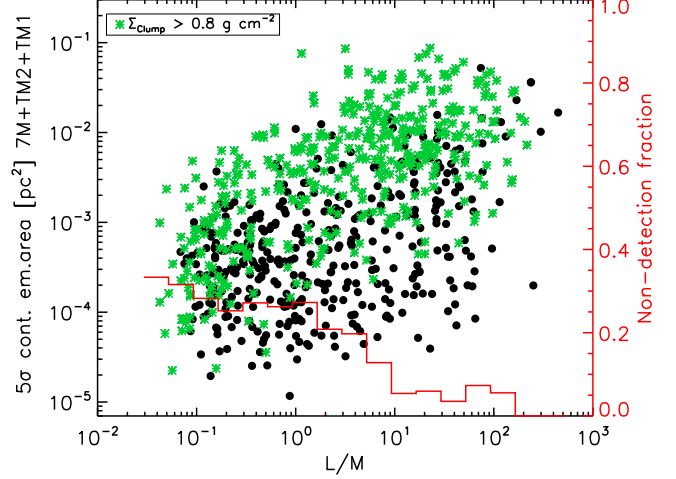
quantity averaged over the extent of the clump. It is then reassuring to verify that in spite of the missing flux problem and taking into account evolutionary effects, the two parameters are indeed correlated. In addition, the fractional distribution of Σ_{Clump} for non-detected fields in ALMA is heavily skewed toward the lower end of the range and this suggests that the choice of the target sensitivity adopted for the ALMAGAL observations is very well matched to the average properties of selected targets.

Second, since the millimeter continuum in interferometric observations traces denser and compact rather than diffuse and extended ISM (as suggested in Fig. 8), the degree of mass concentration as traced by the clump-averaged Σ_{Clump} has a role in determining the amount of dense material that is found at smaller spatial scales probed by ALMA. Furthermore the correlation present in Fig. 8 completely disappears if we use the clump mass instead of its surface density (see Fig. B.6), confirming that is not simply the available mass that drives the process by which the clump ISM is shaped at smaller and smaller scales, but the shape of the gravitational potential. This was anticipated by, for instancem, McKee & Tan (2002) and will be explored in more detail in Elia et al. (in prep.). To illustrate how we excluded the possible presence of a bias introduced by the different distances of the targets, in Fig. B.3 we report the same two quantities of Fig. 8, but in separate distance bins, showing that the relationship holds irrespectively of the distance to the sources.

To illustrate more clearly the degree to which the extent of dense ISM in the clumps fields depends on their evolutionary stage, we report in Fig. 9 the 1.38 mm ALMA 5σ emission area as a function of the clumps, L/M. The two parameters appear indeed correlated, and the large scatter in the y-axis is this time due to the spread in surface density. Indeed in the figure we outline with green symbols the clumps with higher surface density, finding that the larger 1.38 mm emission areas are indeed found on average toward clumps with higher surface density. The correlation coefficient for the entire dataset in Fig. 9 is r_P=0.53, confirming a relatively good correlation. However, this is mostly driven by the higher Σ_{Clump} points for which we have r_P=0.66 and for which the correlation is therefore stronger. The correlation for the subset with lower Σ_{Clump} is only moderate, with correlation coefficient dropping to r_P=0.42.
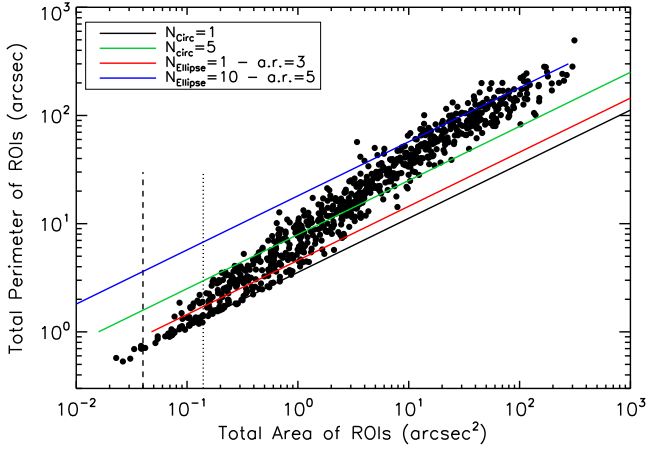
Fig. 10. Perimeter-area relationship where the two parameters have been summed over all the 5σ-RoIs identified in each ALMAGAL field. The black and green lines represent Eq. (1) in the case of a single circular shape or where the assumed perimeter is distributed for five circular shapes. The red and blue lines indicate the cases for a single ellipse with aspect ratio a.r. = 3 and 10 elliptical shapes with a.r. = 5. The vertical dashed(dotted) lines show the median area of the 7M+TM2+TM1 beam for the far/(near) sources (from Sanchez-Monge et al. 2025).

Since the clump surface density does not appear to be an evolutionary indicator based on the L/M (Elia et al. 2017), we conclude from this preliminary global analysis of the continuum emission in ALMAGAL fields that the presence of dense ISM substructures in the target clumps depends both on the initial conditions for clumps fragmentation (as traced by the clump-averaged surface density) and the global evolutionary stage of the clumps as traced by the L/M parameter.

4.2. The morphology of 1.38 mm emission in ALMAGAL

Given this large variance in the morphological appearance of the 1.38 mm continuum emission (Fig. 4), we would like to verify whether there is a metric that is able to clearly categorize the morphology into different classes that could be then related to physical parameters. In the following, we explore the relationship between the area of the emission with its perimeter and with the area of its convex hull.

4.2.1. Area vs. perimeter

A first simple approach in the characterization of the morphology of the 1.38 mm continuum's spatial distribution, is to analyse the relationship between the area A and the perimeter P of the RoIs identified by thresholding images above the 5σ level. In generalized terms, the two quantities are related as:

$$P = kA^{d/2}, \tag{1}$$

with k depending on the specific geometric figure describing the RoI delimiting shape; for instance, in 2D we have $k=4$ for a square or $k = 2\sqrt{\pi}$ for a circle.

The exponent d is the dimension of the shape bounding the RoI; it is $d = 1$ for linear contours in 2D space and $d = 2$ for surfaces enclosing volumes in 3D. As we depart from regular Euclidean planar figures, replicating so-called generator patterns in a self-similar way at smaller and smaller scales to describe increasingly irregular contours, d can assume non-integer values between 1 and 2 and in this we recognize the fractal dimension.

Figure 10 illustrates the perimeter-area plot for the 5σ RoIs identified in each ALMAGAL field, where the two parameters
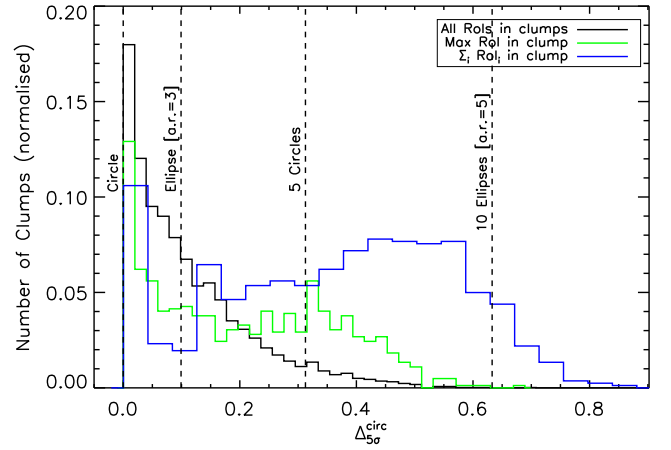
Fig. 11. Normalized histograms of the $\Delta_{5\sigma}^{\mathrm{circ}}$ parameter for the largest RoI in each field (black line) and for all 5σ RoIs in the fields. The $\Delta_{5\sigma}^{\mathrm{circ}}$ values corresponding to the loci indicated by the colored lines in Fig. 10 are also reported here for reference.

have been summed over all RoI. To understand the meaning of the spread of the points in this plot, it is useful to mark locations that would be compatible with simple structures or their combination. For example, the black and red lines corresponds to Eq. (1) in the case of one circle ($k = 2\sqrt{\pi}$) or of one ellipse[3] with aspect ratio a.r. = 3. We see that only in a small number of cases the emission area is compatible with a single circular or elliptical shape and only for small areas of a few arcsec^2, corresponding to about four to five times the maximum beam area represented by the vertical dashed line (from Sanchez-Monge et al. 2025). As another example, if we assume that the total emission area is divided in, say, five equal circular RoIs then the correspondent locus in the plot is given by the green line; similarly, the blue line shows the locus in case the emission area was composed of ten equal elliptical RoIs with a.r. = 5.

The distribution of points shows a continuum of situations. The limited vertical scatter for relatively small emission areas (left of the vertical dashed line) is a construction bias, as for these small areas, there is low maximum number of RoIs it can be divided into and still be at least the size of the beam. The vertical scatter becomes much wider as we increase the total RoIs area. All clumps below and about the green line would be compatible with small clusters of compact objects even for relatively large total emission areas (up to $\sim$100 sq. arcsec.). Clumps with emission showing a larger departure from the circularity could be compatible with even larger cluster of compact roughly circular objects, or a more limited number of filamentary patches idealized as the combination of ten high-a.r. ellipses and represented by the blue line. Of course, large deviations from a single circular RoI may also be compatible with a few patches of very irregular emission; we present an additional characterization in this respect below.

To quantify this behavior, we defined a parameter $\Delta_{5\sigma}^{\mathrm{circ}}$ computed as the distance between each point and the black line (along the direction perpendicular to the line) in Fig. 10.

In Fig. 11, we report the distribution of the $\Delta_{5\sigma}^{\mathrm{circ}}$ parameters in our fields, computed individually for all the RoIs in the fields (black line), for the largest RoI in each field (green line), and for

---

[3] For the ellipse we adopted an approximation for k obtained from the approximation formula $P_{\mathrm{ellipse}} \sim \pi\sqrt{2(a^2 + b^2)}$.
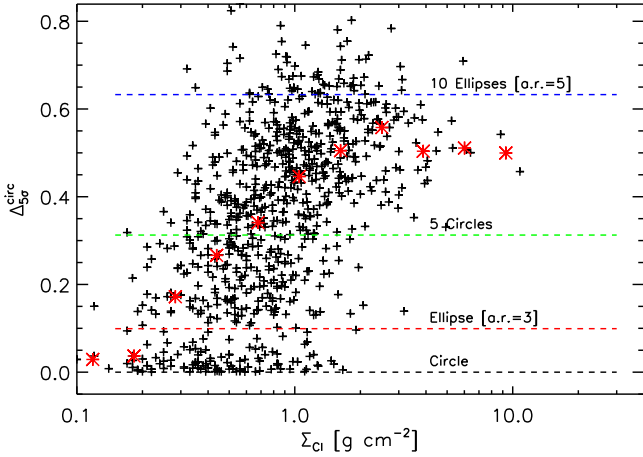
Fig. 12. Departure from circularity, $\Delta_{5\sigma}^{\mathrm{circ}}$, of the total of the 5σ ALMA emission RoIs in each field as a function of the clump surface density. The red asterisks are the medians of $\Delta_{5\sigma}^{\mathrm{circ}}$ in logarithmic bins of surface density.

Fig. 13. Departure from circularity, $\Delta_{5\sigma}^{\mathrm{circ}}$, of the total of the 5σ ALMA emission RoIs in each field as a function of the clump, L/M. The color scale represents the clump surface density as reported in the lateral bar, while the black asterisks are the medians of $\Delta_{5\sigma}^{\mathrm{circ}}$ in logarithmic bins of L/M for the clumps with $\Sigma_{\mathrm{Clump}} \geq 1$ g cm^{-2}.

the sum of the RoIs in each field (blue line). Histograms are normalized to their integral to make their comparison meaningful. The distribution for all RoIs (black) shows a clear predominance of circular and slightly elongated shapes, with a monotonically decreasing trend toward larger values of $\Delta_{5\sigma}^{\mathrm{circ}}$. This might suggest that in many cases the secondary RoIs at the 5σ level are dominated by small circular areas that are either fainter patches of extended emission or compact objects isolated with respect to the main emission area in the field.

When considering only the largest RoI in each field (green), the distribution is clearly different and shows a first component peaked around 0 (circular shape) decreasing down to about $\Delta_{5\sigma}^{\mathrm{circ}} \sim 0.2$, followed by a second smaller and broader peak down to $\Delta_{5\sigma}^{\mathrm{circ}} \sim 0.5$. If, instead, we consider all the RoIs in each field and we compute the total perimeter and total area, the corresponding $\Delta_{5\sigma}^{\mathrm{circ}}$ distribution is the blue line; in addition to this blue line, a narrow component skewed toward 0 (i.e., circular structure) shows an even broader (compared to the green distribution) component extending to ~ 0.8. This metric seems therefore able to select fields that are showing either complex patterns of extended emission or collections of roughly circular or elongated structures (or, mostly likely, both). Values of $\Delta_{5\sigma}^{\mathrm{circ}}$ for all clumps are reported in Table 2.

It is interesting to verify whether these metrics are related to global clump properties. The relationship of the $\Delta_{5\sigma}^{\mathrm{circ}}$ parameter with the clump-averaged surface density in Fig. 12 shows a positive correlation, although as usual with a large scatter around the median values computed in bins of Σ_{Clump} (the red asterisks). While clumps with $\Sigma_{\mathrm{Clump}} \leq 1$ g cm^{-2} span the entire range of $\Delta_{5\sigma}^{\mathrm{circ}}$, almost the totality of denser clumps have $\Delta_{5\sigma}^{\mathrm{circ}}$ higher than values compatible with circular or slightly elliptical shapes. Similarly to Fig. 8, we exclude the presence of remarkable distance biases in Fig. 12 as similar trends are found if the two parameters are plotted for sources in 1 kpc-bin distances (Fig. B.5).

A similar situation is found when relating $\Delta_{5\sigma}^{\mathrm{circ}}$ with L/M. Figure 13 shows that the scatter in the Y axis grows with evolution, with median-averaged values clearly rising with L/M, depicting the more and more frequent presence of complex morphologies with evolution. The points in the figure are color-coded with the clump surface density, and confirm (see also Fig. 12) that this parameter has a strong role in modulating the
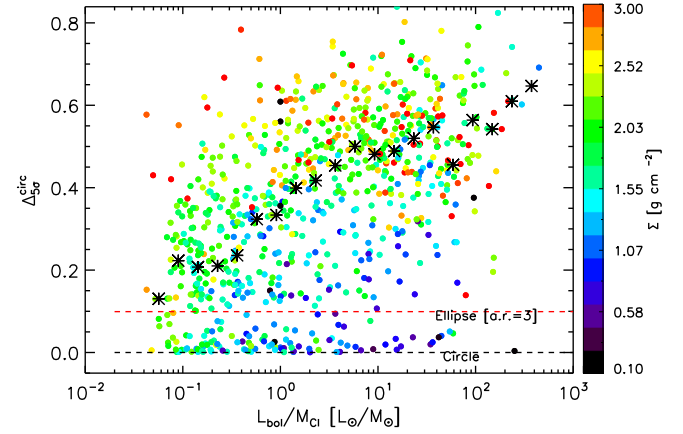
scatter in the Y axis of Fig. 13. For example, if we restrict to the clumps with $\Sigma_{\mathrm{Clump}} \geq 1$ g cm^{-2} (the cyan to red points), we are left with a moderate but significant $\Delta_{5\sigma}^{\mathrm{circ}}$-$L/M$ trend with $r_P \sim 0.46$.

4.2.2. Area versus convex hull

Another interesting metric to characterize the complexity of the emission morphology is offered by the comparison of the areas of the RoIs and the respective convex hull areas. For a bounded subset of the plane, the convex hull may be visualized as the shape enclosed by a rubber band stretched around the subset (Berg et al. 2008). As it essentially represents the smallest convex geometric figure containing the emission RoIs, we expect the two areas to be very similar in case of very regular (roughly circular or elliptical) shapes. On the contrary,. the two areas will be different in case where the RoI is the result of the merging of two or more distinct regions, for instance, or in the case of hub-like filamentary structure. In the following, we define the parameter $Q_{5\sigma}^{\mathrm{hull}}$ as the ratio A_{RoI}/A_{C-Hull} between the area of the largest RoI and its convex hull area computed over the areas with signal above the 5σ level.

Figure 14 shows an example of two extreme cases in which the emission area above 5σ has a similar extent (inside the blue contour), but $Q_{5\sigma}^{\mathrm{hull}}$ is relatively high (~ 0.85) for the relatively regular source (top), and drops to 0.5 for the more complex structure field (bottom). Figure 15 reports the histogram of this parameter over the entire sample where 5σ continuum emission is detected, and suggests that there may possibly be two peaks in the distribution, separated at a value of ~ 0.8. We ran the Silverman test (Silverman 1981) to characterize multi-modality using kernel density estimates over the distribution of the original data, and hence not on binned histograms. The test confirms that the distribution reported in Fig. 15 has a 100% probability of being better described by two modes instead of one. $Q_{5\sigma}^{\mathrm{hull}}$ then offers a simple metric to classify fields based on the convex (or not) nature of their continuum emission.

The relationship of this parameter with the clump's evolutionary stage in Fig. 16 shows a broadly decreasing trend with L/M, indicating that the emission morphology of the largest emission area in each field gets more and more complex as the
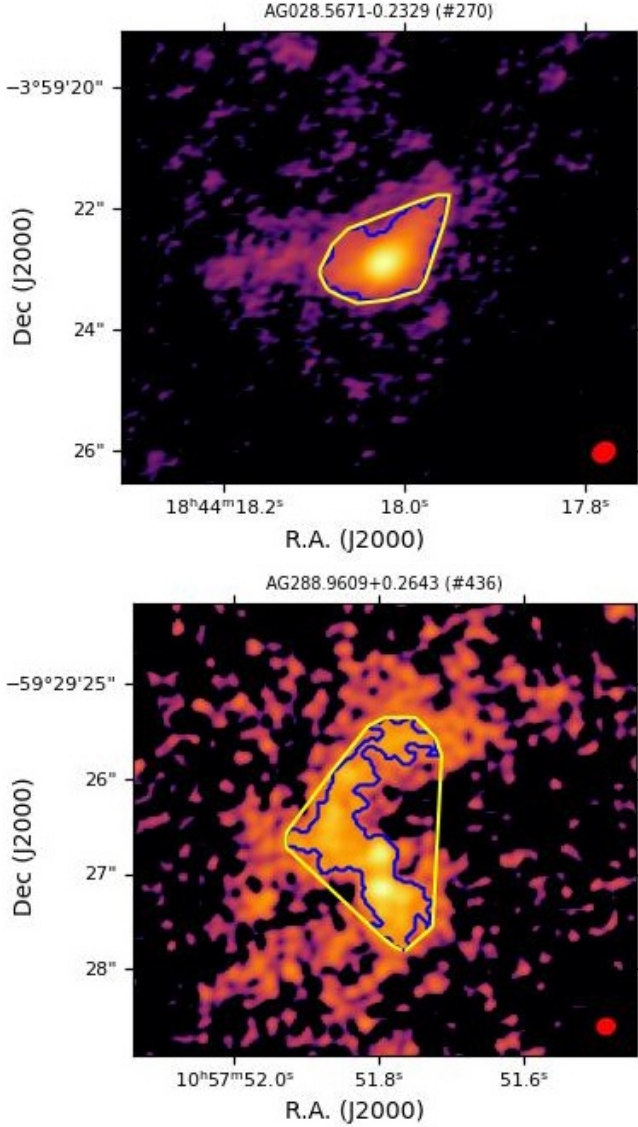
AG028.5671-0.2329 (#270)

AG288.9609+0.2643 (#436)

Fig. 14. Two ALMAGAL fields showing different morphologies for similar 5σ emission areas. The blue contours in each field correspond to the 5σ noise level, while the yellow polygons represent the computed convex hull for the largest RoI in the fields. We have $Q_{5\sigma}^{\mathrm{hull}}$ is ~0.85 for the field in the top map, and ~0.5 for the field in the bottom map. In both plots, the red ellipse in the bottom-right corner is the beam. Source AG name and running number as in Table 1.

field is more evolved. This is also clearly shown in Fig. 17, where the distribution of L/M is reported separately for the fields with "regular" ($Q_{5\sigma}^{\mathrm{hull}} \geq 0.8$, black line) and "irregular" ($Q_{5\sigma}^{\mathrm{hull}} \leq 0.8$, red line) continuum emission morphologies. The L/M for the two fields in Fig. 14 is ~0.05 for AG028.5671-0.2329 and ~26 for AG288.9609+0.2643, respectively.

There is, of course, a bias introduced by the finite resolution of the images. Fields with only compact beam-like emission patches will naturally be roughly circular in morphology, making this metric not informative. These are, however, only a portion of the fields with high $Q_{5\sigma}^{\mathrm{hull}}$. Figure 18 illustrates the relationship between $Q_{5\sigma}^{\mathrm{hull}}$ and the 5σ emission area. The horizontal dashed line marks the $Q_{5\sigma}^{\mathrm{hull}} \sim 0.8$ threshold identified in Fig. 15 to distinguish approximately circular areas of emission from more complex morphologies, while the vertical dashed line indicates
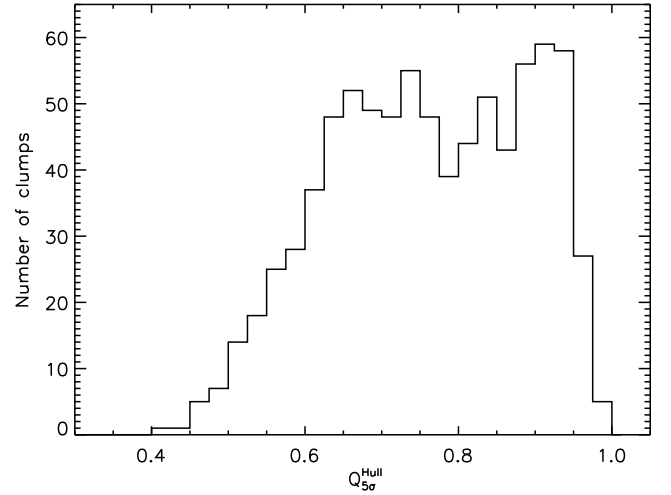
Fig. 15. Distribution of the $Q_{5\sigma}^{\mathrm{hull}}$ parameter for all ALMAGAL fields where 5σ continuum emission is detected.
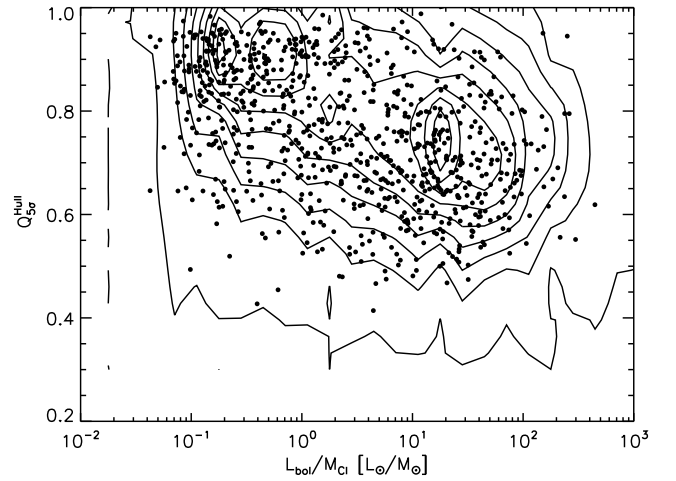
Fig. 16. Distribution of $Q_{5\sigma}^{\mathrm{hull}}$ for all ALMAGAL fields with respect to the evolutionary stage of the clump as traced by the L/M. Contours represent the normalized source density in steps of 10%.
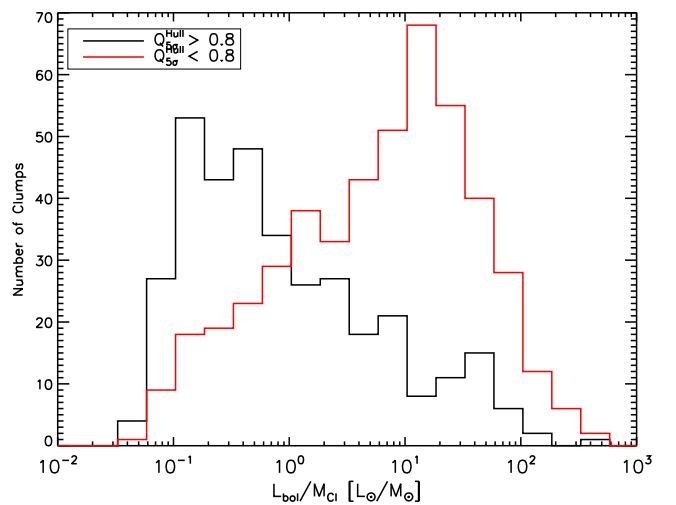
Fig. 17. Distribution of the L/M parameters for fields with $Q_{5\sigma}^{\mathrm{hull}}$ parameter above (black line) or below (red line) the value of 0.8.
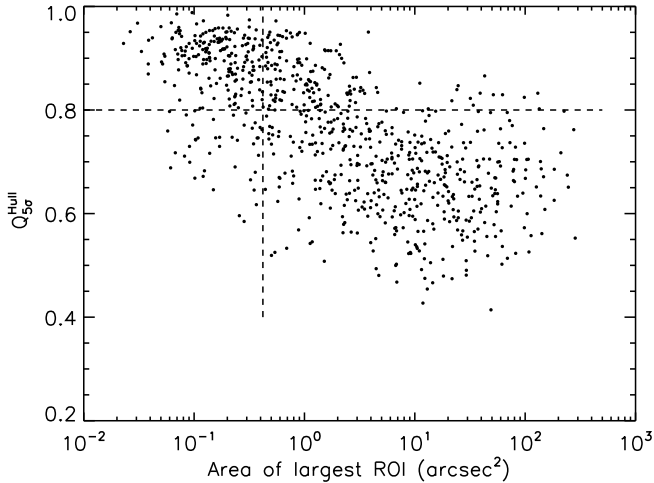
Fig. 18. Distribution of the $Q_{5\sigma}^{\text{hull}}$ parameter as a function of the area of emission above 5σ. The horizontal dashed line marks the $Q_{5\sigma}^{\text{hull}} \sim 0.8$ threshold identified in Fig. 15 to distinguish roughly circular areas of emission from more complex morphologies. The vertical dashed line indicates the maximum area of the beam for all fields, showing that a significant fraction of approximately circular emission areas (above the horizontal line) have areas much larger than the beam.

the maximum area of the beam for all fields. If roughly circular areas were only due to beam-like or modestly resolved compact emission, we would not expect to see points in the top-right quadrant (defined by the two dashed lines) that instead contains ~44% of the fields with $Q_{5\sigma}^{\text{hull}} \geq 0.8$. We then conclude that the $Q_{5\sigma}^{\text{hull}}$ metric contains valuable information to classify the continuum emission morphology.

From a qualitative viewpoint, both the $\Delta_{5\sigma}^{\text{circ}}$ and the $Q_{5\sigma}^{\text{hull}}$ metrics suggest that a more complex dust spatial structure is observed as the clumps evolve. This is fully compatible with a concurrent contribution from increasing column density and/or increasing temperature of the dust with L/M due to radiative (and possibly dynamical) feedback from the forming protostars. In both cases, the millimeter flux would increase, causing more and more material to become detectable above the noise. Figure 9 shows that indeed the area of detectable emission increases with L/M (or evolution), and Figs. 13, 16, and 17 show that this larger area deviates more and more from a roughly circular shape and gets more and more irregular and complex with L/M. In other words as the clumps evolve we do not find relevant occurrences of large and circular areas of emission. This certainly agrees with a dynamically evolving clump fragmentation, also influenced by increasing levels of feedback. Conversely, these findings do not support the notion of relatively massive seeds of fragmentation that regularly grow in extension with increasing column density or temperatures. In Table 2, available at the CDS, contains some of the parameters we have presented in this section; namely, for each field, the emission area above 5σ level both in physical units and as fraction of the field of view, plus the $\Delta_{5\sigma}^{\text{circ}}$ and the $Q_{5\sigma}^{\text{hull}}$ parameters determined for the largest detected structure in each field.

5. Forward outlook

ALMAGAL is a game changer in areas like fragmentation statistics and morphologies, gas dynamics in infall and outflow motions in the clumps, chemistry, and systematic disk detections, among others. Here, we present a brief summary of the more

immediate works that are already published or will be submitted and published in the next six months. Wells et al. (2024) published a first study of flow dynamics in a subsample of 100 ALMAGAL fields based on H_2CO lines. Coletta et al. (2025) presents the first release of a continuum compact source catalog, which includes the core physical properties and a first discussion of the core mass function in evolutionary terms. Elia et al. (in prep.) will discuss core properties in the context of the hosting clumps, while Schisano et al. (in prep.) will present an in-depth analysis of the morphological distribution of cores and their mass segregation. Wallace et al. (in prep.) will investigate the multiscale morphological properties of continuum emission, greatly extending the preliminary analysis presented in this paper. Mininni et al. (2025) discusses the multiscale morphological relationship between the dust continuum and molecular line emission of CH_3CN, CH_3OH, H_2CO, HC_3N, and other tracers. Benedettini et al. (in prep.) analyses the relationship between the distribution of v_{LSR} of molecular lines and the dynamical properties of the hosting clumps gas. An investigation of the $^{13}C/^{12}C$ Galactic gradient will be presented by Law et al. (in prep.), while Stroud et al. (in prep.) will discuss the evolutionary properties of continuum cores in the context of the association with methanol masers. Finally, Jones et al. (in prep.) and Allande et al. (in prep.) will present a first catalog of molecular lines, complete with the derived gas physical properties, toward the 1.38 mm cores of Coletta et al. (2025).

Extending the forward look toward 2026, we anticipate studies on complex organic molecules (Allande et al., in prep.), line-based physical and evolutionary studies based on machine-learning methodologies (Sanchez-Monge et al., in prep.) and cores virial analysis (Mininni et al., in prep.). Finally, the ALMAGAL data have also proved effective for preliminary studies of disks towards the brightest cores that led to follow-up studies that will be presented by Ahmadi et al. (in prep.).

A first complete public release of the 7M+TM2+TM1 and 7M+TM2 continuum images will be done after the acceptance of the first round of continuum-based papers introduced above, and we anticipate this to take place no later than December 2025. A first round of papers with a preliminary exploitation of the 7M+TM2+TM1 and 7M+TM2 spectroscopic datacubes is in relatively less advanced stage, so that their public release takes place no later than December 2026.

6. Conclusions

The ALMAGAL Large Program has mapped 1013 dense star-forming clumps in the 1.38 mm continuum and lines to obtain the first statistically significant demographic description of the fragmentation process that leads to the formation of dense cores. In this first paper, we present a complete review of the target clumps physical properties, with a first analysis of the morphological properties of the dust continuum emission. We find that metrics such as the perimeter and area ratio, along with the convex hull and area ratio, are effective in mapping the emergence of complex irregular morphologies proportionally to the evolutionary stage of the clumps and their surface density.

Data availability

Tables 1 and 2 are available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/696/A149

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
Anderson, M., Peretto, N., Ragan, S. E., et al. 2021, MNRAS, 508, 2964
Avison, A., Fuller, G. A., Frimpong, N. A., et al. 2023, MNRAS, 526, 2278
Berg, M., van Kreveld, M., Overmars, M., & Schwarzkopf, O. 2008, Computational Geometry: Algorithms and Applications, 3rd edn. (Springer)
Beuther, H., Henning, T., Linz, H., et al. 2015, A&A, 581, A119
Beuther, H., Mottram, J. C., Ahmadi, A., et al. 2018, A&A, 617, A100
Bolatto, A. D., Warren, S. R., Leroy, A. K., et al. 2013, Nature, 499, 450
Bonnell, I. A., & Bate, M. R. 2006, MNRAS, 370, 488
Bonnell, I. A., Larson, R. B., & Zinnecker, H. 2007, Protostars and Planets V, 149
Cesaroni, R., Pestalozzi, M., Beltrán, M. T., et al. 2015, A&A, 579, A71
Chen, H.-R. V., Zhang, Q., Wright, M. C. H., et al. 2019, ApJ, 875, 24
Coletta, A., Molinari, S., Schisano, E., et al. 2025, A&A, 696, A151
Commercon, B., Hennebelle, P., & Henning, T. 2011, ApJ, 742, L9
Contreras, Y., Sanhueza, P., Jackson, J. M., et al. 2018, ApJ, 861, 14
Csengeri, T., Bontemps, S., Wyrowski, F., et al. 2017, A&A, 600, L10
Draine, B. T. 2003, ARA&A, 41, 241
Duarte-Cabral, A., Bontemps, S., Motte, F., et al. 2013, A&A, 558, A125
Elia, D., Molinari, S., Schisano, E., et al. 2017, MNRAS, 471, 100
Elia, D., Merello, M., Molinari, S., et al. 2021, MNRAS, 504, 2742
Girichidis, P., Federrath, C., Banerjee, R., & Klessen, R. S. 2012, MNRAS, 420, 613
Gong, H., & Ostriker, E. C. 2009, ApJ, 699, 230
He, Y.-X., Zhou, J.-J., Esimbek, J., et al. 2015, MNRAS, 450, 1926
Hunter, G. H., Sormani, M. C., Beckmann, J. P., et al. 2024, A&A, 692, A216
Irabor, T., Hoare, M., Burton, M., et al. 2023, MNRAS, 520, 1073
Ishihara, K., Sanhueza, P., Nakamura, F., et al. 2024, ApJ, accepted [arXiv:2407.06845]
Jeans, J. H. 1902, Philos. Trans. Roy. Soc. Lond. Ser. A, 199, 1

Jijina, J., & Adams, F. C. 1996, ApJ, 462, 874
Kauffmann, J., & Pillai, T. 2010, ApJ, 723, L7
Keto, E., & Wood, K. 2006, ApJ, 637, 850
Klessen, R. S. 2000, ApJ, 535, 869
Klessen, R. S., & Glover, S. C. O. 2016, Saas-Fee Adv. Course, 43, 85
Klaassen, P. D., Testi, L., & Beuther, H. 2012, A&A, 538, A140
Konyves, V., André, P., Men'shchikov, A., et al. 2010, A&A, 518, L106
Krumholz, M. R., Klein, R. I., McKee, C. F., Offner, S. S. R., & Cunningham, A. J. 2009, Science, 323, 754
Krumholz, M. R., Bate, M. R., Arce, H. G., et al. 2014, Protostars and Planets VI, 243
Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57
Larson, R. B. 1985, MNRAS, 214, 379
Liu, T., Evans, II, N. J., Kim, K., et al. 2020, MNRAS, 496, 2790
Lumsden, S. L., Hoare, M. G., Urquhart, J. S., et al. 2013, ApJS, 208, 11
Maeder, A., & Meynet, G. 2000, ARA&A, 38, 143
McKee, C. F., & Tan, J. C. 2002, Nature, 416, 59
McKee, C. F., & Tan, J. C. 2003, ApJ, 585, 850
Mege, P., Russeil, D., Zavagno, A., et al. 2021, A&A, 646, A74
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, submitted
Molinari, S., Pezzuto, S., Cesaroni, R., et al. 2008, A&A, 481, 345
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., Baldeschi, A., Robitaille, T. P., et al. 2019, MNRAS, 486, 4508
Motte, F., Nony, T., Louvet, F., et al. 2018, Nat. Astron., 2, 478
Motte, F., Bontemps, S., Csengeri, T., et al. 2022, A&A, 662, A8
Nonhebel, M., Barnes, A. T., Immer, K., et al. 2024, A&A, 691, A70
Nony, T., Louvet, F., Motte, F., et al. 2018, A&A, 618, L5
Ossenkopf, V., & Henning, T. 1994, A&A, 291, 943
Palau, A., Ballesteros-Paredes, J., Vázquez-Semadeni, E., et. al. 2015, MNRAS, 453, 3785
Peretto, N., Fuller, G. A., Duarte-Cabral, A., et al. 2013, A&A, 555, A112
Peters, T., Banerjee, R., Klessen, R. S., et al. 2010, ApJ, 711, 1017
Peters, T., Banerjee, R., Klessen, R. S., & Mac Low, M.-M. 2011, ApJ, 729, 72
Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, ARA&A, 48, 431
Preibisch, T., Ossenkopf, V., Yorke, H., & Henning, T. 1993, A&A, 279, 577
Purcell, C. R., Hoare, M. G., Cotton, W. D., et al. 2013, ApJS, 205, 1
Rees, M. J. 1976, MNRAS, 176, 483
Rigby, A. J., Peretto, N., Anderson, M., et al. 2024, MNRAS, 528, 1172
Rygl, K. L. J., Wyrowski, F., Schuller, F., & Menten, K. M. 2013, A&A, 549, A5
Sanchez-Monge, A., Brogan, C. L., Hunter, T. R., et al. 2025, A&A, 696, A150
Sanhueza, P., Contreras, Y., Wu, B., et al. 2019, ApJ, 886, 102
Silverman, B. W. 1981, J. Roy. Stat. Soc., 43, 97
Smith, R. J., Clark, P. C., & Bonnell, I. A. 2009, MNRAS, 396, 830
Svoboda, B. E., Shirley, Y. L., Traficante, A., et al. 2019, ApJ, 886, 36
Tackenberg, J., Beuther, H., Henning, T., et al. 2012, A&A, 540, A113
Tan, J. C., Beltrán, M. T., Caselli, P., et al. 2014, in Protostars and Planets VI, eds. H. Beuther, R. Klessen, C. P. Dullemond, & T. Henning, 149
Traficante, A., Fuller, G. A., Billot, N., et al. 2017, MNRAS, 470, 3882
Traficante, A., Duarte-Cabral, A., Elia, D., et al. 2018, MNRAS, 477, 2220
Traficante, A., Fuller, G. A., Duarte-Cabral, A., et al. 2020, MNRAS, 491, 4310
Traficante, A., Jones, B. M., Avison, A., et al. 2023, MNRAS, 520, 2306
Urquhart, J. S., Koenig, C., Giannetti, A., et al. 2018, MNRAS, 473, 1059
Urquhart, J. S., Wells, M. R. A., Pillai, T., et al. 2022, MNRAS, 510, 3389
Vázquez-Semadeni, E., Palau, A., Ballesteros-Paredes, J., Gómez, G. C., & Zamora-Avilés, M. 2019, MNRAS, 490, 3061
Veilleux, S., Cecil, G., & Bland-Hawthorn, J. 2005, ARA&A, 43, 769
Wells, M. R. A., Urquhart, J. S., Moore, T. J. T., et al. 2022, MNRAS, 516, 4245
Wells, M. R. A., Beuther, H., Molinari, S., et al. 2024, A&A, 690, A185
Wu, J., & Evans, II, N. J. 2003, ApJ, 592, L79
Xu, F., Wang, K., Liu, T., et al. 2024, ApJS, 270, 9
Yorke, H. W., & Sonnhalter, C. 2002, ApJ, 569, 846
Zhang, Q., Wang, K., Lu, X., & Jimenez-Serra, I. 2015, ApJ, 804, 141
Zinnecker, H., & Yorke, H. W. 2007, ARA&A, 45, 481

[1] Istituto Nazionale di Astrofisica (INAF)-Istituto di Astrofisica e Planetologia Spaziale, Via Fosso del Cavaliere 100, 00133 Roma, Italy
[2] Physikalisches Institut der Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany
[3] University of Connecticut, Department of Physics, 196A Auditorium Road, Unit 3046, Storrs, CT 06269, USA
[4] Institute of Astronomy and Astrophysics, Academia Sinica, 11F of ASMAB, AS/NTU No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan

5 Institut de Ciències de l'Espai (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, 08193 Bellaterra (Barcelona), Spain

6 Institut d'Estudis Espacials de Catalunya (IEEC), 08860, Castelldefels (Barcelona), Spain

7 Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany

8 SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, SK11 9FT, UK

9 Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

10 Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Straße 2, 69120 Heidelberg, Germany

11 Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

12 Center for Data and Simulation Science, University of Cologne, Cologne, Germany

13 Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

14 Radcliffe Institute for Advanced Studies at Harvard University, 10 Garden Street, Cambridge, MA 02138, USA

15 National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

16 Dipartimento di Fisica, Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy

17 Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

18 Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany

19 Istituto Nazionale di Astrofisica (INAF), Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, Florence, Italy

20 Max-Planck-Institute for Extraterrestrial Physics (MPE), Garching bei München, Germany

21 Laboratoire d'Études du Rayonnement et de la Matière en Astrophysique et Atmosphères (LERMA), Observatoire de Paris, Meudon, France

22 UK Astronomy Technology Centre, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK

23 School of Engineering and Physical Sciences, Isaac Newton Building, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK

24 School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK

25 Université Paris-Saclay, Université Paris-Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France

26 Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, China

27 INAF-Istituto di Radioastronomia & Italian ALMA Regional Centre, Via P. Gobetti 101, 40129 Bologna, Italy

28 Faculty of Physics, University of Duisburg-Essen, Lotharstraße 1, 47057 Duisburg, Germany

29 Jülich Supercomputing Centre, Forschungszentrum Jülich, Wilhelm-Johnen-Straße, Jülich, 52428, NRW, Germany

30 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

31 Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Meguro, Tokyo 152-8551, Japan

32 Cardiff Hub for Astrophysics Research & Technology, School of Physics & Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff CF24 3AA, UK

33 National Astronomical Observatory of Japan, National Institutes of Natural Sciences, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

34 National Radio Astronomy Observatory, PO Box O, Socorro, NM 87801, USA

35 SRON Netherlands Institute for Space Research, Groningen, The Netherlands

36 Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands

37 Universidad Autonoma de Chile, Pedro de Valdivia 425, Providencia, Santiago de Chile, Chile

38 Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80389, USA

39 Dipartimento di Fisica e Astronomia, Alma Mater Studiorum – Università di Bologna, Bologna, Italy

40 Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea

41 University of Science and Technology, Korea (UST), 217 Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

42 European Southern Observatory, Karl-Schwarzschild Str. 2, 85748 Garching bei München, Germany

43 Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada

44 Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile

45 Centro de Astro-Ingeniería (AIUC), Pontificia Universidad Católica de Chile, Av. Vicuña Mackena 4860, Macul, Santiago, Chile

46 Department of Astronomy, Yunnan University, Kunming 650091, PR China

47 Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Rome, Italy

Appendix A: ALMAGAL Clumps Properties

Here, we give a brief description of the table, available at the CDS, containing the consolidated properties of the ALMAGAL targets revised as explained in the main text:

- Col. 1: an ordinal number of the field
- Cols. 2–3: unique ALMAGAL source name file ID together with the original designations either in the Hi-GAL band-merged catalog ("HBM" prefix) ot the RMS survey ("G" prefix)
- Cols. 4–5: Galactic coordinates of the fields center
- Col 6: the target membership to either "near" or "far" ALMAGAL subsamples (col. 6) defining the ALMA antenna configurations used and hence the minimum achieved angular resolution ($0''.15$ and $0''.3$ respectively)
- Col. 7: revised v_{LSR}
- Col. 8: revised heliocentric distance
- Cols. 9–10: Clump mass and bolometric luminosity from Elia et al. (2021), or derived with the same methodology (see Sect. 3.2.2 above) rescaled at the revised distance of Col. 8
- Cols. 11–12: cold dust temperature and bolometric temperature as in Elia et al. (2021), or derived with the same methodology.
- Col. 13: clump surface density reported from Elia et al. (2021), with the difference that the area used for the computation is based on the measured clump size without deconvolving the *Herschel* beam.
- Col. 14: Evolutionary class after Urquhart et al. (2022): Q (quiescent), P (protostellar), Y (young stellar object), and H (HII region). Cases where the methodology failed to identify a specific class are labeled "n/a".

The last column may contain a series of strings describing some specificities about the revision process of the target properties. The strings have the following meanings:

- **-kda_defnear**: source originally set at FAR kinematic distance solution by Mege et al. 2021 in absence of any KDA discriminator. In this work the NEAR distance decision is instead adopted for these sources, for consistency with the approach followed for the sources from the RMS sample.
- **-V7**: the new ALMAGAL-based v_{LSR} differs from the original by more than 7 km s^{-1}
- **-D500**: the new ALMAGAL-based distance differs from the original by more than 500pc.
- **-Dn2f**: the new distance determined from the v_{LSR} of ALMAGAL dense tracer lines, switches this source from the NEAR to the FAR sample. As originally belonging to the NEAR sample it was observed with a maximum 0.3 arcsec resolution, therefore now resulting in a sub-optimal linear resolution.
- **-Df2n**: the new distance determined from the v_{LSR} of ALMAGAL dense tracer lines, switches this source from the FAR to the NEAR sample. As originally belonging to the FAR sample it was observed with a maximum 0.15 arcsec resolution, therefore now resulting in a super-optimal linear resolution.
- **-D>8**: the new distance determined from the v_{LSR} of ALMAGAL dense tracer lines, puts the source beyond the distance limit used to select the original sample.
- **-refit40k**: the dust temperature estimated from *Herschel* fluxes has been updated here with respect to Elia et al. (2017, 2021) due to a 40K hard limit that was reached in the grey-body fit.
- **-sed_irr**: RMS source with a Hi-GAL counterpart in the *Herschel* 5 bands, but the SED is irregular (not a single downward-facing concavity) and did not pass criteria to be included in Elia et al. (2017, 2021). This criteria have been relaxed for the present work to estimate physical parameters.
- **-sed_rebuilt**: RMS source not in Elia et al. (2017, 2021), that needed a custom build of the *Herschel* SED to estimate physical parameters in the present work. Possible reasons: source had multiple counterparts in one *Herschel* band, or counterparts were not linked in a single SED.
- **-sed_sat**: RMS source not in Elia et al. (2017, 2021) due to saturation in some *Herschel* bands. To estimate physical parameters for the present work the SED has been manually built allowing missing bands.
- **-sed_noband**: RMS source not in Elia et al. (2017, 2021) because of missing 350–500 µm*Herschel* bands due to source faintness. Physical parameters for the present work are estimated using three remaining bands
- **-no_props**: RMS source not in Elia et al. (2017, 2021) because has a *Herschel* counterpart in 1 or 2 bands only. It was not possible to estimate physical parameters.

Appendix B: Plots for distance bias characterization and other supporting material

This section reports useful plots for the characterization of observational biases and to support the conclusions in the various relationships explored in this work. They are referenced where needed in the body of the paper.
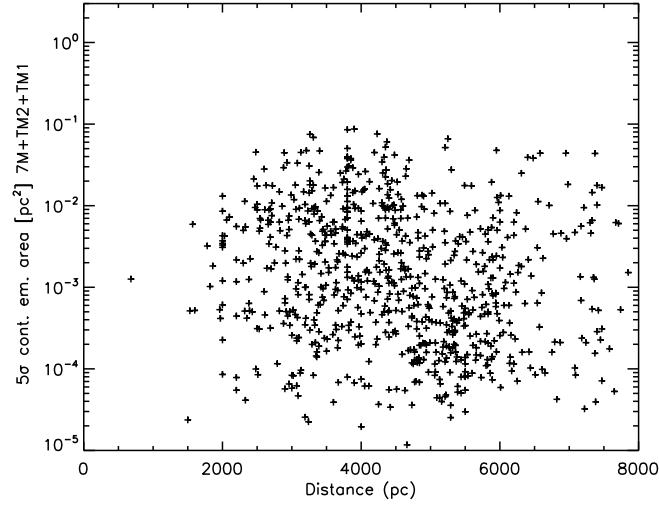

Fig. B.1: ALMA 1.38'mm total emission area above 5σ in the 7M+TM2+TM1 images as a function of source distance. No relationship can be seen.
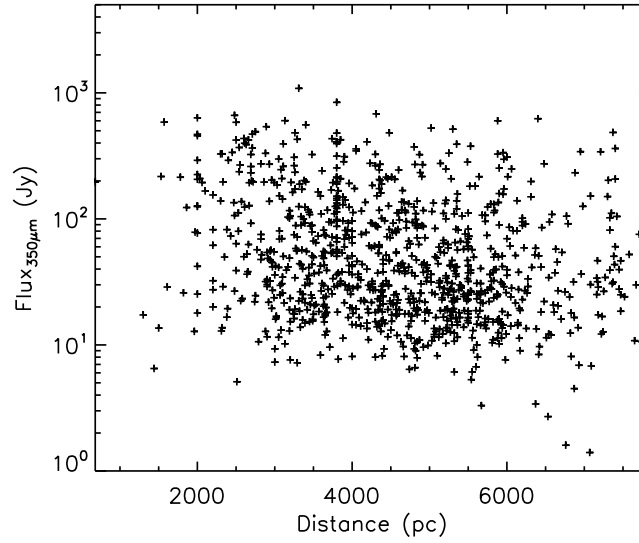

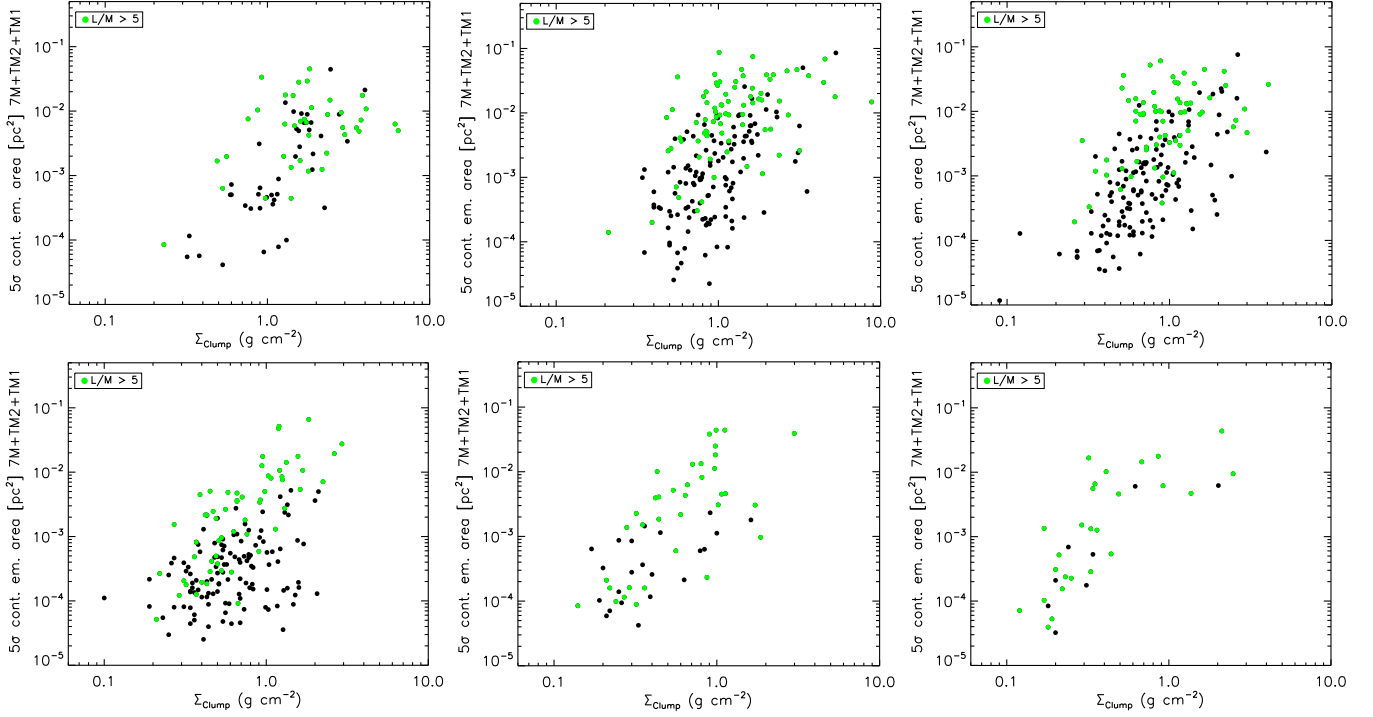Fig. B.2: *Herschel* 350µm flux for each clumps as a function of clump distance. No relationship can be seen.

Fig. B.3: 1.38mm continuum emission area in 5σ contour vs clump surface density, as in Fig.8, but separately in different distance 1kpc-bins. Bins center distance is from 2.5 to 4.5kpc (top row, left to right), and from 5.5 to 7.5kpc (bottom row, left to right).
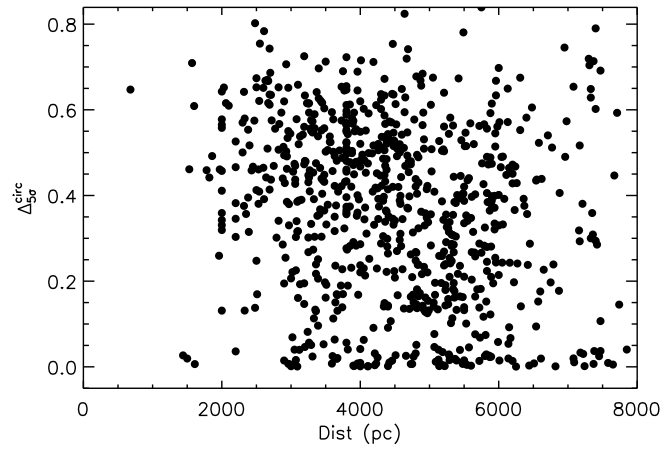

Fig. B.4: $\Delta_{5\sigma}^{circ}$ parameter for the total RoIs 5σ area as a function of clump distance. No relationship can be seen.
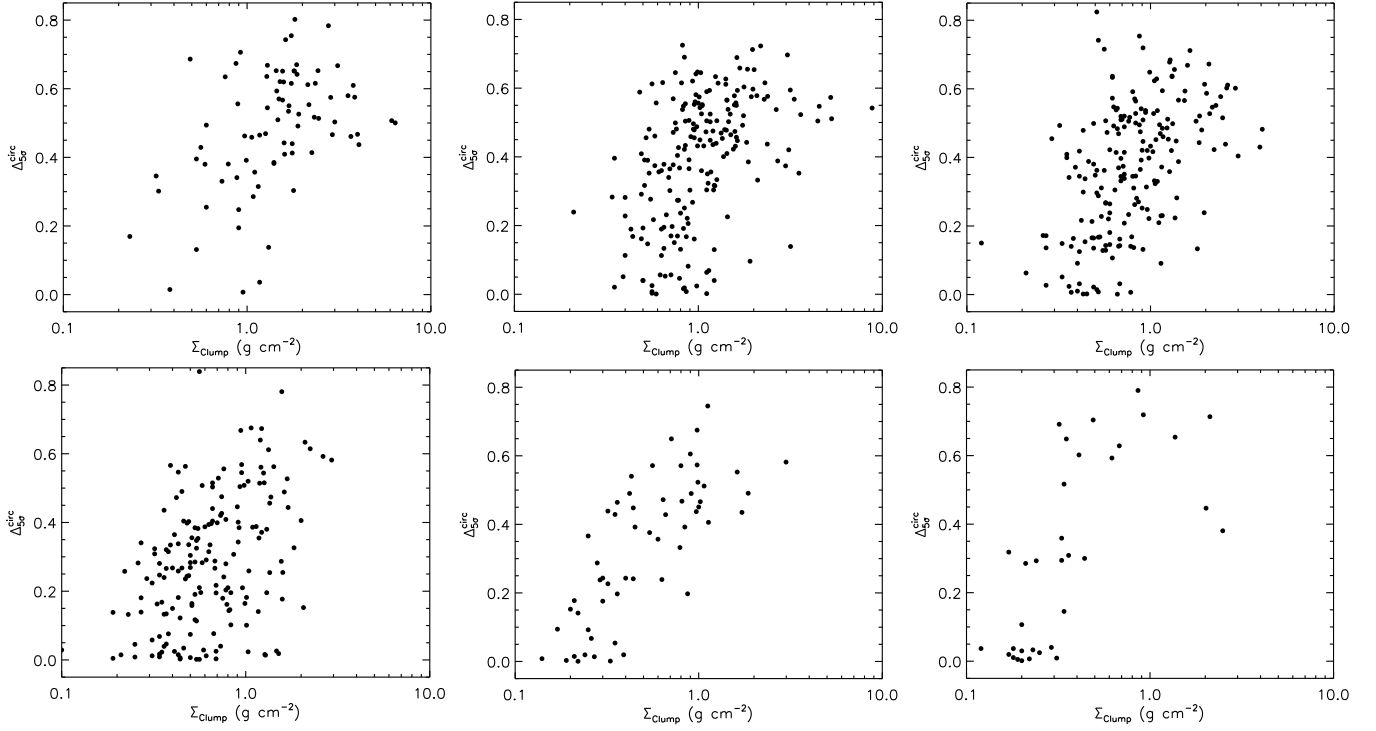
Fig. B.5: Departure from circularity, $\Delta_{5\sigma}^{\rm circ}$, of the total of the 5σ ALMA emission RoIs in each field as a function of the clump surface density, as in Fig.12, but separately in different distance 1kpc-bins. Bins center distance is from 2.5 to 4.5kpc (top row, leftto right), and from 5.5 to 7.5kpc (bottom row, left to right).
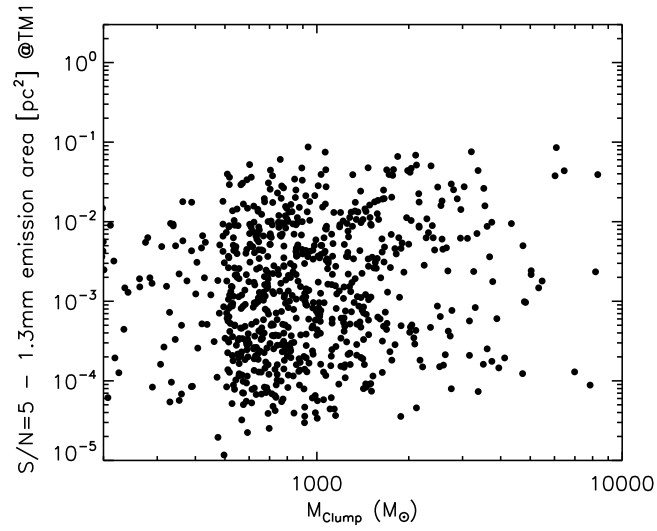

Fig. B.6: 1.38mm continuum emission area in 5σ contour of 7M+TM2+TM1 images versus clump Mass. No correlation is present.