LIVERPOOL JOHN MOORES UNIVERSITY

LJMU Research Online

Urquhart, JS, Wells, MRA, Pillai, T, Leurini, S, Giannetti, A, Moore, TJT, Thompson, MA, Figura, C, Colombo, D, Yang, AY, Koenig, C, Wyrowski, F, Menten, KM, Rigby, AJ, Eden, DJ and Ragan, SE

 ATLASGAL - evolutionary trends in high-mass star formation

http://researchonline.ljmu.ac.uk/id/eprint/17533/

Article

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/

ATLASGAL – evolutionary trends in high-mass star formation

J. S. Urquhart [1]★ M. R. A. Wells,[1] T. Pillai,[2] S. Leurini,[3] A. Giannetti,[4] T. J. T. Moore,[5]
M. A. Thompson,[6] C. Figura,[7] D. Colombo,[8] A. Y. Yang,[8] C. König,[8] F. Wyrowski,[8] K. M. Menten,[8]
A. J. Rigby [9] D. J. Eden [5] and S. E. Ragan [9]

[1]*Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK*
[2]*Institute for Astrophysical Research, Boston University, Boston, MA 02215, USA*
[3]*INAF – Osservatorio Astronomico di Cagliari, Via della Scienza 5, I-09047 Selargius (CA), Italy*
[4]*INAF – Istituto di Radioastronomia, Via P. Gobetti 101, I-40129 Bologna, Ital*
[5]*Astrophysics Research Institute, Liverpool John Moores University, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[6]*School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK*
[7]*Wartburg College, Waverly, IA 50677, USA*
[8]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-121 Bonn, Germany*
[9]*School of Physics and Astronomy, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, UK*

ABSTRACT

ATLASGAL is an 870-μm dust survey of 420 $\deg^2$ the inner Galactic plane and has been used to identify ∼10 000 dense molecular clumps. Dedicated follow-up observations and complementary surveys are used to characterize the physical properties of these clumps, map their Galactic distribution, and investigate the evolutionary sequence for high-mass star formation. The analysis of the ATLASGAL data is ongoing: We present an up-to-date version of the catalogue. We have classified 5007 clumps into four evolutionary stages (quiescent, protostellar, young stellar objects and H II regions) and find similar numbers of clumps in each stage, suggesting a similar lifetime. The luminosity-to-mass (L_{bol}/M_{fwhm}) ratio curve shows a smooth distribution with no significant kinks or discontinuities when compared to the mean values for evolutionary stages indicating that the star formation process is continuous and that the observational stages do not represent fundamentally different stages or changes in the physical mechanisms involved. We compare the evolutionary sample with other star formation tracers (methanol and water masers, extended green objects and molecular outflows) and find that the association rates with these increases as a function of evolutionary stage, confirming that our classification is reliable. This also reveals a high association rate between quiescent sources and molecular outflows, revealing that outflows are the earliest indication that star formation has begun and that star formation is already ongoing in many of the clumps that are dark even at 70 μm.

Key words: stars: early-type – stars: formation – ISM: evolution – Galaxy: kinematics and dynamics – submillimetre: ISM.

1 INTRODUCTION

High-mass stars are a key component of many areas of astrophysics. They dominate the energy budget of galaxies, driving their evolution and ultimately that of the Universe as a whole (Kennicutt & Evans 2012). They are also responsible for the production of all heavy elements, which are returned to the interstellar medium through their strong stellar winds and in supernovae. This action changes the composition and chemistry of their local environments and provides the raw material required for new generations of star and planet formation (Zinnecker & Yorke 2007). It is therefore crucial to understand the high-mass star formation process, not only as a stand-alone theory, but also as a tool to understand both the history and future of massive star clusters and to allow us to construct complex models of the evolution of the Universe via numerical simulations (Motte, Bontemps & Louvet 2018).

High-mass stars are rare and only contribute a few per cent of the stellar population, normally characterized by the initial mass function (IMF; Kroupa 2001). A consequence of their rarity is that regions of high-mass star formation are statistically less likely to be found nearby, with most having distances >2 kpc. They evolve very rapidly, reaching the main sequence while still deeply embedded, and so the earliest stages take place behind many hundreds of magnitudes of visual extinction and thus remain hidden at visual and near- and even mid-infrared wavelengths. High-mass stars nearly always form in clusters, which significantly complicates the identification and analysis of individual protostellar objects. All of these complications have hindered the development of a comprehensive evolutionary framework for high-mass star formation.

A first step to improve our understanding of high-mass star formation is the availability of high-resolution images at submillimetre wavelengths over large areas (Zinnecker & Yorke 2007). These are now readily available for the whole Galactic mid-plane (HiGAL; Molinari et al. 2010) and the inner Galactic plane (APEX Telescope

Large Area Survey of the Galaxy ATLASGAL; Schuller et al. 2009) with resolutions between ∼10 and 40 arcsec. These provide unbiased samples of dense clumps identified by their dust emission that are a fundamental starting point for star formation as well as being intimately associated with all stages in the process.

The ATLASGAL survey is an unbiased 870-μm submillimetre continuum survey of the inner Galactic plane ($300° < \ell < 60°$ and $|b| < 1°.5$, with a spatial resolution of 19 arcsec) conducted with the 12-m Atacama Pathfinder Experiment (APEX) telescope (Güsten et al. 2006). This survey was later extended to include the Carina tangent ($280° < \ell < 300°$ and $-2° < b < 1°$). It has provided a large inventory of dense molecular clumps (∼10 000 clumps; Contreras et al. 2013; Csengeri et al. 2014; Urquhart et al. 2014b) and includes samples of sources in all of the early evolutionary stages associated with high-mass star formation, from very young starless objects to evolved H II regions that are starting to break out of their natal clumps (König et al. 2017; Urquhart et al. 2014a, 2018).

The ATLASGAL compact source catalogue (CSC; Contreras et al. 2013; Urquhart et al. 2014b) consists of ∼10 000 dense clumps. Accurate categorisation of the ATLASGAL catalogue has the potential not only to provide large samples of sources in all of the main evolutionary stages and allow them to be robustly compared to each other, but will also provide a more complete picture of star formation activity across the inner Galactic disc. The ATLASGAL CSC has therefore been the focus of an intensive campaign to characterize the properties and evolutionary state of the clumps. These have included dedicated molecular-line studies to produce the kinematics, temperatures, chemistry and determine kinematic distances (Wienen et al. 2012, 2015, 2018; Giannetti et al. 2014; Csengeri et al. 2016; Kim et al. 2017, 2018; Tang et al. 2018; Navarete et al. 2019; Urquhart et al. 2019) and detailed analysis of clumps associated with star formation tracers such as H II regions (Urquhart et al. 2013b), methanol masers (Urquhart et al. 2013a, 2015; Billington et al. 2019, 2020), massive young stellar objects (MYSO; Urquhart et al. 2014b; König et al. 2017; Urquhart et al. 2018) and filamentary structures (Li et al. 2016; Mattern et al. 2018; Lin et al. 2019). This programme of follow-up observations and use of complementary multiwavelength studies has resulted in the ATLASGAL sample being the most well-characterized sample of high-mass star-forming clumps currently available.

This characterization of the sample is ongoing as new data become available and our analysis improves. In this paper, we provide an updated version of the catalogue presented in Urquhart et al. (2018). We have used new radial velocity measurements taken from the SEDIGISM (Structure, Excitation and Dynamics of the Inner Galactic Interstellar Medium) survey (Schuller et al. 2017, 2021; Duarte-Cabral et al. 2021) to derive distances to ∼900 clumps (∼600 new distances and ∼300 distances have been updated). We also describe a robust set of criteria that has been applied to multiwavelength mid- and far-infrared images to determine accurately the evolutionary stages and properties of the clumps. The main aim of this work is to produce a high-reliability sample of star-forming clumps that are well characterized in terms of their physical properties and their evolutionary stage, and will serve as a starting point for more detailed investigations.

The structure of this paper is as follows: We describe how the velocities and distances have been determined in Section 2. In Section 3, we describe how the physical parameters are derived. We outline our evolutionary classification criteria in Section 4 and compare the physical properties characterizing various stages identified. In Section 5, we present the correlation between various parameters and investigate the evolutionary sequence for high-mass

star formation. We summarize the work presented and outline future work in Section 6.

2 RADIAL VELOCITIES AND DISTANCES

2.1 Radial velocities

Urquhart et al. (2018) combined the results from a number of molecular-line surveys and targeted observations towards ∼1000 ATLASGAL clumps to provide velocities and distances to nearly 8000 clumps. This study focused on clumps located more than 5° away from the Galactic Centre (GC) as few high-resolution molecular-line surveys were available for this region. This situation has improved significantly in recent years with data now becoming available from the SEDIGISM and CHIMPS2 surveys (Schuller et al. 2021 and Eden et al. 2020, respectively). These new surveys allow us to extend our previous work into the GC region.

The SEDIGISM survey is a ^{13}CO and C^{18}O (2–1) survey conducted with the APEX telescope.[1] In a previous paper (Urquhart et al. 2021), we combined the ATLASGAL sample of clumps with the reduced data cubes provided by the SEDIGISM catalogue to extract spectra towards clumps where the two surveys overlapped (i.e. $300° < \ell < 18°$ and $|b| < 0°.5$). This comparison resulted in 5148 spectra being extracted and analysed, allowing us to assign velocities to an additional 1108 ATLASGAL clumps and extend our velocity coverage into the GC region (i.e. $|\ell| < 5°$). In addition, we also include the 31 clumps for which radial velocities have been taken from the literature.

A comparison of the velocities assigned from the analysis of the SEDIGISM data to those previously assigned found agreement within 3 km s^{-1} for 3608 of the 3936 clumps (∼92 per cent; see Urquhart et al. 2021 for details) in the SEDIGISM region. The disagreements are due to multiple line-emission components being seen towards the clumps, which reduce the reliability of identifying the correct velocity component. In these cases, we were able to take advantage of the SEDIGISM data to create maps of the different velocity components, which were then compared to the dust emission to identify the component with the best morphological correlation. Lower resolution CO (1-0) data had been used to make the initial velocity assignments to many of the clumps with disagreements and the higher-resolution SEDIGISM observations were better able to resolve the line-of-sight confusion (see Urquhart et al. 2021 for more details). The high resolution and higher critical density of the ^{13}CO (2–1) line and the ability to match the morphology of gas and dust means velocities based on SEDIGISM data are more reliable than values from some of the other surveys used previously. We therefore adopted the SEDIGISM velocities for 269 of the 318 clumps where the velocities do not agree. The allocation of a velocity component was uncertain for the remaining 49, and so the previous velocities assigned to these clumps were discarded.

In total, there are 9817 clumps within the main ATLASGAL region (i.e. $300° < \ell < 60°$ and $|b| < 1°.5$) and we have now managed to assign a velocity to 8899 of these (∼90 per cent of the sample). Many of the remaining clumps are located within the GC region where the spectra have either poor baselines or are very broad (>30 km s^{-1}) and often blended components making any velocity allocation difficult (Urquhart et al. 2021). We are now, therefore, as

---

[1]Other transitions have also been covered. These are much weaker in general and only detected towards the brightest ATLASGAL sources and are less useful; for more details, see Schuller et al. (021).

Table 1. Summary of the kinematic distance solutions for the 8417 AT-LASGAL sources located in the Galactic disc ($\ell = 300°$–$357°$ and $3°$–$60°$).

Step	Description of method	Total number distances assigned
(i)	Parallax/spectroscopic	114
(ii)	Outer galaxy	136
(iii)	Tangent	766
(iv)	Z distance	1280
(v)	H I EA	76
(vi)	H I SA near (?)	2725 (174)
(vi)	H I SA far (?)	904 (182)
(vii)	IRDC associations	39
(viii)	Literature	1707
(ix)	Ambiguous (solar circle)	325 (61)
(x)	No V_{LSR} available	345

Note. The Roman numerals (i)–(viii) given in Column 1 refer to the various steps described in Urquhart et al. (2018) (see the text for brief summary), while the (ix) and (x) are used to identify sources where a distance could not be determined and clumps for which no velocity is available, respectively.

complete with regards to velocities for the ATLASGAL catalogue as we are likely to get.

2.2 Kinematic distances

We have new or updated velocities for 1408 clumps taken from a combination of velocities obtained from the literature and from SEDIGISM (Urquhart et al. 2021). This sample has allowed us to determine radial velocities for many sources toward the GC ($355°$ $< \ell < 5°$) that were excluded from our previous paper. We will use these radial velocities to determine kinematic distances, which will subsequently be used to calculate their physical properties and investigate their Galactic distribution. We still, however, need to exclude sources within a few degrees of the GC (i.e. $357° < \ell < 3°$) because kinematic distances for sources in this region are extremely unreliable. This exclusion reduces the number of sources for which we need to determine distances to 581.

We have given a detailed description of the method used to determine distances to the ATLASGAL clumps in Urquhart et al. (2018), and so we only provide a brief outline of the main steps here. We combine the radial velocity measurements with a model of the Galactic rotation to obtain a kinematic distance. We have used the Galactic rotation curve of Reid et al. (2016), which has utilized the results of maser parallax distances measured by the Bar and Spiral Structure Legacy (BeSSeL) Survey (Reid et al. 2019).[2] An unavoidable consequence of using a rotation model is that there are two possible solutions for any given radial velocity for sources located inside the solar circle: These are located at equal distances on either side of the tangent distance, and are commonly referred to as the near and far distances.

In Table 1, we provide a summary of the different methods used and the number of sources to which these have been applied. Here we give a brief description of the methods: (i) For sources with a reliable spectroscopic (Moisés et al. 2011) or maser parallax distance (Reid et al. 2019), these have been adopted; (ii) sources located outside the Solar circle (i.e. Galactocentric radius (R_{gc}) >8.35 kpc) are not affected by distance ambiguity; (iii) sources located within 10 km s^{-1} of the tangent velocity are placed at

the tangent distance; (iv) sources where a far distance would place them more than 120 pc from the Galactic mid-plane are placed at the near distance as a far distance is very unlikely; (v) distance ambiguity resolved using the H I Emission-absorption method (H I EA; e.g. Kolpak et al. 2003; Araya et al. 2002; Urquhart et al. 2011); (vi) distance ambiguity resolved using the H I self-absorption method (H I SA; e.g. Jackson et al. 2002; Roman-Duval et al. 2009; Wienen et al. 2015): question marks indicate the number of sources where the resolution is less certain. We use archival data from the Southern and VLA Galactic Plane surveys (SGPS; McClure-Griffiths et al. 2005 and VGPS; Stil et al. 2006) for the H I distance resolutions; (vii) sources associated with an infrared dark cloud (IRDC), which are seen in silhouette against the bright mid-infrared Galactic background, are considered to be at the near distance; and (viii) distance solutions taken from the literature (see Table 2 for references).

We have conducted a comprehensive literature review of similar studies to confirm our distance analysis and to allocate distances where our methods have been unsuccessful. Comparisons with the distances reported by other studies were made in Urquhart et al. (2018) and distances were found to agree in ∼70–80 per cent of cases. As an additional check, we compare our distances to those reported by Elia et al. (2021) based on analysis of the HiGAL survey (Molinari et al. 2010). Restricting the matches to those within 20 arcsec and radial velocities within 5 km s^{-1}, we find agreement in the assigned velocities to be ∼72 per cent (i.e. distances agree within 2 kpc to allow for slight differences in the assigned radial velocities and the ways the distances are determined).

2.3 Identification of complexes

We have grouped the individual clumps together into complexes. This grouping is done by using friends-of-friends clustering analysis to associate clumps that are localized in longitude, latitude, and velocity (ℓbv-space). The initial groups identified by the algorithm are subsequently inspected to ensure reliability and that the distance solutions are self-consistent within the uncertainties associated with the different methods used. For example, the H I SA method is approximately 80 per cent reliable (Busfield et al. 2006) and so the individual distances within a grouping should have a similar level of agreement. For complexes where the distances are not consistent, we have repeated the clustering analysis using increasingly tighter angular and velocity offsets between clumps until structures that are coincident in position and distance are identified. Where a reliable distance can be assigned to a complex, it is also assigned to the clumps associated with it. This process is more fully described in Urquhart et al. (2018) and we refer the reader to that paper for more details.

We have re-run the analysis of the 776 associations identified in Urquhart et al. (2018), updating velocities for ∼300 clumps that were included in the previous analysis, and providing new velocities and distances for a similar number of clumps. This new analysis has identified 877 associations, 684 of which were previously identified, with 47 located in the $\ell = 355°$–$357°$ and $3°$–$5°$ ranges and are therefore new identifications. The modified and new velocities have resulted in the modification of 146 associations, in most cases, only very slightly. Of the 684 associations that were previously identified, the distances agree for all but 49, and so distances have been amended for ∼7 per cent of the associations. We provide a list of the associations identities and their positions, velocities, and derived distances in Table 3.

Table 2. Summary of the kinematic distance analysis.

ATLASGAL CSC name	V_{LSR} (km s^{-1})	V_{LSR} Ref.	Reid distances			Adopted kinematic solution			Association name	Distance (kpc)
			Bayesian (kpc)	Near (kpc)	Far (kpc)	Distance (kpc)	Solution flag^a	Dist. ref.^{b}		
AGAL003.613−00.104	36.2	Urquhart et al. (2021)	10.9	6.7	9.9	10.9	(vi)	...	...	...
AGAL003.651−00.127	7.9	Urquhart et al. (2021)	2.9	2.5	14.6	2.9	(vi)	...	G003.705−00.068	2.9
AGAL003.661−00.114	3.6	Urquhart et al. (2021)	2.9	1.5	16.7	2.9	(viii)	...	G003.705−00.068	2.9
AGAL003.683−00.096	7.5	Urquhart et al. (2021)	2.9	2.4	14.7	2.9	(vi)	...	G003.705−00.068	2.9
AGAL003.688−00.147	4.3	Urquhart et al. (2021)	2.9	1.6	16.3	2.9	(vi)	...	G003.705−00.068	2.9
AGAL003.689−00.274	−30.9	Urquhart et al. (2021)	18.8	...	...	18.8	(ii)	...	...	...
AGAL003.698−00.102	7.8	Urquhart et al. (2021)	2.9	2.4	14.6	2.9	(vi)	...	G003.705−00.068	2.9
AGAL003.719+00.016	9.1	Urquhart et al. (2021)	2.9	2.7	14.1	2.9	(viii)	...	G003.705−00.068	2.9
AGAL003.744+00.019	7.8	Urquhart et al. (2021)	2.9	2.4	14.6	2.9	(vi)	...	G003.705−00.068	2.9
AGAL003.746−00.071	4.0	Urquhart et al. (2021)	2.9	1.5	16.4	2.9	(vi)	...	G003.705−00.068	2.9

Notes. Only a small portion of the data is provided here. The full table is available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.125.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/.

^{a} The distance solution flags refer to the different steps described in Section 2 and Table 1.

^{b} References for distance solutions: (1) Anderson & Bania (2009): (2) Araya et al. (2002); (3) Battisti & Heyer (2014); (4) Urquhart et al. (2013b); (5) Dunham et al. (2011); (6) Fish et al. (2003), (7) Reid et al. 2014, (8) Sanna et al. (2014), (9) Wu et al. (2014), (10) Xu et al. (2009), (11) Fish et al. (2003), (12) Downes et al. 1980, (13) Giannetti et al. (2015), (14) Green & McClure-Griffiths (2011), (15) Immer et al. (2012), (16) Kolpak et al. (2003); (17) Pandian, Menten & Goldsmith (2009); (18) Sewilo et al. (2004); (19) Moisés et al. (2011); (20) Pandian, Momjian & Goldsmith (2008); (21) Roman-Duval et al. (2009); (22) Stead & Hoare (2010); (23) Sanna et al. (2009); (24) Sato et al. (2010); (25) Urquhart et al. (2012); (26) Watson et al. (2003); (27) Xu et al. (2011); (28) Zhang et al. (2013); (29) (Nagayama et al. (2011).

Table 3. Derived association parameters.

Association name	Literature name	# of members	ℓ (°)	b (°)	V_{LSR} (km s^{-1})	ΔV_{LSR} (km s^{-1})	Distance (kpc)	ΔDistance (kpc)	Ref.
G005.950−01.183	M8 – Lagoon Nebula	24	5.950	−1.183	12.4	2.0	0.95	...	Moisés et al. (2011)
G006.183−00.358	W28	78	6.183	−0.358	12.6	4.9	3	0.2	Reid et al. (2019)
G008.581−00.329	W30 (SNR?)	38	8.581	−0.329	37.1	1.3	4.45	0.23	...
G010.147−00.237	W31-south	65	10.147	−0.237	16.9	11.6	2.9	0.3	Moisés et al. (2011)
G010.474−00.012	G010.47+00.02	10	10.474	−0.012	68.4	2.8	8.5	0.6	Reid et al. (2019)
G010.639−00.400	W31C (North)	17	10.639	−0.400	−2.9	1.2	3.9	0.5	Reid et al. (2019)
G010.901−00.070	Snake	34	10.901	−0.070	25.8	4.9	4.1	0.2	Reid et al. (2019)
G011.916+00.726	N4 Bubble	8	11.916	0.726	26.9	3.8	2.81	0.29	...
G012.910+00.460	IRAS−18089−1732	8	12.910	0.460	32.8	1.3	2.5	0.3	Reid et al. (2019)
G012.960−00.168	W33 (G012.80−00.20)	85	12.960	−0.168	37.8	7.6	2.6	0.2	Reid et al. (2019)
G015.079−00.625	M17	55	15.079	−0.625	19.5	1.9	2	0.1	Reid et al. (2019)
G016.996+00.719	M16	62	16.996	0.719	22.1	2.8	1.86	0.13	...
G018.936−00.350	W39	58	18.936	−0.350	63.8	3.4	4	1.1	...

Note. A more complete table that includes parameters for all 877 clusters is available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.125.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/.

2.4 Summary of distances

There are a total of 8417 clumps in the $3° < |\ell| < 60°$ region. 6673 of these clumps have been associated with one of the 877 associations identified by the ℓbv friends-of-friends analysis. The remaining 1744 are considered to be isolated. We have allocated distances through the methods described above to 8130 clumps in this region (∼97 per cent). This re-analysis represents a modest increase (∼10 per cent) in the number of distances presented in Urquhart et al. (2018) and so the overall Galactic distribution shown in Fig. 1 is broadly unchanged. The overall reliability and completeness of the ATLASGAL catalogue, however, have been improved.

3 PHYSICAL PROPERTIES

In this section, we describe the methods used to determine physical properties for the 8417 clumps located within the $3° < |\ell| < 60°$ region for which a distance has been determined. We have previously derived the clump masses, peak column densities, bolometric luminosity, sizes, and virial parameter in Urquhart et al. (2018), and

we refer the reader to that paper for a detailed description. One issue we noted in that paper was a trend for higher masses and sizes as a function of evolution (i.e. larger for H II regions compared to quiescent sources, suggesting an evolutionary trend). This trend is caused by the heating of the clump by the embedded young star(s), which results in more of the clump's envelope being raised above the detection threshold. This effect produces an observational bias in which the clumps appear to increase in size and their volume densities decrease with evolution (see fig. 33 in Urquhart et al. 2018).

Billington et al. (2019) demonstrated that the observational bias can be completely eliminated by using the FWHM source sizes (determined from the pixels above the half power level) and masses (see fig. 9 of that paper). These FWHM parameters are therefore much more reliable when comparing properties of different evolutionary subsamples. Following the method described by Billington et al. (2019), we have recalculated the FWHM sizes, masses, H$_2$ volume density (n), and free-fall times for the clumps using the updated distances discussed in the previous section. We provide a brief description of how these source properties have been determined below.
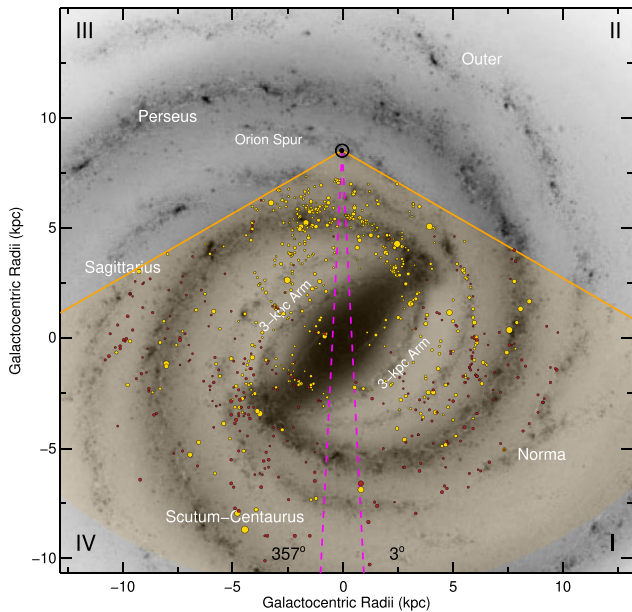
Figure 1. 2D images showing expected large-scale features of the Milky Way as viewed from the Galactic pole. The positions of ATLASGAL sources have been overlaid to facilitate comparison of their distribution to the large-scale structure of the Galaxy. The yellow circles show positions of the clusters while the red circles show the positions of individual clumps. The sizes of the circles give an indication of the masses. The background image is a schematic of the Galactic disc as viewed from the Northern Galactic Pole (courtesy of NASA/JPL-Caltech/R. Hurt (SSC/Caltech)). The Sun is located at the apex of the wedge and is indicated by the ⊙ symbol. The spiral arms are labelled in white and Galactic quadrants are given by the Roman numerals in the corners of the image. The magenta line shows the innermost region toward the GC where distances are not reliable.

The sizes of clumps are determined from the number of pixels within the FWHM contour, i.e. above 50 per cent of the peak of the ATLASGAL dust continuum emission (cf. Shirley et al. 2003):

$$R_{\text{fwhm}} = \sqrt{\frac{A}{\pi}}, \quad (1)$$

where A is the number of pixels within the 50-per-cent flux contour multiplied by the square of the pixel size in arcsec (i.e. 36 arcsec^2). The 50 per cent flux level is below the detection threshold for sources with an S/N below 6σ and cannot be reliably determined and so these sources are excluded. We also exclude clumps with fewer pixels than the beam integral ($\Theta_{\text{fwhm}}^2 \times 1.133 = 11.3$ pixels, where Θ_{fwhm} is 19.2 arcsec) as the flux measurements will be overestimated. Of the 8417 clumps considered here, 5866 satisfy these two conditions and are deconvolved from the ATLASGAL beam using

$$R_{\text{fwhm,decon}} = \sqrt{R_{\text{fwhm}}^2 - \left(\frac{\Theta_{\text{fwhm}}}{2}\right)^2} \quad (2)$$

The distances determined in Section 2 are used to convert the deconvolved radius into a physical radius, $R_{\text{fwhm, pc}}$. The FWHM clump mass, M_{fwhm}, is determined using

$$M_{\text{fwhm}} = \frac{D^2 \, S_{\nu,\text{fwhm}} \, \gamma}{B_\nu(T_{\text{dust}}) \kappa_\nu}, \quad (3)$$

where $S_{\nu,\text{fwhm}}$ is the integrated 870-μm flux density within the 50 per cent contour, D is the distance to the source, γ is the gas-to-dust mass ratio, B_ν is the Planck function for a dust temperature T_{dust}, and κ_ν is the dust absorption coefficient taken as 1.85 cm^2 g^{-1} (Schuller et al.

2009 and references therein). We use the dust temperatures derived from fits to the spectral-energy distributions of the clumps (SEDs; Urquhart et al. 2018) and assume that these values are reasonable estimates of the average temperatures of the inner parts of the clumps (see also König et al. 2017). The gas-to-dust ratio is calculated using the following relationship empirically determined by Giannetti et al. (2017):

$$\log(\gamma) = \left(0.087 \left[^{+0.045}_{-0.025}\right] \pm 0.007\right) R_{\text{gc}} + \left(1.44 \left[^{-0.45}_{+0.21}\right] \pm 0.03\right), \quad (4)$$

where R_{gc} is the Galactocentric distance expressed in kpc. The systematic uncertainties are given in the square brackets. This prescription gives a value of γ between 130 and 145 at the distance of the Sun, which is in very good agreement with the local value of 136 (see Giannetti et al. 2017 for a detailed discussion).

We calculate the mean FWHM volume density by dividing the M_{fwhm} by the volume:

$$n(\text{H}_2) = \frac{3}{4\pi} \frac{M_{\text{fwhm}}}{\mu m_{\text{p}} R_{\text{fwhm}}^3}, \quad (5)$$

where $n(\text{H}_2)$ is the hydrogen particle density per cm^{-3}, μ is the mean molecular weight per hydrogen atom (taken as 2.8; Kauffmann et al. 2008), m_{p} is the mean proton mass, and M_{fwhm} and R_{fwhm} are as previously defined. We make the assumption that each clump is generally spherical and is not extended along the line of sight.

The clump free-fall times are useful in providing a lower limit to the star formation time-scales and can be readily determined from the mean densities using

$$\tau_{\text{ff}} = \sqrt{\frac{3\pi}{32 G \bar{\rho}}}, \quad (6)$$

where $\bar{\rho} = 3 M_{\text{fwhm}}/4\pi R_{\text{fwhm}}^3$ is the mean density of the clump. The free-fall times are typically within (1–3) $\times$ 10^5 yr, which compare well with the H II region phase lifetimes derived empirically by Mottram et al. (2011) and from a numerical model by Davies et al. (2011) ($\sim$(2–4) $\times$ 10^5 yr).

Finally, we define two additional distance-independent parameters, these being the luminosity-to-mass ratio (L/M_{fwhm}) and the mass surface density ($M_{\text{fwhm}}/\pi R_{\text{fwhm}}^2$). The luminosities have been taken directly from Urquhart et al. (2018); these have been rescaled to the distances presented here. The first of these two ratios is considered to be a good diagnostic for evolution and measure of the instantaneous star formation efficiency (Molinari et al. 2008; Urquhart et al. 2014a, 2018) and has been used and discussed extensively in previous ATLASGAL papers.

We provide a full set of the physical properties described in this section in Table 4. Distances are available for 8130 clumps in the sample (96.7 per cent) and masses, sizes and volume densities have been determined for 7992 (95 per cent) clumps, 5696 (67.8 per cent) clumps, and 5576 (66.2 per cent) clumps, respectively. We show the distributions of the parameters described in the paragraphs above in Fig. 2 and, for completeness, we also include the dust temperature and luminosity derived in Urquhart et al. 2018. The grey histograms show the distribution of all clumps for which a value is available, while the light-red shaded histograms show the distribution of the clumps classified as being either quiescent, protostellar, YSO, or H II regions. These groupings constitute our high-reliability 'evolutionary' sample of clumps that are described in the next section. We present a summary of the properties of the clumps in Table 5 280.0pt. We note that the clumps have sizes of $\sim$0.5 pc and masses of $\sim$650 M$_\odot$, and so are capable of forming a cluster of stars with at least one high-mass star (Csengeri et al. 2014).

Table 4. Derived clump parameters.

CSC name	Evolution type	T_{dust} (K)	V_{LSR} (km s^{-1})	Distance (kpc)	R_{GC} (kpc)	R_{fwhm} (pc)	Log[L_{bol}] (L$_\odot$)	Log[M_{fwhm}] (M$_\odot$)	Log[n_{fwhm}(H$_2$)] (cm^{-3})	τ_{ff} (Myr)	L_{bol}/M_{fwhm}-ratio (L$_\odot$/M$_\odot$)
AGAL003.008+00.111	Ambiguous	15.2	152.4	15.4	7.03	0.75	3.298	4.000	4.907	0.115	0.20
AGAL003.016+00.141	Ambiguous	12.7	151.1	15.4	7.03	1.78	2.831	4.734	4.522	0.179	0.01
AGAL003.016−00.179	Ambiguous	17.8	−4.8	21.3	12.97	4.31	4.595	5.066	3.703	0.834	0.34
AGAL003.021−00.067	Ambiguous	12.2	−0.9	2.9	5.44	0.35	1.852	2.955	4.856	0.104	0.08
AGAL003.028−00.094	Ambiguous	11.9	9.3	2.9	5.42	0.51	2.314	2.820	4.248	0.209	0.31
AGAL003.034+00.402	Quiescent	13.9	13.0	2.9	5.43	0.34	2.318	2.422	4.356	0.185	0.79
AGAL003.049+00.392	Ambiguous	15.7	11.6	2.9	5.43	0.27	1.656	2.502	4.740	0.119	0.14
AGAL003.049−00.047	Ambiguous	19.1	9.0	2.9	5.42	0.18	2.379	2.033	4.805	0.110	2.22
AGAL003.056+00.151	Ambiguous	13.5	153.4	15.4	7.03	1.18	3.047	4.201	4.525	0.179	0.07
AGAL003.093+00.422	Quiescent	9.7	22.7	10.7	2.40	1.87	3.035	4.127	3.855	0.243	0.08

Notes. Only a small portion of the data is provided here. The full table is available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.125.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/.
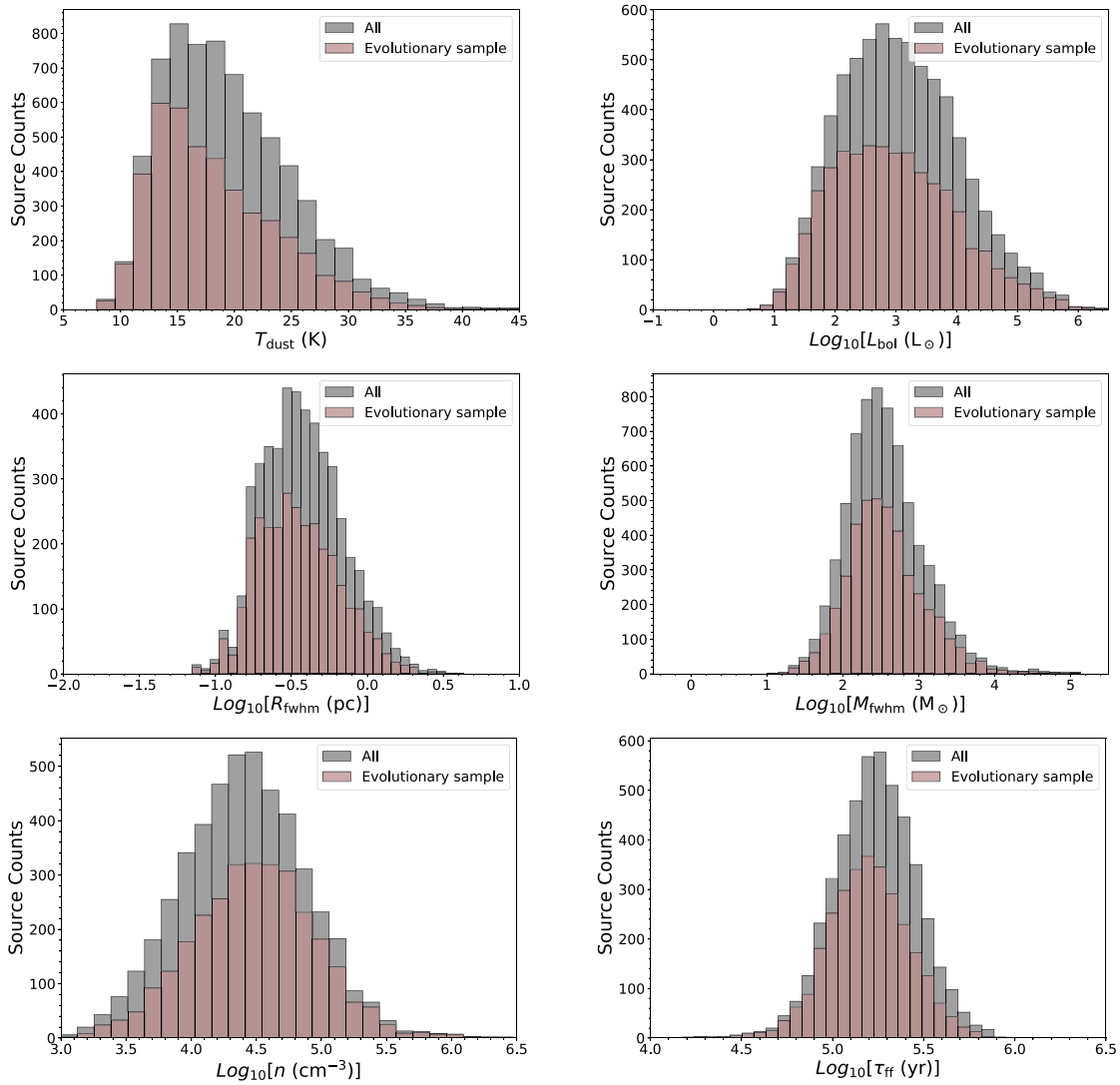

Figure 2. Distributions of clump properties discussed in Section 3, namely dust temperature and bolometric luminosity (top panels), FWHM size and mass (middle panels), and average density and free-fall time (bottom panels). The grey histograms show the distribution for the whole sample while the light-red shaded histograms show the distribution for the clumps identified in one of four evolutionary stages that are described in Section 4. We have excluded clumps within 2 kpc of the Sun to avoid biasing the distribution to nearby low-mass, low-luminosity clumps.

Table 5. Summary of physical properties.

Parameter	#	$\bar{x}$	$\sigma/\sqrt{(N)}$	σ	$x_{\min}$	$x_{\max}$
T_{dust} (K)	6853	19.416	0.069	5.678	7.9	56.1
$\mathrm{Log}_{10}[M_{fwhm} \, (\mathrm{M}_\odot)]$	6853	2.823	0.007	0.539	1.159	5.036
$\mathrm{Log}_{10}[L \, (\mathrm{L}_\odot)]$	6853	3.077	0.012	0.971	0.552	6.906
R_{fwhm} (pc)	4859	0.459	0.005	0.338	0.07	4.31
$\mathrm{Log}_{10}[n \, (\mathrm{cm}^{-3})]$	4779	4.512	0.007	0.453	2.942	6.895
τ_{ff} (Myr)	4779	0.195	0.002	0.105	0.011	1.041
$\mathrm{Log}_{10}[\Sigma \, (\mathrm{M}_\odot/\mathrm{pc}^2)]$	4708	1.837	0.006	0.441	0.219	4.072

Note. In Column 2, we give the number of clumps in each subsample; in Columns 3–5, we give the mean values, the error in the mean and the standard deviation, respectively; and in Columns 6–7, we give the minimum and maximum values of the samples, respectively.

We present an updated version of the 30 most massive complexes located within the Solar circle in Table 6. The content of this table was originally presented in table 8 of Urquhart et al. (2018) and has been modified to reflect the changes in association membership, distances and mass measurements. The majority of the most prominent complexes are unchanged and this table is primarily included for completeness.

4 EVOLUTIONARY SEQUENCE

4.1 Classification scheme

The large number of sources in the catalogue previously inhibited our ability to visually inspect the multiwavelength data to classify every clump reliably. To date, only a relatively modest fraction of the clumps has been reliably classified through visual inspection ($\sim$100; König et al. 2017) or by cross-matching with reliable H II region and YSO catalogues ($\sim$1000; Urquhart et al. 2013b, 2014a) such as the Red MSX survey (RMS; Lumsden et al. 2013),[3] and the CORNISH survey (Hoare et al. 2012; Purcell et al. 2013). The majority of the rest of the catalogue, therefore, has been automatically classified into one of three categories (quiescent, protostellar or YSO) depending on their spectral energy distributions (Urquhart et al. 2018). YSOs and H II regions have very similar SEDs and so, where available, radio emission was used to distinguish between these two evolutionary stages (e.g. Becker et al. 1994; Urquhart et al. 2007, 2009; Purcell et al. 2013; Medina et al. 2019).

Although these automatic classifications are useful for deriving statistical properties, they are not always reliable indicators of the various evolutionary stages, particularly towards intense star-forming complexes where clumps in different evolutionary stages are often found in close proximity to each other. Furthermore, as the brightness of the dust emission traced by ATLASGAL is a function of both the column density and the temperature, a detection can indicate a combination of high column density and low temperatures or vice versa. The catalogue is therefore also likely to contain sources with significant emission from warm diffuse gas and photon-dominated regions (PDRs) that are often found in star forming regions. These kinds of sources are not part of the evolutionary sequence and need to be removed to avoid unduly biasing the derived properties of any particular stage, and the only way this exclusion can be reliably achieved is through visual inspection of multiwavelength images.

The MALT90 team (Jackson et al. 2013) used mid-infrared images (3–24 µm) from the GLIMPSE (Churchwell et al. 2009) and MIPSGAL (Carey et al. 2009) surveys to classify $\sim$3000 clumps into four categories (i.e. PDR, H II region, protostellar, and quiescent; Rathborne et al. 2016). Our classification scheme follows a similar set of criteria but also takes advantage of the availability of the 70-µm images and catalogue that are now available from HiGAL (Molinari et al. 2010, 2016). These data are indispensable in revealing embedded protostars at longer wavelengths in clumps that may be missed, even at 24 µm. This scheme has been applied to all 8417 clumps located outside the GC region, regardless of whether they satisfy the FWHM criterion discussed in the previous section.

The SEDs of quiescent and star-forming clumps can be described as a modified version of blackbody radiation, which is referred to as a greybody. The temperatures of quiescent clumps are $\sim$10–15 K and, at these temperatures and with no internal heating source, these clumps are dark at wavelengths of 70 µm or less. The formation of a protostellar object inside a clump results in an increase in the global temperature of the clump due to accretion on to the embedded source. This luminosity warms the envelope and therefore shifts the peak of the SED to shorter wavelengths. As a result, the protostar becomes detectable at 70 µm, but not at mid-infrared wavelengths (i.e. $\leq$24 µm). As the protostar continues to accrete material, it will increase the global temperature of the clump and will start to appear first at mid-infrared and then near-infrared wavelengths. At this stage, we refer to the embedded protostar as a young stellar object (YSO). For a low- or intermediate-mass star, accretion will end and the YSO will continue to contract and disperse its circumstellar envelope and eventually join the main sequence. High-mass stars, however, produce a significant amount of their flux at UV wavelengths, which ionizes their local environments. This ionization creates a rapidly expanding bubble of hot, ionized gas, an H II region that can be easily distinguished by its compact radio emission and mid-infrared morphology.

Given the gradual changes in the SED as the embedded object first appears at 70 µm and then evolves towards the main sequence, it is possible to use a carefully selected set of far- and mid-infrared wavelength images (3–8, 24, and 70 µm) to initially distinguish between quiescent and star-forming clumps and, for the latter, to classify the embedded objects (e.g. protostellar, YSO, or H II region). This method has been successfully applied by König et al. (2017) to a small subsample of ATLASGAL clumps that had been specially selected to include examples of the four evolutionary stages discussed in this paragraph.

Fig. 3 illustrates the sequence of steps used to classify the clumps into one of five categories. This scheme is similar to the steps described by König et al. (2017) with one small modification. We previously distinguished between protostars and YSOs using an $\sim$22-

---

[3]The RMS catalogue has produced the largest and most reliable catalogue of H II regions and massive-YSOs to date.

Table 6. Physical parameters for the 30 most massive complexes identified, ordered by their total dense-gas mass.

Complex name (1)	Literature name (2)	# of members (3)	ℓ (°) (4)	b (°) (5)	V_{LSR} (km s^{-1}) (6)	Distance (kpc) (7)	R_{gc} (kpc) (8)	T (K) (9)	Log(L_{bol}) (L$_\odot$) (10)	L_{bol} (per cent) (11)	Log(M_{fwhm}) (M$_\odot$) (12)	M_{fwhm} (per cent) (13)	L_{bol}/M_{fwhm} (L$_\odot$/M$_\odot$) (14)
G030.653−00.017	W43	250	30.65±0.47	−0.02±0.17	97.5±9.0	4.90	4.83	19.9±6.1	6.56	2.91	5.10	1.83	29.12
G043.141−00.018	W49	14	43.14±0.05	−0.02±0.03	9.5±4.0	11.10	7.60	25.6±4.6	7.18	12.02	5.04	1.60	137.95
G049.261−00.318	W51	93	49.26±0.22	−0.32±0.10	60.9±6.9	5.20	6.33	22.3±5.1	7.04	8.72	5.01	1.50	106.82
G336.996−00.002		107	337.00±0.19	−0.00±0.14	−73.9±4.6	6.36	4.51	21.6±6.1	6.51	2.60	4.99	1.42	33.63
G024.322+00.187		61	24.32±0.16	0.19±0.09	113.3±3.9	7.60	3.44	17.9±5.4	5.73	0.43	4.89	1.13	6.90
G333.121−00.353	G333	245	333.12±0.44	−0.35±0.23	−50.8±4.3	3.60	5.41	21.5±5.7	6.52	2.64	4.82	0.96	50.41
G033.527−00.011		26	33.53±0.19	−0.01±0.03	103.3±2.2	8.80	4.96	18.1±5.5	5.47	0.24	4.79	0.91	4.78
G305.453+00.065	G305	102	305.45±0.27	0.06±0.20	−37.5±3.9	4.00	6.85	22.9±6.4	6.34	1.74	4.76	0.84	38.33
G035.569+00.015	G35.58−0.03	15	35.57±0.05	0.01±0.08	52.4±3.2	10.40	6.06	21.2±4.8	5.69	0.38	4.57	0.54	13.18
G331.389−00.106		87	331.39±0.28	−0.11±0.12	−89.0±5.0	3.90	5.27	21.6±6.1	6.24	1.37	4.50	0.46	54.70
G336.469−00.166		29	336.47±0.08	−0.17±0.09	−84.5±5.2	10.20	4.21	23.6±2.6	6.20	1.25	4.47	0.43	53.42
Bania Clump 1		17	354.75±0.08	0.32±0.06	89.9±9.8	8.40	0.76	16.6±5.4	5.40	0.20	4.46	0.42	8.81
G023.392−00.166	W41	35	23.39±0.07	−0.17±0.08	99.3±3.4	5.90	3.76	20.6±6.3	5.99	0.78	4.45	0.42	34.44
G020.737−00.089		17	20.74±0.04	−0.09±0.06	56.6±1.8	11.70	4.86	23.1±3.7	5.76	0.46	4.44	0.40	21.12
G023.072−00.349	W41	41	23.07±0.23	−0.35±0.10	73.8±7.3	5.00	4.23	16.9±4.1	5.10	0.10	4.42	0.38	4.83
G331.094−00.414		53	331.09±0.21	−0.41±0.06	−65.8±2.2	4.00	5.23	20.0±6.2	5.82	0.52	4.39	0.36	26.94
G006.183−00.358	W28	78	6.18±0.29	−0.36±0.17	12.6±4.9	3.00	5.38	18.4±4.6	5.67	0.37	4.37	0.34	20.18
G336.951+00.000		30	336.95±0.16	0.00±0.08	−118.3±3.4	7.70	3.27	21.6±4.9	5.86	0.58	4.36	0.33	31.74
G037.684−00.209	W47	18	37.68±0.18	−0.21±0.13	58.0±6.5	9.80	6.03	23.9±4.2	5.99	0.77	4.34	0.32	44.10
G045.486+00.071		13	45.49±0.05	0.07±0.06	58.6±2.2	7.70	6.23	23.1±7.0	6.17	1.17	4.32	0.31	69.99
G019.627−00.110		14	19.63±0.06	−0.11±0.03	59.4±2.8	11.60	4.63	21.6±5.1	5.53	0.27	4.32	0.31	16.29
G008.581−00.329	W30	38	8.58±0.17	−0.33±0.05	37.1±1.3	4.40	4.00	17.4±5.8	5.36	0.18	4.32	0.31	10.94
G018.936−00.350	W39	58	18.94±0.12	−0.35±0.20	63.8±3.4	4.00	4.75	19.3±4.8	5.41	0.20	4.30	0.29	12.84
G321.082−00.516		13	321.08±0.06	−0.52±0.03	−60.8±0.7	9.30	5.93	24.0±4.0	5.73	0.42	4.30	0.29	26.88
G024.049+00.495		31	24.05±0.27	0.50±0.06	96.5±3.7	5.80	3.85	16.9±4.5	5.72	0.41	4.27	0.27	27.96
G010.903−00.068	Snake	35	10.90±0.19	−0.07±0.11	25.7±4.9	4.10	4.39	15.8±6.2	4.72	0.04	4.26	0.27	2.88
G340.248−00.273		61	340.25±0.13	−0.27±0.13	−49.0±3.5	3.60	5.14	17.7±4.7	5.46	0.23	4.26	0.26	16.01
G332.124−00.070		67	332.12±0.30	−0.07±0.12	−49.2±1.8	3.10	5.79	17.7±5.7	5.00	0.08	4.25	0.26	5.71
G350.216+00.082		33	350.22±0.08	0.08±0.05	−65.5±5.2	6.00	2.67	20.4±4.5	5.99	0.77	4.24	0.25	55.53
G012.960−00.168	W33	85	12.96±0.26	−0.17±0.15	37.8±7.6	2.60	5.85	16.3±4.2	4.98	0.08	4.24	0.25	5.50

Notes. In the first column, we give the complex name, which is constructed from the mean longitude and latitude position of member clumps. In Column 2, we give the more common name found in the literature. In Columns 3–6, e give the number of clumps in the association, and the mean position and velocity of the associated members and their standard deviation. In Columns 7 and 8, we give the distance to the complex and the corresponding distance from the GC. In Column 8 we give the mean and standard deviation for the temperature determined by averaging over the associated clumps. In Columns 10 and 12, we give the total luminosity and mass of the clumps in each complex, while in Columns 11 and 13, we give the total luminosity and mass as a percentage of total luminosity and mass in the disc (as determined in this paper). In the final column, we give the luminosity to mass ratio determined by summing up the total luminosity of all clumps and dividing it by the total mass of the associated clumps (i.e. dividing Column 10 by Column 12). A more complete table that includes parameters for all 877 is available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.125.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/.
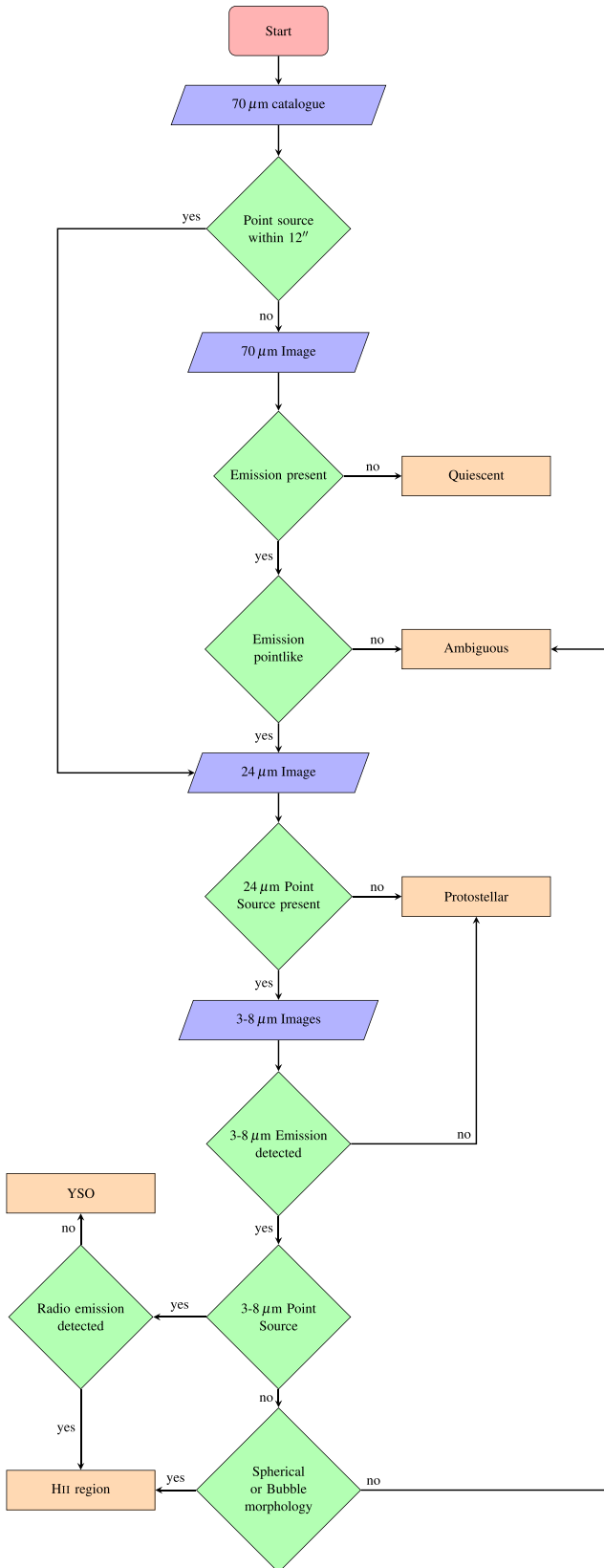
Figure 3. Flowchart showing the sequence of steps taken and the images used to classify the ATLASGAL clumps into one of five categories.

μm flux threshold of 2.7 Jy,[4] but this is somewhat arbitrary, given the improved sensitivities of WISE (Wright et al. 2010) and MIPSGAL (Carey et al. 2009) and so we have relaxed this criterion and simply use the presence or absence of an 8 μm point source to distinguish between a YSO and protostar.

We have applied this set of criteria to GLIMPSE 3–8-μm, MIPSGAL 24-μm, and Hi-GAL 70-μm postage-stamp images of all of the ATLASGAL sources for which the data are available (i.e. $|b| < 1°$) to produce reliable samples of the different evolutionary stages. The assignment of evolutionary stages described above works well when the regions are relatively isolated and are dominated by a single infrared object. Given the sizes and masses of these clumps ($\sim$0.5 pc and $\sim$500 M$_\odot$), however, it is reasonable to assume that these will all form clusters and so we should expect to find different evolutionary stages present in the same clump. In these cases, we classify the clump evolutionary stage according to the most evolved embedded object, assuming that it will dominate the clump's physical properties.

Below, we provide a brief description of the six of the most common categories identified using 8, 24, and 70 μm. The first four of these represent an evolutionary sequence (see Fig. 4 for examples):

(i) Quiescent: These clumps are cold (10–15 K), devoid of any embedded objects and dark at 70 μm (see the upper panel of Fig. 4 for an example). A first step was to cross-match the ATLASGAL clumps with the Hi-GAL 70-μm point source catalogue (Elia et al. 2017) and exclude any clumps where a match was found within 12 arcsec of the submm emission peak.[5] The rationale for this criterion is that these clumps are already undergoing star formation. This step reduced the sample that needed to be inspected by a factor of 2. The next step was to inspect the 70-μm maps to ensure that the central part of the clumps (as bounded by the 50 per cent contour of the 870-μm emission) is clear of any large-scale extended 70-μm emission that might hide an embedded point source.

(ii) Protostellar: This stage is the earliest in the star formation process when the protostar begins to warm its natal clump and becomes visible at 70 μm. A comparison with the HiGAL 70-μm catalogue revealed the presence of a 70-μm point source located within 12 arcsec of the centre of the clump, causing them to be immediately flagged as being associated with star formation. If these sources have no counterpart at 3–8 μm within 12 arcsec of the centre of the clump, they are classified as protostellar (see upper-middle panel of Fig. 4 for an example). We initially made a distinction between 70-μm sources that had a 24-μm counterpart and those that did not. We found no significant difference in the physical properties of these two groups (i.e. dust temperature, luminosity, and luminosity-to-mass ratio, all of which are considered to be good evolutionary diagnostics) and so decided that the distinction was not useful and merged them.

(iii) YSO: At this point in the protostar's evolution, it has warmed its natal clump sufficiently enough to be visible at both 3–8 and 24 μm (see the lower middle panel of Fig. 4 for an example). The protostar and YSOs are much smaller than the resolution of the GLIMPSE image ($\sim$2 arcsec) and so the emission at all mid-infrared
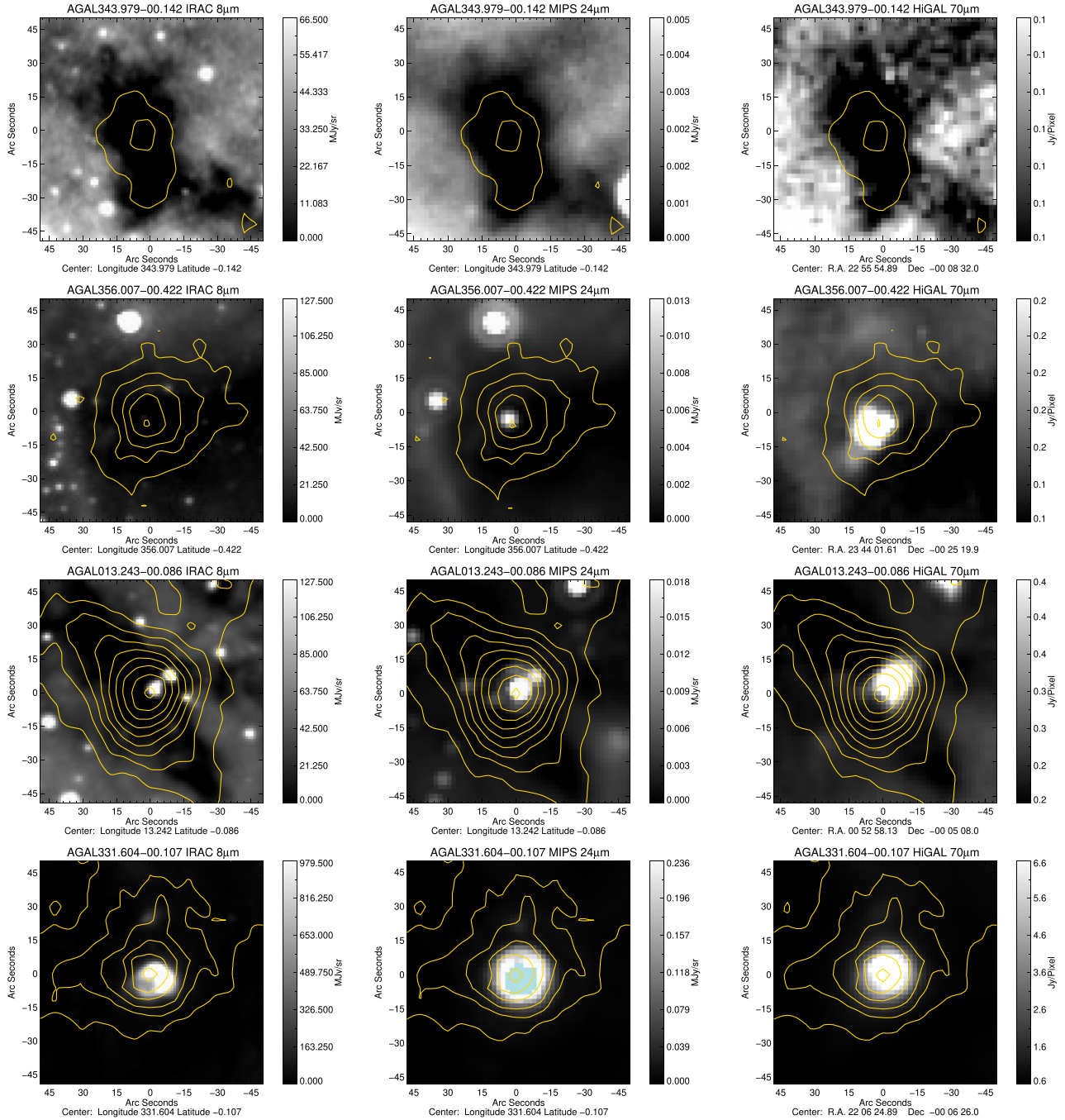

Figure 4. Examples of the 8-μm IRAC, 24-μm MIPS, and 70-μm HiGAL images used in the classification process. From the top to down, we show an example of a quiescent, protostellar, YSO, and H II region clump. The contours show the distribution of the 870-μm dust emission traced by ATLASGAL. These start at 3σ and increase in steps determined using the dynamic range power-law fitting scheme where $D = N^i + 3$, where D is the dynamic range, S_{peak}/σ, N is the number of contours, in this case 8, and i is the power law that determines the separation between consecutive contours. This approach is a modified version of a scheme developed by Thompson et al. (2006). The light-blue coloured region seen in the lower middle panel indicates the presence of saturated pixels.

wavelengths will be unresolved (Mottram et al. 2007; De Buizer et al. 2005).

(iv) H II region: These are the most evolved sources in embedded stages investigated and as such will be the warmest and most clearly observable at all of the wavelengths inspected. Due to the fact that they are surrounded by a rapidly expanding ionized nebula, we would expect all but the very youngest of these regions to be extended in the 3–8 μm images and many will be clearly associated with bubbles

(see the lower panel of Fig. 4 for an example). The only reliable way of identifying the youngest (hypercompact and ultracompact) H II regions is to search for a correlation between the peak of the submillimetre emission and compact radio emission at frequencies of ~5 GHz. Coincidence of radio emission with the extended 3–8 μm is also useful in confirming that they are indeed H II regions. Our analysis has identified two populations of H II regions, which we refer to as radio-loud and radio-quiet. Comparing the properties of these
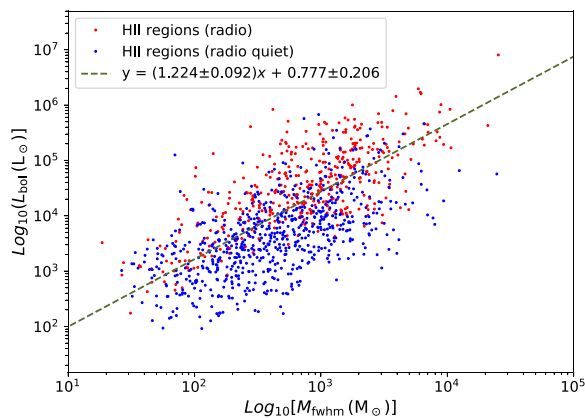
Figure 5. Correlation between luminosity and mass for all H II regions identified in the sample. The dashed green line shows the results of a linear least-squares fit to the distance limited sample of H II regions (i.e. distances between 2 and 4 kpc).

two types of H II regions, we noticed that the former are warmer and more luminous and consequently have higher luminosity-to-mass ratios. Looking at the luminosity–mass plot shown in Fig. 5, however, we see that the two samples form a continuous distribution in this parameter space. We note that the radio-loud H II regions are almost an order of magnitude more luminous than the radio-quiet H II regions. This difference suggests that, while all of these H II regions are associated with high-mass stars, the radio-quiet H II regions are primarily associated with early B-type stars, based on their luminosities seen in Fig. 5 ($M_\star < 20\,\mathrm{M_\odot}$). For lower mass stars, while the intensity of the radio emission quickly drops below the detection limit of the radio surveys.

(v) Ambiguous: In many cases, the ATLASGAL clump is coincident with significant amounts of extended mid-infrared emission across most of one or more images, and no distinguishable embedded point source is seen. These sources are then classified as 'ambiguous.' This appearance is indicative of regions of warm diffuse gas that may obscure the presence of any protostellar objects at a similar or cooler temperature. There a few different kinds of sources that fall into this category such as clumps associated with PDRs found around the edges of evolved H II regions, and star formation complexes where many sub-regions at different stages of evolution are located in close proximity to each other. This ambiguity is not to say that many of these regions are not associated with star formation, but that it is impossible to estimate the fraction that are so associated, or their evolutionary state. It is, therefore, not possible to make a reliable determination about the current level of star formation in these clumps and so they are excluded from the statistical analysis that follows. In total, 3395 clumps have been classified as ambiguous.

(vi) Planetary nebulae (PNe): These form during the latest stages in the life cycle of lower-mass evolved stars when they are shedding their outer layers. These objects are therefore enshrouded in an envelope of hot dust that can be detected in the ATLASGAL survey. PNe are the only unresolved sources detected in the ATLASGAL catalogue and appear to be isolated and featureless in mid-infrared images (star-forming regions are nearly always associated with dust lanes and/or nebulosity) and their SEDs peak at mid-infrared wavelengths. Only four of these have been detected in ATLASGAL. Given their rarity, these are not shown explicitly in the flowchart presented in Fig. 3.
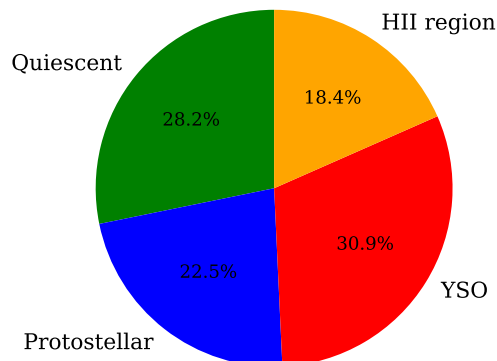

Figure 6. Breakdown of the proportions of the clumps from the distance-limited sample in the four evolutionary stages. The Poisson uncertainties on these values are of the order of 1 per cent and so the differences are significant.

4.2 Classification statistics for embedded objects

We have four reliable evolutionary subsamples, each comprising a similar number of clumps. If we look at a distance-limited sample (i.e. see Fig. 6), however, we find that quiescent and YSO-associated clumps are the most numerous, each with approximately 30 per cent, with the protostellar and H II regions having ∼20 per cent each. Together, these make up approximately 60 per cent of the classified sample, with the remaining 40 per cent being classified as Ambiguous. In Fig. 2, we show the distributions of physical properties for the whole ATLASGAL sample and for those sources associated with one of the four evolutionary stages described in the previous section. It is interesting to note the similarities between the distributions of the two samples and this suggests that many of the clumps excluded from one of the evolutionary samples may also be associated with star formation but that the images are complicated or hard to interpret. We present a summary of results of the classification process giving the statistics and mean and standard deviation for the physical properties in Table 7.

If we assume that the fraction of sources in each stage is proportional to a similar fraction of the star formation time-scale, the statistics suggest that the lifetimes of the quiescent and YSO stages are similar to each other but 50 per cent longer than the protostellar and H II region stages. Although the fraction of quiescent sources is approximately a factor of 2 higher than reported in Urquhart et al. (2018), the number of quiescent sources is actually similar. The reason for the increase in this fraction is that a larger proportion of the sources previously assigned to one of the other three source types is now classified as being Ambiguous, as these are more often associated with diffuse or complicated far- and mid-infrared emission. The impact of this reclassification can be seen in the temperature distribution plot shown in the upper left-hand panel of Fig. 2, which shows that the fraction of clumps classified as one of the four star-forming types decreases with increasing temperature. The fraction of quiescent sources given here is therefore likely to be an upper limit and the true fraction is likely to be somewhere between between 12 and 28 per cent.

In the left-hand panels of Fig. 7, we show the cumulative distribution functions for the sizes, masses, and volume densities for all four evolutionary stages. Inspection of the sizes reveals a significant difference between the quiescent clumps and the other three evolutionary stages (KS-test returns a *p*-value ≪0.0013). The sizes of the other three stages are all similar. Turning our attention to the clump masses, we notice a trend for decreasing mass in the centres of the clumps as the embedded objects evolve

Table 7. Summary of mean and standard deviations for the clump properties by evolutionary stage.

Classification type	Number	T_{dust} (K)	Distance (kpc)	R_{fwhm} (pc)	Log$[M_{fwhm}]$ ($M_\odot$)	Log$[n_{H_2}]$ (cm^{-3})	Log$[L_{bol}]$ ($L_\odot$)	Log$[L_{bol}/M_{fwhm}]$ ($L_\odot\,M_\odot^{-1}$)
				Full sample				
Quiescent	1218	13.8 ± 2.56	4.4 ± 2.78	0.43 ± 0.35	2.511 ± 0.515	4.537 ± 0.449	2.09 ± 0.76	-0.42 ± 0.59
Protostellar	1010	15.7 ± 3.46	4.5 ± 3.13	0.34 ± 0.25	2.408 ± 0.558	4.702 ± 0.492	2.42 ± 0.88	0.02 ± 0.70
YSO	1543	18.4 ± 3.89	5.1 ± 3.64	0.35 ± 0.28	2.350 ± 0.627	4.614 ± 0.509	2.86 ± 0.85	0.51 ± 0.65
H II Region	1236	24.4 ± 4.87	6.5 ± 4.33	0.46 ± 0.34	2.517 ± 0.687	4.445 ± 0.574	3.82 ± 0.95	1.30 ± 0.64
				Distance limited sample (2–4 kpc)				
Quiescent	572	13.7 ± 2.74	3.1 ± 0.49	0.30 ± 0.11	2.384 ± 0.291	4.660 ± 0.336	1.92 ± 0.60	-0.47 ± 0.60
Protostellar	457	15.5 ± 3.55	3.1 ± 0.50	0.25 ± 0.09	2.301 ± 0.334	4.790 ± 0.337	2.26 ± 0.71	-0.04 ± 0.71
YSO	626	17.9 ± 3.84	3.1 ± 0.52	0.23 ± 0.09	2.189 ± 0.348	4.780 ± 0.367	2.63 ± 0.67	0.44 ± 0.66
H II Region	373	24.5 ± 5.50	3.2 ± 0.52	0.24 ± 0.09	2.166 ± 0.373	4.698 ± 0.463	3.49 ± 0.80	1.33 ± 0.66

towards the main sequence (see the middle left-hand panel of Fig. 7; the *p*-values for the quiescent and protostellar clumps is 0.0006 and for the protostellar and YSOs it is 0.0002, confirming the observed difference is statistically significant). Looking at the volume density distribution (lower left-hand panel of Fig. 7) we see that the protostellar and YSO distributions are indistinguishable from each other. The quiescent clumps have a similar-shaped curve to that of the YSOs and protostars but shifted to lower densities (*p*-value $\ll 0.0013$). The H II region distribution is significantly different from the other three stages, having a similar shape as the YSOs and protostars at high volume densities but a shallower slope, and extends to include some of the lowest volume densities. The differences observed for some of these parameters could be attributed to the evolution of the embedded object. We will see in the next section, however, that these differences are in fact the result of observational biases.

In the right-hand panels of Fig. 7, we show the cumulative distribution functions for temperature, luminosity, and luminosity-to-mass ratio. All of these parameters increase as a function of evolution, as expected. The increase in these parameters has been reported in previous studies (Elia et al. 2017; König et al. 2017; Urquhart et al. 2018) and so these results are not new but demonstrate that we have produced a large and reliable set of evolutionary subsamples of clumps.

5 DISCUSSION

5.1 Physical correlations

We illustrate the luminosity–mass and mass–radius distributions by evolutionary source type (as identified in Section 4) in Fig. 8. The linear least-squares fit to the luminosity–mass distribution of a distance-limited sample (2–4 kpc) of the H II regions (shown in Fig. 5) gives a slope of 1.224 ± 0.0916 and intercept of 0.777 ± 0.206 compared with the value of 1.314 ± 0.019 determined in Urquhart et al. (2018). The change in the way the mass is determined has resulted in a decrease in the steepness of the slope, but the new slope is consistent with the older one within 3σ. The slope fit to the mass–radius distribution of the full star-forming sample shown in Fig. 8 is 1.448 ± 0.039, which is again slightly shallower than the value of 1.647 ± 0.012 determined in Urquhart et al. (2018).

We show the radius, mass, and volume density distributions as functions of the luminosity-to-mass (L_{bol}/M_{fwhm}) ratio in Fig. 9. The distributions of the mass and volume density are similar, with the evolutionary type changing with increasing L_{bol}/M_{fwhm}-ratio as we would expect. In the previous section, we noted some significant

trends in the cumulative distribution functions of these parameters but cautioned that these are likely to be due to observational bias rather than the result of evolution of an embedded object. There is no obvious correlation between size and evolution (upper panel; the Spearman rank correlation coefficient is $r = -0.087$). A cursory glance at the mass and density distribution, however, suggests that these parameters are correlated with evolution, particularly in the mass-L_{bol}/M_{fwhm}-ratio plot. This trend is largely due to the presence of a significant deficit of clumps towards the lower left region of the parameter space (middle and lower panels of Fig. 9). ATLASGAL is a flux-limited survey. Since the clump mass is a function of flux and dust temperature, it has a poorer mass sensitivity for colder objects leading to the deficits in the Fig. 9 and resulting in the perceived change in mass as a function of evolution noted in the middle and lower left-hand panels of Fig. 7. One final interesting feature seen in the volume density distribution plot (lower panel of Fig. 9) is a small but noticeable increase in the volume density above a L_{bol}/M_{fwhm}-ratio of $10\,L_\odot/M_\odot$. This is possibly related to the faster free-fall collapse times found for higher-density clumps, resulting in lifetimes so short that they are very unlikely to be observed in the earlier evolutionary stages. These high-density and rapidly evolving clumps are only seen towards the end of their evolution because it takes them a while to disperse their natal material once the H II region has formed.

If we limit the analysis to a mass and density range in which we are complete (i.e. $M_{fwhm} > 200\,M_\odot$ and n_{fwhm} between 3×10^4 and $4 \times 10^5\,cm^{-3}$), then the differences between the different evolutionary stages disappear (the Spearman rank correlation coefficients are $r = -0.03$ and $r = -0.02$ for the M_{fwhm}-L_{bol}/M_{fwhm}-ratio and n_{fwhm}-L_{bol}/M_{fwhm}-ratio, respectively, for the ranges mentioned). There is, therefore, no obvious correlation between the radius, mass, or volume density with evolution (i.e. the overall distributions are flat). The lack of any significant change in the clump mass and density as a function of the L_{bol}/M_{fwhm}-ratio further supports these findings (see also Urquhart et al. 2018 and Billington et al. 2019). This result is somewhat surprising, given that significant global infall motions in clumps are often reported in the literature (e.g. (0.3–16) $\times 10^{-3}\,M_\odot\,yr^{-1}$; Wyrowski et al. 2016), which would lead to a significant increase in the clump mass if these rates were to continue over the star formation time frame (i.e. several 10^5 yr).

Recent analysis by Jackson et al. (2019) of a sample of ~1000 MALT90 clumps found that, while high-mass stars are undergoing gravitational collapse, the blue asymmetry in the line emission measured for quiescent, protostellar and compact H II regions was significantly larger than for more extended H II regions and PDRs. These results suggest that global collapse is underway even in the earliest quiescent stage but decreases in the later stages. This
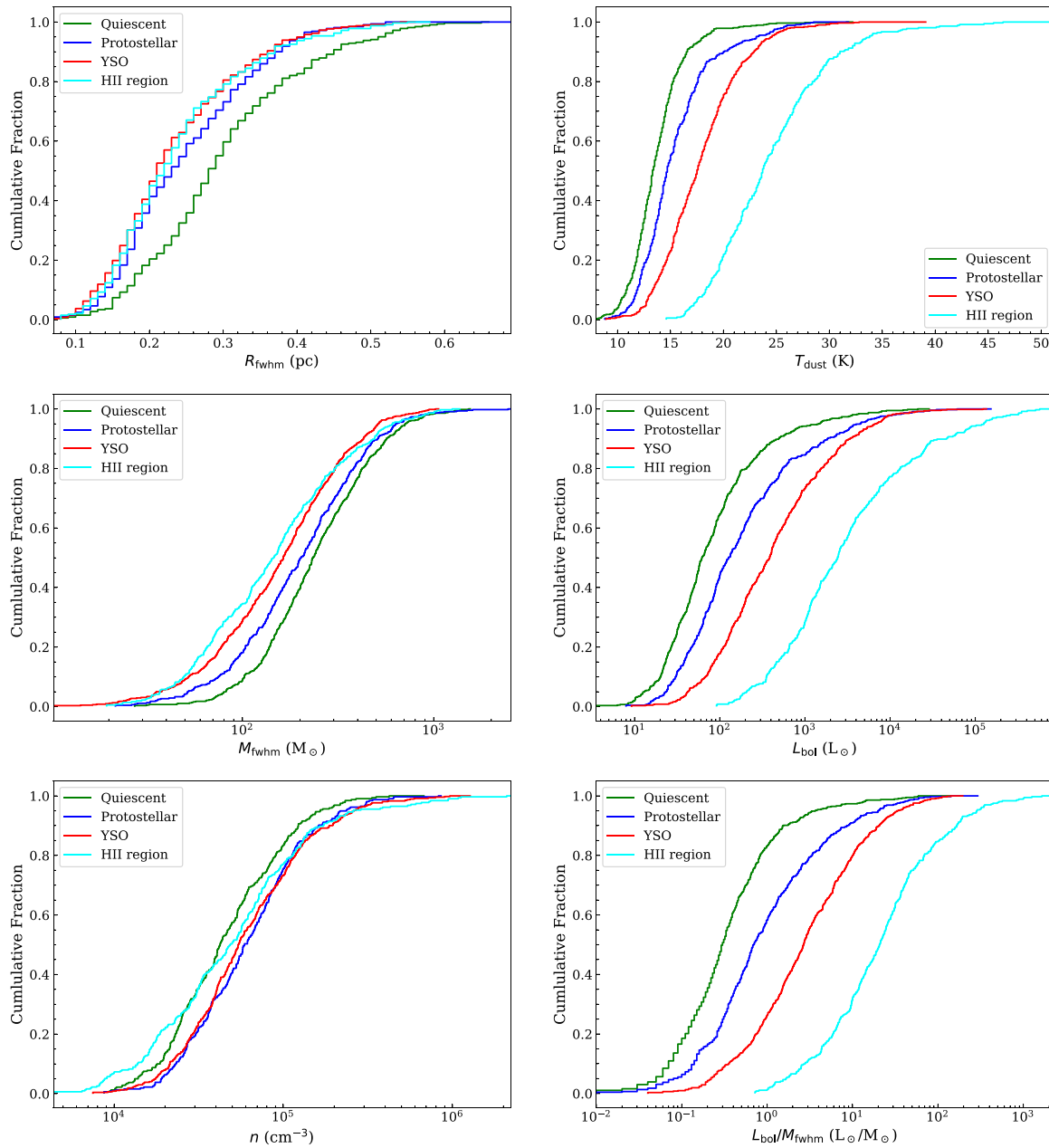
Figure 7. Cumulative distribution functions for the main evolutionary categories. The left-hand panels show the distributions for the FWHM radius, mass, and volume densities, which are similar for all of the subsamples. The right-hand panels show the dust temperature, bolometric luminosity, and the luminosity-to-mass ratio. These latter values are all good evolutionary diagnostics and show very clear differences between the different evolutionary subsamples. For all of these plots, we have used a distance-limited sample of 2–4 kpc to avoid any possible distance bias.

hypothesis is supported by a more recent set of observations of the $J = 3$–2 and 4–3 rotational transitions of the HCO^+, HNC, HCN, and N_2H^+ towards a sample of ATLASGAL clumps that found the incidence of infall motions decreases as the evolutionary stage of the clumps increases (Neupane et al. in preparation), presumably due to the increasing radiative and mechanical feedback from the forming proto-cluster. In fact, the gas motions of three of the nine sources investigated by Wyrowski et al. (2016) were found to be dominated by strong outflowing/expanding motions. Two of these were classified as UC H II regions and one as an MYSO and so these can all be considered to be relatively evolved. A decreasing infall rate may significantly limit any increase in the clump masses and would

be consistent with our findings. The results reported by Jackson et al. (2019), however, are indicative rather than definitive and, while the results obtained from the higher excitation transitions reported by Neupane et al. (in preparation) show a clear trend, they are drawn from a relatively small number of sources and further work is required confirm this hypothesis.

The process of turning the gas and dust into stars will also lead to a decrease in the clump mass over time; however, some of the gas used to create stars is expected to be replaced by infalling material on to the clump from the parental GMC. Star formation efficiencies (ε) are expected to range from 5 to 30 per cent (e.g. Lada & Lada 2003; Rugel et al. 2019). If we assume an efficiency to be ~20 per cent during the
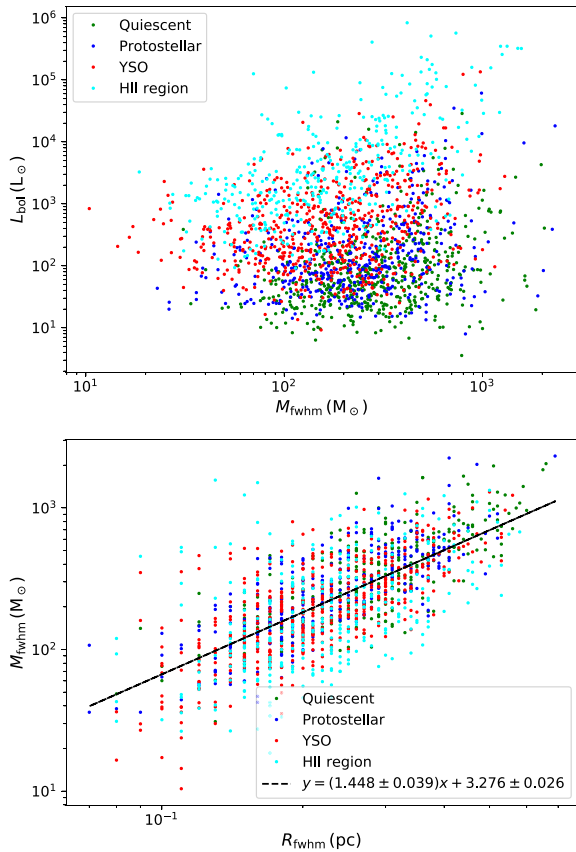
Figure 8. Luminosity–mass and mass–radius relationships are shown as a function of the evolutionary source type for all clumps in the distance-limited sample.

star formation process and compare this to the uncertainties in clump masses (which is a factor of a few) we find that we are unlikely to notice this change, particularly since any decrease in mass is likely to be counteracted by an increase in the mass due to infall. Although there is no evidence of any significant change in the clump mass during the star formation process it does not preclude changes in the internal structure on smaller scales. In fact, we expect gas to continue to contract to form higher density substructures and to feed accretion on to forming protostellar objects (e.g. Csengeri et al. 2017, 2018). Rather, it is just that these motions are not observed on the scales ATLASGAL is probing (i.e. $R_{fwhm} \approx 0.5$–1 pc). Indeed, a recent deep study of a $2° \times 1°$ field by Rigby et al. (2021) reported a small but significant increase in the average clump mass from the earliest to the middle stages, before decreasing towards the later stages. Their observations have a modest improvement in resolution compared to ATLASGAL (12 arcsec, corresponding to 0.2–0.4 pc), which might explain the difference in the results. The way the masses are measured, frequency of the observations and coverage, however, are all quite different making a direct comparison difficult.

5.2 Linking the L_{bol}/M_{fwhm}-ratio to a star formation time-scale

The right-hand panels of Fig. 7 show the cumulative distribution functions for temperature, luminosity, and luminosity-to-mass ratio for the four evolutionary subsamples. All of these distributions look relatively smooth as one would expect if the properties of these subsamples are normally distributed. If, however, these observation-
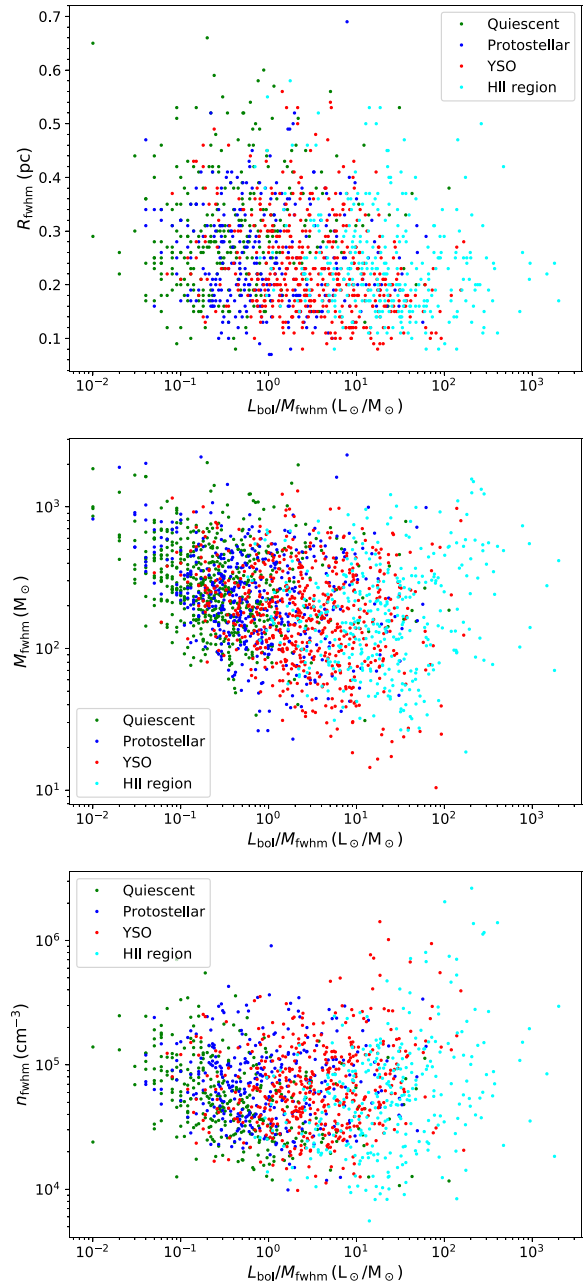

Figure 9. The radius, mass, and volume density for the the evolutionary source types as a function of the luminosity-to-mass ratio. These plots include all clumps in the distance-limited sample.

ally classified subsamples do represent distinct stages in the star formation process, we might expect to see points of inflection and jumps in the cumulative distribution function at the transition points of a clump's evolution.

We show the cumulative distribution for the combined evolutionary sample in Fig. 10. In this plot and those that follow, we use the full evolutionary sample as L_{bol}/M_{fwhm}-ratio is a distance-independent parameter and there is no significant difference between the full sample and the distance-limited sample. This plot shows that the central part of the curve is still smooth and devoid of any features that are coincident with the mean values of L_{bol}/M_{fwhm}-ratios for the four evolutionary stages. This lack of abrupt changes suggests that star
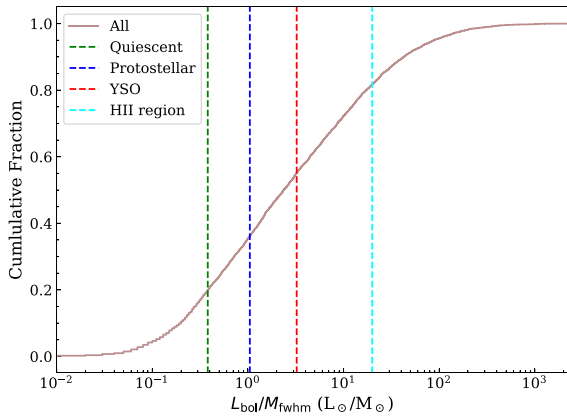
Figure 10. Cumulative distribution function for the L_{bol}/M_{fwhm}-ratio for the combined star formation sample (rose curve). The vertical lines indicate the mean of the lognormal distribution of the four subsamples identified that constitute the star formation sample.

formation is a rather smooth and continuous process. Also, the stages are useful in identifying groups of protostellar objects with similar properties and/or ages but do not themselves represent fundamentally different stages or changes in the physical mechanisms involved. The curve flattens out at both ends but the source numbers are too low in these regions of the parameter space to draw any conclusions from that behaviour. We can speculate as to their nature, however. The steepening of the slope at the low-L_{bol}/M_{fwhm}-ratio end of the curve is possibly related to the onset of star formation and the fact this steepening is seen below the mean of the quiescent stage indicates that star formation is ongoing in many of these apparently 'starless' clumps. This hypothesis is strongly supported by the association of many of the quiescent clumps with outflows as discussed in the following subsection. On the other hand, the turnover of the slope at the high-L_{bol}/M_{fwhm}-ratio end starts just after the mean of the H II region stage, which is possibly related to the increasingly disruptive influence of high Lyman photon flux and expanding ionization front.

We can take our analysis of the L_{bol}/M_{fwhm}-ratio a step further by linking it to evolutionary time-scales. If we assume that the number of sources in each L_{bol}/M_{fwhm} bin is proportional to the fraction of the star formation time-scale that a protostar spends in each L_{bol}/M_{fwhm} range, then the y-axis of the cumulative distribution function is also proportional to the star's formation time. In Fig. 11, we have re-plotted Fig. 10 but have switched the axes to have the independent variable on the x-axis and have changed the label to accordingly. These plots show how the L_{bol}/M_{fwhm}-ratio changes during the star formation process. Looking in a little more detail at the curve presented in the lower panel of Fig. 11 we see that nearly all clumps fall on a smooth curve of increasing L_{bol}/M_{fwhm}-ratio and so are therefore increasing their luminosity very slowly in the very early stages. This also reveals that the rate of increase in the L_{bol}/M_{fwhm}-ratio accelerates as the embedded object evolves through the protostellar and YSO stages before turning into what appears to be a runaway process when it reaches the H II region stage. This is to be expected, as rapidly expanding ionized bubbles begin to disrupt and dissipate their natal environment.

In the early stages of a protostar's evolution, the majority of the luminosity is generated by accretion rather than fusion and so the acceleration of the L_{bol}/M_{fwhm}-ratio with evolution is therefore related to an increase in the accretion rate over time for the earliest stages. High-mass stars reach the main sequence while still deeply embedded
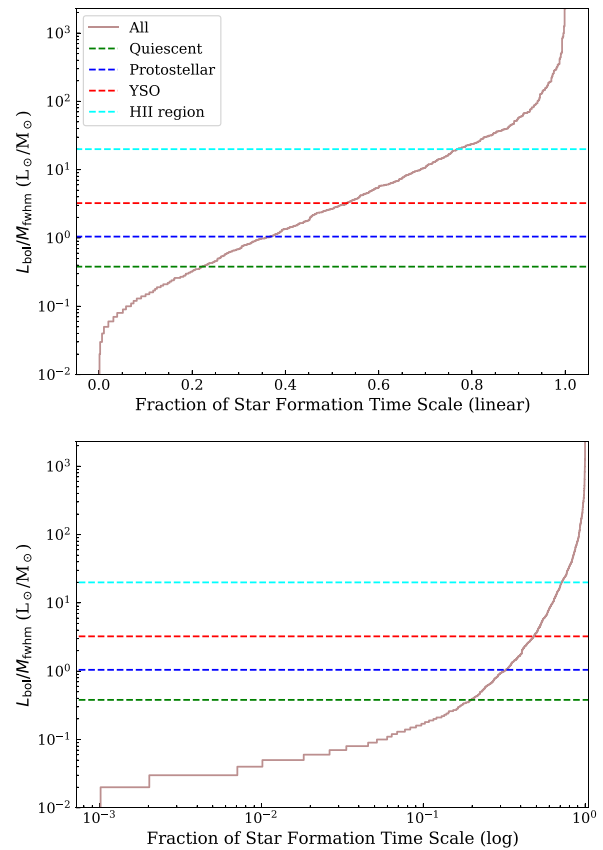

Figure 11. The L_{bol}/M_{fwhm}-ratio as a function of the fractional free-fall collapse times of the clumps. The two plots show the same data, but in the upper panel, the x-axis is linear, and in the lower panel, it is plotted on a log scale. The horizontal lines indicate the mean of the lognormal distribution of the four evolutionary types.

and can continue to accrete material and so in the later stages an increasing proportion of their luminosity will be generated by fusion, which is perhaps why we see a shape increase in the L_{bol}/M_{fwhm}-ratio for the most evolved sources. The steady increase can be seen in the upper panel of Fig. 11 lends support for high-mass star formation models where the rate of accretion also increases over time, which were previously found to be consistent with the results reported by Davies et al. (2011) from the modelling of the RMS survey (Lumsden et al. 2013).

In Section 3, we used the mean volume densities to estimate the free-fall time-scales for the clumps. The calculation of the free-fall time by definition does not take account of internal support or feedback, both of which will slow down the collapse process and result in longer collapse times. The free-fall time-scales, however, were a few times 10^5 yr, i.e. similar to the H II region formation time-scales derived empirically by Mottram et al. (2011) and from a numerical model by Davies et al. (2011). This similarity suggests that, although the free-fall times are a lower limit, they are perhaps only a few times shorter than the star formation time-scale and can therefore be useful to gain some insight to how the star formation process proceeds.

In Fig. 12, we have plotted the L_{bol}/M_{fwhm} curve for two small ranges in volume density as a function of the star formation time-scale (upper panel) and as a function of the corresponding free-fall time calculated in Section 3 (lower panel). A linear-least-square fit
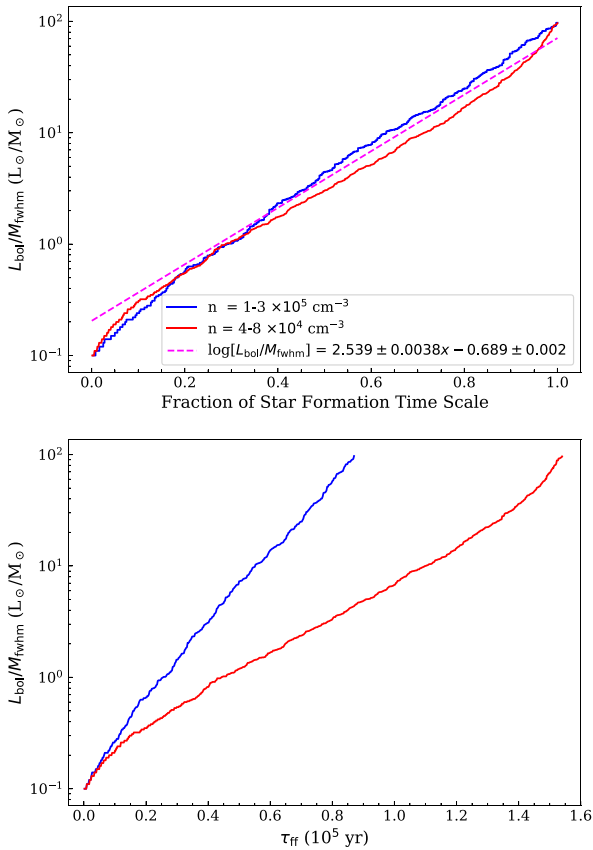
Figure 12. The L_{bol}/M_{fwhm}-ratio as a function of the star formation time-scale (upper panel) and as a function of the free-fall collapse times of the clumps (lower panel) for two volume density ranges. We have restricted the L_{bol}/M_{fwhm} range to be between 0.1 and 100 $L_\odot/M_\odot$ to concentrate the analysis on the region of parameter space for which we have reliable statistics. A linear least-squares fit to the L_{bol}/M_{fwhm}-ratio track of all clumps with volume densities within $(1-50) \times 10^5 \, \mathrm{cm}^{-3}$ is shown in the upper panel.

to the log of the L/M-ratio is shown in the upper panel and gives a power-law relation between the L/M-ratio and the star formation time-scale of $10^{2.5 \times \tau_{ff} - 0.7}$. Comparing the tracks for the two different density ranges reveals them to be very similar and a KS-test returns a *p*-value of 0.03 confirming that the tracks are not significantly different. From this similarity, we conclude that the process by which stars accrete material is not influenced by the mean volume density of the clumps. Hence, the star formation time-scale and the relative time in the different stages both scale with clump density and so with free-fall time. The process is therefore similar and scaleable, it just goes faster in denser clumps. The lower panel of Fig. 12 reveals

how the L_{bol}/M_{fwhm}-ratio changes as a function of free-fall time for different density ranges, with the denser clumps forming stars much more quickly, as expected.

5.3 Correlation with other samples

We have cross-matched the ATLASGAL sample with four other star formation tracers to test the validity of the evolutionary sequence identified in this paper. We have used the catalogue of ∼300 extended green objects (EGOs; Cyganowski et al. 2008), catalogues of molecular outflows produced by Maud et al. (2015), de Villiers et al. (2014), Yang et al. (2018), and Yang et al. (2022) methanol multibeam (MMB) survey (Green et al. 2009) and water masers identified by HOPS (Walsh et al. 2011, 2014). EGOs are identified by their enhanced emission in the 4.5-μm band that contains the rotationally excited H_2 ($v = 0-0$, S(9, 10, 11)) and CO ($v = 1-0$) band-head lines, which are indicative of outflow activity and active accretion (Cyganowski et al. 2008). Molecular outflows are intimately associated with accretion discs and are therefore a strong indication that star formation is taking place in a clump. Class II Methanol masers (Menten 1991) are thermally pumped and require the presence of high densities and a strong mid-infrared radiation source and so are thought to be associated exclusively with the closest environment of high-mass protostellar objects (Minier et al. 2003; Breen et al. 2013; Billington et al. 2019). Water masers are collisionally pumped and, when encountered in star formation environments, are thought to be excited in the cavities of molecular outflows.

We give the statistics of these matches as a function of the evolutionary stage in Table 8. For each tracer, we give the total number of associations, the overall percentage, and the percentage for the region where all of the surveys overlap. It is the latter of these that is most useful when comparing trends in the association rates as a function of evolution. We show a Venn diagram of the correlation between the different tracers in Fig. 13. There are no strong correlations between the different tracers, which would suggest they are somewhat random.

It is clear from Table 8 that the association rates increase as a function of evolution, and this trend holds even up to the H II region stage for all but the EGOs. For these, the association rate peaks at the YSO stage and then begins to decrease. It is also interesting to note that the association rates for EGOs, water, and methanol masers with clumps classified as quiescent are effectively zero, which is consistent with their classification as starless (as suggested by the lack of a 70-μm point source). We do, however, find that a significant fraction are associated with molecular outflows, which is a strong indication that star formation is already underway in them. Fig. 14 shows the cumulative distribution functions for the L_{bol}/M_{fwhm}-ratio of clumps associated with each of the four star

Table 8. Summary of associations with other star formation tracers by evolutionary stage.

Classification type	Number	EGOs (#)	(per cent)	(per cent)	H₂O masers (#)	(per cent)	(per cent)	Methanol masers (#)	(per cent)	(per cent)	Outflows (#)	(per cent)	(per cent)
Quiescent	1206	1	0.1	0	1	0.1	0.3	2	0.2	0.0	106	8.8	15.7
Protostellar	941	46	4.9	7.4	37	3.9	5.9	97	10.3	11.5	120	12.8	19.7
YSO	1510	120	7.9	8.9	90	6.0	10.9	231	15.3	16.8	257	17.0	23.8
H II	1222	51	4.2	5.6	116	9.5	14.8	242	19.8	19.2	359	29.4	41.6
Total	8417	257	3.1		317	3.8		723	8.6		1458	17.3	

Note. For each tracer, we give the total number of associations, the overall percentage and the percentage for the region where all of the surveys overlap (i.e. $300° < \ell < 350°$ and $|b| < 0.5°$).
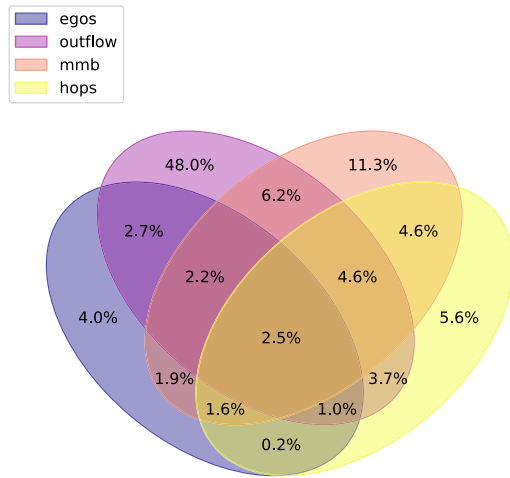
Figure 13. Venn diagram showing the correlation between the four different star formation tracers associated with ATLASGAL clumps.

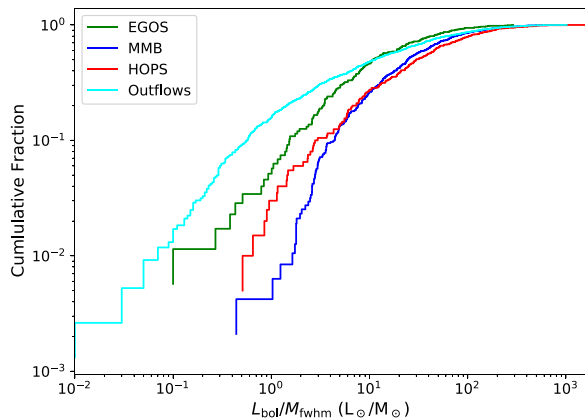**Figure 14.** Rate of change in the $L_{\rm bol}/M_{\rm fwhm}$-ratio as a function of the star formation tracers. See Section 5.3 for details of the surveys used and how the associations have been made.

formation tracers. This plot reveals that molecular outflows are the earliest signpost for star formation, followed by EGOs, and then by water and methanol masers (for more detailed analysis of methanol and water masers and star formation evolutionary sequences, see Breen et al. 2018; Billington et al. 2020; Ladeyschikov et al. 2020).

The number of clumps associated with molecular outflows is likely a lower limit, as the sensitivity of the available CO data is often not sufficient to detect high-velocity wings in the spectra. If one includes only the clumps towards which good CO data are available in the statistical analysis, the association rates approximately three times larger (Yang et al. 2018). This result suggests that star formation is already under way in perhaps half of the quiescent clumps. This behaviour is consistent with the observed increases in the $L_{\rm bol}/M_{\rm fwhm}$-ratio seen in Fig. 10 prior to the appearance of a protostellar object and explains why there is no sharp jump in the cumulative distribution curve between the quiescent and protostellar stages. Comparing the $L_{\rm bol}/M_{\rm fwhm}$-ratio for quiescent clumps with and without an associated outflow (see Fig. 15) reveals that those associated with outflows have significantly larger ratios (p-value $\ll 0.0013$) and are therefore more evolved.
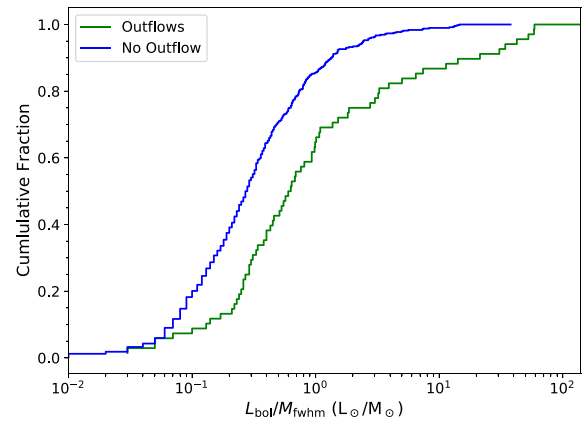

Figure 15. The $L_{\rm bol}/M_{\rm fwhm}$-ratio for quiescent clumps with and without outflow associations. The mean values for the $L_{\rm bol}/M_{\rm fwhm}$-ratio of clumps with and without outflows are 7.8 and 0.8 $L_\odot/M_\odot$, respectively.

We therefore conclude that the presence of molecular outflows is the earliest signpost for star formation – even preceding maser activity – and that star formation is already underway in a significant number of quiescent clumps identified in the ATLASGAL catalogue.

6 SUMMARY AND CONCLUSIONS

The ATLASGAL compact source catalogue consists of approximately 10 000 clumps located across the inner Galactic plane. In this paper, we use new molecular-line information to determine reliable velocities and distances to many hundreds of clumps located outside of the GC ($3° < |\ell| < 60°$). We used these new distances to refine the membership of associations of clumps determined using a friends-of-friends analysis. In total, 877 associations with two or more clump members have been identified. We also use the distances to determine a complete set of unbiased physical properties for the clumps (masses, luminosities, radii, volume densities, and free-fall times). Finally, we have performed a visual inspection of multiwavelength images (8–870 μm, including *Spitzer*, *WISE*, and *HiGAL* bands) of all sources located outside the GC to identify robust evolutionary samples (quiescent, protostellar, YSO, and H II region) and compared their physical properties to identify trends in the data that provide insight into the star formation process.

The main results of this analysis are as follows:

(i) We have allocated velocities to 8899 clumps of the 9817 clumps located in the main ATLASGAL survey region (i.e. $300° < \ell < 60°$ and $|b| < 1°.5$), corresponding to ∼90 per cent of the whole sample. Excluding clumps within $3°$ of the GC, where kinematic distances are unreliable, results in a sample of 8417 clumps. We have determined distances to 8130 of these (∼97 per cent).

(ii) The visual classification has resulted in the identification of 5007 clumps in one of four evolutionary stages (1218 quiescent, 1010 protostellar, 1543 YSO, and 1236 H II region). The fact that we have identified roughly equal numbers in each of the four evolutionary stages suggests that the statistical lifetimes of each stage are comparable. These are considered to be a high-reliability sample of clumps with well-determined physical properties and visual classification.

(iii) A comparison of the physical properties reveals that the masses, sizes, and densities are similar for all four evolutionary stages, suggesting that the mass and structure of the clumps do

not change significantly during the star formation process (typically a few times 10^5 yr). Conversely, we find significant and systematic increases in the temperature, luminosity, and luminosity-to-mass ratio that are consistent with the evolutionary stages identified.

(iv) We have investigated the correlations between the luminosity, mass and size and find that the luminosity–mass and mass–radius relations are broadly consistent with previously reported values. These relations are therefore fairly robust and are relatively insensitive to the method used to determine the parameter values.

(v) We find that the cumulative distribution function of the L_{bol}/M_{fwhm}-ratio is very smooth and reveals no significant changes in the slope or jumps at the mean values determined for the four evolutionary stages or in the intervening regions where one would expect the embedded objects transition between one stage and another. This behaviour leads us to conclude that star formation is a rather monotonic and continuous process and the observational stages, while useful in identifying groups of protostellar objects with similar properties and/or ages, do not in themselves represent fundamentally different stages or changes in the physical mechanisms involved on clump scales.

(vi) If we assume that the number of clumps in each L_{bol}/M_{fwhm}-ratio interval reflects the fraction of the statistical lifetime spent in each interval, we are able to equate the cumulative distribution function to a statistical lifetime. This step allows us to investigate how the L_{bol}/M_{fwhm}-ratio changes with time, and doing so we find that the rate of change of L_{bol}/M_{fwhm}-ratio continues to accelerate as the clump evolves, turning into a runaway process towards the end. This behaviour supports star formation models in which the accretion rate increases over time.

(vii) Cross-matching the evolutionary sample with four other star formation samples (outflows, EGOs, and methanol and water masers) we find an overall association rate of ~15–20 per cent. There is a clear correlation between association rate and evolutionary stage, with the former increasing as a function of evolution. The association rate for EGOs and masers is effectively zero for the quiescent stage, as one would expect for starless clumps. A significant fraction of these sources are associated with molecular outflows, however, indicating that star formation is already underway in many such clumps. Molecular outflows are therefore the earliest manifestations of star formation, appearing before the protostar becomes visible in the far-infrared or maser emission is detected.

ACKNOWLEDGEMENTS

[6]http://www.astropy.org..

DATA AVAILABILITY

The data presented in this paper is available from a dedicated website (https://atlasgal.mpifr-bonn.mpg.de). The full version of Tables 2–4 and 6 are only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.125.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/MNRAS/.

REFERENCES

Anderson L. D., Bania T. M., 2009, ApJ, 690, 706
Araya E., Hofner P., Churchwell E., Kurtz S., 2002, ApJS, 138, 63
Astropy Collaboration et al., 2013, A&A, 558, A33
Astropy Collaboration et al., 2018, AJ, 156, 123
Battisti A. J., Heyer M. H., 2014, ApJ, 780, 173
Becker R. H., White R. L., Helfand D. J., Zoonematkermani S., 1994, ApJS, 91, 347
Billington S. J. et al., 2019, MNRAS, 490, 2779
Billington S. J. et al., 2020, MNRAS, 499, 2744
Breen S. L. et al., 2018, MNRAS, 474, 3898
Breen S. L., Ellingsen S. P., Contreras Y., Green J. A., Caswell J. L., Stevens J. B., Dawson J. R., Voronkov M. A., 2013, MNRAS, 435, 524
Busfield A. L., Purcell C. R., Hoare M. G., Lumsden S. L., Moore T. J. T., Oudmaijer R. D., 2006, MNRAS, 366, 1096
Carey S. J. et al., 2009, PASP, 121, 76
Churchwell E. et al., 2009, PASP, 121, 213
Contreras Y. et al., 2013, A&A, 549, A45
Csengeri T. et al., 2014, A&A, 565, A75
Csengeri T. et al., 2016, A&A, 586, A149
Csengeri T. et al., 2017, A&A, 600, L10
Csengeri T. et al., 2018, A&A, 617, A89
Cyganowski C. J. et al., 2008, AJ, 136, 2391
Davies B., Hoare M. G., Lumsden S. L., Hosokawa T., Oudmaijer R. D., Urquhart J. S., Mottram J. C., Stead J., 2011, MNRAS, 416, 972
De Buizer J. M., Radomski J. T., Telesco C. M., Piña R. K., 2005, ApJS, 156, 179
de Villiers H. M. et al., 2014, MNRAS, 444, 566
Downes D., Wilson T. L., Bieging J., Wink J., 1980, A&AS, 40, 379
Duarte-Cabral A. et al., 2021, MNRAS, 500, 3027
Dunham M. K., Rosolowsky E., Evans N. J. II, Cyganowski C., Urquhart J. S., 2011, ApJ, 741, 110
Eden D. J. et al., 2020, MNRAS, 498, 5936
Elia D. et al., 2017, MNRAS, 471, 100
Elia D. et al., 2021, MNRAS, 504, 2742
Fish V. L., Reid M. J., Wilner D. J., Churchwell E., 2003, ApJ, 587, 701
Giannetti A. et al., 2014, A&A, 570, A65
Giannetti A. et al., 2017, A&A, 606, L12
Giannetti A., Wyrowski F., Leurini S., Urquhart J., Csengeri T., Menten K. M., Bronfman L., van der Tak F. F. S., 2015, A&A, 580, L7
Green J. A. et al., 2009, MNRAS, 392, 783
Green J. A., McClure-Griffiths N. M., 2011, MNRAS, 417, 2500
Güsten R., Nyman L. Å., Schilke P., Menten K., Cesarsky C., Booth R., 2006, A&A, 454, L13
Hoare M. G. et al., 2012, PASP, 124, 939
Hunter J. D., 2007, Compu. Sci. Eng., 9, 90
Immer K., Schuller F., Omont A., Menten K. M., 2012, A&A, 537, A121
Jackson J. M. et al., 2013, Publ. Astron. Soc. Aust., 30, e057
Jackson J. M. et al., 2019, ApJ, 870, 5
Jackson J. M., Bania T. M., Simon R., Kolpak M., Clemens D. P., Heyer M., 2002, ApJ, 566, L81
Kauffmann J., Bertoldi F., Bourke T. L., Evans N. J. II, Lee C. W., 2008, A&A, 487, 993
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kim W. J., Urquhart J. S., Wyrowski F., Menten K. M., Csengeri T., 2018, A&A, 616, A107
Kim W. J., Wyrowski F., Urquhart J. S., Menten K. M., Csengeri T., 2017, A&A, 602, A37

Kolpak M. A., Jackson J. M., Bania T. M., Clemens D. P., Dickey J. M., 2003, ApJ, 582, 756

König C. et al., 2017, A&A, 599, A139

Kroupa P., 2001, MNRAS, 322, 231

Lada C. J., Lada E. A., 2003, ARA&A, 41, 57

Ladeyschikov D. A., Urquhart J. S., Sobolev A. M., Breen S. L., Bayandina O. S., 2020, AJ, 160, 213

Li G.-X., Urquhart J. S., Leurini S., Csengeri T., Wyrowski F., Menten K. M., Schuller F., 2016, A&A, 591, A5

Lin Y., Csengeri T., Wyrowski F., Urquhart J. S., Schuller F., Weiss A., Menten K. M., 2019, A&A, 631, A72

Lumsden S. L., Hoare M. G., Urquhart J. S., Oudmaijer R. D., Davies B., Mottram J. C., Cooper H. D. B., Moore T. J. T., 2013, ApJS, 208, 11

Mattern M. et al., 2018, A&A, 619, A166

Maud L. T., Moore T. J. T., Lumsden S. L., Mottram J. C., Urquhart J. S., Hoare M. G., 2015, MNRAS, 453, 645

McClure-Griffiths N. M., Dickey J. M., Gaensler B. M., Green A. J., Haverkorn M., Strasser S., 2005, ApJS, 158, 178

Medina S. N. X. et al., 2019, A&A, 627, A175

Menten K. M., 1991, ApJ, 380, L75

Minier V., Ellingsen S. P., Norris R. P., Booth R. S., 2003, A&A, 403, 1095

Moisés A. P., Damineli A., Figuerêdo E., Blum R. D., Conti P. S., Barbosa C. L., 2011, MNRAS, 411, 705

Molinari S. et al., 2010, PASP, 122, 314

Molinari S. et al., 2016, A&A, 591, A149

Molinari S., Pezzuto S., Cesaroni R., Brand J., Faustini F., Testi L., 2008, A&A, 481, 345

Motte F., Bontemps S., Louvet F., 2018, ARA&A, 56, 41

Mottram J. C. et al., 2011, ApJ, 730, L33 +

Mottram J. C., Hoare M. G., Lumsden S. L., Oudmaijer R. D., Urquhart J. S., Sheret T. L., Clarke A. J., Allsopp J., 2007, A&A, 476, 1019

Nagayama T., Omodaka T., Handa T., Honma M., Kobayashi H., Kawaguchi N., Ueno Y., 2011, PASJ, 63, 719

Navarete F. et al., 2019, A&A, 622, A135

Pandian J. D., Menten K. M., Goldsmith P. F., 2009, ApJ, 706, 1609

Pandian J. D., Momjian E., Goldsmith P. F., 2008, A&A, 486, 191

Price S. D., Egan M. P., Carey S. J., Mizuno D. R., Kuchar T. A., 2001, AJ, 121, 2819

Purcell C. R. et al., 2013, ApJS, 205, 1

Rathborne J. M. et al., 2016, Publ. Astron. Soc. Aust., 33, e030

Reid M. J. et al., 2014, ApJ, 783, 130

Reid M. J. et al., 2019, ApJ, 885, 131

Reid M. J., Dame T. M., Menten K. M., Brunthaler A., 2016, ApJ, 823, 77

Rigby A. J. et al., 2021, MNRAS, 502, 4576

Roman-Duval J., Jackson J. M., Heyer M., Johnson A., Rathborne J., Shah R., Simon R., 2009, ApJ, 699, 1153

Rugel M. R. et al., 2019, A&A, 622, A48

Sanna A. et al., 2014, ApJ, 781, 108

Sanna A., Reid M. J., Moscadelli L., Dame T. M., Menten K. M., Brunthaler A., Zheng X. W., Xu Y., 2009, ApJ, 706, 464

Sato M., Reid M. J., Brunthaler A., Menten K. M., 2010, ApJ, 720, 1055

Schuller F. et al., 2009, A&A, 504, 415

Schuller F. et al., 2017, A&A, 601, A124

Schuller F. et al., 2021, MNRAS, 500, 3064

Sewilo M., Watson C., Araya E., Churchwell E., Hofner P., Kurtz S., 2004, ApJ, 154, 553

Shirley Y. L., Evans Neal J. I., Young K. E., Knez C., Jaffe D. T., 2003, ApJS, 149, 375

Stead J. J., Hoare M. G., 2010, MNRAS, 407, 923

Stil J. M. et al., 2006, AJ, 132, 1158

Tang X. D. et al., 2018, A&A, 611, A6

Thompson M. A., Hatchell J., Walsh A. J., MacDonald G. H., Millar T. J., 2006, A&A, 453, 1003

Urquhart J. S. et al., 2009, A&A, 501, 539

Urquhart J. S. et al., 2011, MNRAS, 410, 1237

Urquhart J. S. et al., 2012, MNRAS, 420, 1656

Urquhart J. S. et al., 2013a, MNRAS, 431, 1752

Urquhart J. S. et al., 2013b, MNRAS, 435, 400

Urquhart J. S. et al., 2014a, MNRAS, 443, 1555

Urquhart J. S. et al., 2014b, A&A, 568, A41

Urquhart J. S. et al., 2015, MNRAS, 446, 3461

Urquhart J. S. et al., 2018, MNRAS, 473, 1059

Urquhart J. S. et al., 2019, MNRAS, 484, 4444

Urquhart J. S. et al., 2021, MNRAS, 500, 3050

Urquhart J. S., Busfield A. L., Hoare M. G., Lumsden S. L., Clarke A. J., Moore T. J. T., Mottram J. C., Oudmaijer R. D., 2007, A&A, 461, 11

Virtanen P. et al., 2020, Nat. Methods, 17, 261

Walsh A. J. et al., 2011, MNRAS, 416, 1764

Walsh A. J., Purcell C. R., Longmore S. N., Breen S. L., Green J. A., Harvey-Smith L., Jordan C. H., Macpherson C., 2014, MNRAS, 442, 2240

Watson C., Araya E., Sewilo M., Churchwell E., Hofner P., Kurtz S., 2003, ApJ, 587, 714

Wienen M. et al., 2015, A&A, 579, A91

Wienen M., Wyrowski F., Menten K. M., Urquhart J. S., Walmsley C. M., Csengeri T., Koribalski B. S., Schuller F., 2018, A&A, 609, A125

Wienen M., Wyrowski F., Schuller F., Menten K. M., Walmsley C. M., Bronfman L., Motte F., 2012, A&A, 544, A146

Wright E. L. et al., 2010, AJ, 140, 1868

Wu Y. W. et al., 2014, A&A, 566, A17

Wyrowski F., Güsten R., Menten K. M., Wiesemeyer H., Csengeri T., König C., Urquhart J. S., 2016, EAS Publ. Ser., 281

Xu Y., Moscadelli L., Reid M. J., Menten K. M., Zhang B., Zheng X. W., Brunthaler A., 2011, ApJ, 733, 25

Xu Y., Reid M. J., Menten K. M., Brunthaler A., Zheng X. W., Moscadelli L., 2009, ApJ, 693, 413

Yang A. Y. et al., 2022, A&A In Press, preprint (2021arXiv211110850Y)

Yang A. Y., Thompson M. A., Urquhart J. S., Tian W. W., 2018, ApJS, 235, 3

Zhang B., Reid M. J., Menten K. M., Zheng X. W., Brunthaler A., Dame T. M., Xu Y., 2013, ApJ, 775, 79

Zinnecker H., Yorke H. W., 2007, ARA&A, 45, 481

This paper has been typeset from a T$_{\rm E}$X/L$^{\rm A}$T$_{\rm E}$X file prepared by the author.